Spring 5-31-2005

Collaborative examinations in asyncronous learning networks : field experiments on collaborative learning through online assessments

Jia Shen
New Jersey Institute of Technology

ABSTRACT

COLLABORATIVE EXAMINATIONS IN ASYNCHRONOUS LEARNING NETWORKS: FIELD EXPERIMENTS ON COLLABORATIVE LEARNING THROUGH ONLINE ASSESSMENTS

by
Jia Shen

With the proliferation of computer networks and the emergence of virtual teams, learning and knowledge sharing in the online environment has become an increasingly important topic. Applying constructivism and collaborative learning theories to assessment, the *collaborative online exam* is designed featuring students' active participation in various phases of the exam process through small group activities online. A *participatory online exam* process is designed featuring similar procedures except that students' involvement in each phase of the exam is individual. The collaborative online exam and the participatory online exam are investigated regarding student exam study strategies, group process, exam outcomes, faculty satisfaction, and exam efficiency. A 1*3 field experiment was conducted to compare three exam modes: the traditional exam, the participatory exam, and the collaborative exam. Results show that the collaborative examination significantly enhanced interaction and promoted higher order learning. In particular, small group activities in the online learning process significantly increased interactions among students which enhanced their sense of an online learning community. Active involvement in the online exams significantly reduced the use of surface learning in exam study. Overall, students reported significantly higher perceptions of learning in the collaborative exam than the other exam modes.

COLLABORATIVE EXAMINATIONS IN ASYNCHRONOUS LEARNING NETWORKS: FIELD EXPERIMENTS ON COLLABORATIVE LEARNING THROUGH ONLINE ASSESSMENTS

by
Jia Shen

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Information Systems

Department of Information Systems

May 2005

APPROVAL PAGE

COLLABORATIVE EXAMINATIONS IN ASYNCHRONOUS LEARNING NETWORKS: FIELD EXPERIMENTS ON COLLABORATIVE LEARNING THROUGH ONLINE ASSESSMENTS

Jia Shen

Dr. Starr Roxanne Hiltz, Dissertation Advisor — Date
Distinguished Professor of Information Systems, NJIT

Dr. Michael Bieber, Dissertation Co-Advisor — Date
Associate Professor of Information Systems, NJIT

Dr. Il Im, Committee Member — Date
Research Professor of Information Systems, NJIT
Assistant Professor, Yonsei University, Korea

Dr. Katia Passerini, Committee Member — Date
Assistant Professor of Management Information Systems, NJIT

Dr. Julian Scher, Committee Member — Date
Associate Professor of Information Systems, NJIT

Dr. Dorothy Leidner, Committee Member — Date
Randall W. and Sandra Ferguson Professor of Information Systems, Baylor University

BIOGRAPHICAL SKETCH

Author: Jia Shen

Degree: Doctor of Philosophy

Date: May 2005

Undergraduate and Graduate Education:

- Doctor of Philosophy in Information Systems,
  New Jersey Institute of Technology, Newark, NJ, 2005
- Master of Science in Information Systems,
  New Jersey Institute of Technology, Newark, NJ, 2001
- Bachelor of Science in Computer Science and Engineering,
  Beijing Polytechnic University, Beijing, P. R. China, 1999

Major: Information Systems

Presentations and Publications:

Shen, J., Hiltz, S. R. and Bieber, M. (2005). "Collaborative Examinations in Asynchronous Learning Networks: Research in Progress." Elements of Quality Online Education: Engaging Communities. Bourne, J. and Moore, J. C. Needham, MA, Sloan Consortium Publications. 6 in Sloan-C Series.

Shen, J. and Patten, K. (2005). "Faculty Perspectives of Collaborative Online Examinations: Initial Report." ISOneWorld Conference.

Shen, J., Cheng, K.-E., Bieber, M. and Hiltz, S. R. (2004). "Traditional in-Class Examination Vs. Collaborative Online Examination in Asynchronous Learning Networks: Field Evaluation Results." Americas Conference on Information Systems 2004, New York City, NY, Best Paper Nomination.

Shen, J. and Jones, Q. (2003). "In Situ Data Capture and Mobile Knowledge Management: Helping Technicians Share Case Stories." Americas Conference on Information Systems, Florida.

Shen, J. (2003). "Utilizing Mobile Devices to Capture Case Stories for Knowledge Management." ACM CHI Doctoral Consortium - Extended Abstract, Florida.

Shen, J. and Jones, Q. (2002). "Pervasive Story Fair: Capturing and Sharing of Stories for Knowledge Sharing." CSCW, Storytelling and Collaborative Activities Workshop Position Paper, New Orleans.

Shen, J. and Shen, X. (2001). "User Requirements in Mobile Systems." Americas Conference on Information Systems, Boston.

Shen, J., Hiltz, S. R., Cheng, K.-E., Cho, Y. and Bieber, M. (2001). "Collaborative Examinations for Asynchronous Learning Networks: Evaluation Results." 34th Hawaii International Conference on Systems Sciences, Maui, Hawaii, IEEE Computer Society Press.

Shen, J., Cheng, K.-E., Cho, Y., Hiltz, S. R. and Bieber, M. (2000). "Evaluation of an on-Line Collaborative Examination Process." Americas Conference on Information Systems, California.

献给

我的爷爷沈正华，为你的激励人心的意志和品格。
我的父亲沈寄云和母亲宗重华，为你们的爱和信任。
我的哥哥沈晓健，因为我爱你。
我的丈夫潘鹏鹏，为你说发现了我身上的金子，并在我们共同的生命旅程中信守此念。

To my grandpa, Zhenghua SHEN,
for your inspirational strength and integrity.

To my parents, Jiyun SHEN & Zhonghua ZONG,
for loving and believing in me.

To my brother, Xiaojian SHEN,
because I love you.

To my husband, Pengpeng PAN,
who claims he has discovered gold in me, and has been sticking to the belief in our life journey.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to Dr. Starr Roxanne Hiltz, my advisor and mentor, for her guidance, encouragement, and moral support throughout this research and my entire graduate study at NJIT. I would also like to express my deep gratitude to Dr. Michael Bieber, my co-advisor, without whom this work is not possible. I sincerely thank each of my Committee members, Dr. Il Im, Dr. Katia Passerini, Dr. Julian Scher, and Dr. Dorothy Leidner, for their valuable insights to this research and encouragements to me. I also thank Dr. Murray Turoff for offering a college graduate the opportunity and financial assistance to pursue her Ph.D. halfway across the globe.

I am very grateful to all the faculty members who have participated in the experiment section of this study: Karen Patten, Richard Egan, Dr. George Olsen, Dr. Il Im, Benjamin Ngugi, Osama Eljabiri, Dr. Robb Klashner, and Dr. Brian Whitworth. I am also grateful to the students for their participation and inputs.

To all the scholars and Ph.D. fellow students whom I have met at NJIT and through conferences over the years, I express my great appreciation. Special thanks to Priscilla Nainby, whose encouragement helped me to go through the difficult times during the journey. I also thank the anonymous reviewers who have reviewed this work.

Finally, I thank my family and friends. This dissertation is not possible without your love, support, and patience. In addition to my immediate family to whom I dedicate this work, I thank my extended family in the U.S., including uncle Ahlong, my husband's sister, and their families. Special thanks to my wonderful parents-in-law, and to my sister-in-law who brings happiness to my brother and to us. I am grateful to my dear friends all over the world. I love you all very much.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter | Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF TABLES
(Continued)

Table **Page**

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

With the proliferation of computer networks and the emergence of virtual teams, learning and knowledge sharing in the online environment has become an increasingly important topic. While Virtual Classrooms (Hiltz, 1994) and ALN (Asynchronous Learning Networks) have become a reality, the traditional assessment methods still dominate when it comes to exams. Instructors take full control of the exam process by designing questions, grading and administering exams while distance-learning students often have to travel to exam centers or take proctored exams.

With the Total Quality Movement and it adoption to education (Deming, 1986; Olson, 1992), assessment has evolved from providing accountability of students' learning to being increasingly regarded as an important part of the education process (Wright, 2003). The traditional instructor-controlled exam reflects the objectivist learning model (Leidner and Jarvenpaa, 1995), where learning is regarded as the uncritical absorption of objective knowledge transferred from the instructor to students. New assessment approaches are proposed based on constructivism theory (Piaget, 1969; Piaget, 1970) where knowledge is constructed actively by learners. For example, learner-centered assessment (Huba and Freed, 1999) and classroom assessment (Angelo and Cross, 1993) are proposed to shift the attention from instructors and teaching to students and learning through assessment.

Collaborative Learning is a learner-centered and team-oriented 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 in both face-to-face settings and ALN using Group Support Systems (GSS) (Hiltz, 1988; Alavi, 1994; Sloffer, Dueber et al., 1999) and in Knowledge Management in distributed teams (e.g. Distributed Cognition, Hutchins, 1995, and Communities of Practice, Wenger, 1998)

With the recognition of the collaborative nature of ALN, a few 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 the grading scheme (Kwok and Ma, 1999), question composition (Wilson, 2004), collaborative question answering (Shindler, 2003) and peer and self-grading (Topping, 1998; Falchikov and Goldfinch, 2000; Sluijsmans, Brand-Gruwel et al., 2003). One study featuring student active participation in various phases of the exam was conducted at NJIT (Shen, Cheng et al., 2004; Shen, Hiltz et al., 2001; Shen, Cheng et al., 2000; Wu, Bieber et al., 2004). The online exam was adopted in a graduate-level course at NJIT for five semesters, where students designed essay-type exam questions, answered questions designed by peers, and graded answers to the questions they authored. The exams were conducted in 3-4 week's period using asynchronous conferencing systems. Student surveys revealed overall favorable attitudes towards the online exam process, including learning effects and high student satisfaction.

Meanwhile, students also reported concerns on issues such as exam efficiency and peer's ability in designing questions and grading.

1.2 Dissertation Overview

This research investigates virtual teams and Computer Supported Cooperative Work (CSCW) through students' collaborative learning in online assessments. Applying constructivism and collaborative learning theories to assessment, the *collaborative online exam* is designed featuring students' active participation in various phases of the exam process through small group activities online. A *participatory online exam* process is designed featuring similar procedures except that students' involvement in each phase of the exam is individual. The collaborative online exam and the participatory online exam are investigated regarding student exam study strategies, group process, exam outcomes, faculty satisfaction, and exam efficiency. Based on the Online Interaction Learning Model (Benbunan-Fich, Hiltz et al., 2004), an input-process-output collaborative examination research framework was proposed. A 1*3 field experiment was conducted to compare three exam modes: the traditional exam, the participatory exam, and the collaborative exam. Quantitative and qualitative data were collected through pre and post exam student surveys, student and instructor exam logs, and interviews.

Through rigorous field experiments, this study shows that the collaborative examination significantly enhanced interaction and promoted higher order learning. In particular, small group activities in the online learning process significantly increased interactions among students which enhanced their sense of an online learning community. Active involvement in the online exam process significantly reduced the use of surface learning in exam study. Overall, students reported significantly higher perceptions of learning in the collaborative exam than the other exam modes. A number of significant relationships were discovered, which are consistent with findings from other studies in ALN.

1.3 Dissertation Outline

This dissertation includes nine chapters. Chapter 1 introduces the dissertation topic and provides an overview of this study. Chapter 2 provides a literature review of the related theories and research in assessment and collaborative learning. Chapter 3 presents research questions, research framework, and hypotheses. Chapter 4 discusses research methodologies including data collection methods, and provides details of the study. Chapter 5 starts the presentation of study results by providing descriptive statistics and index validation. Chapter 6 presents the overall results of quantitative data analysis. Chapter 7 continues quantitative data analysis and provides detailed hypotheses testing results to answer research questions one through six. Chapter 8 provides the log and interview data analysis to answer research questions seven through nine. Chapter 9 concludes the dissertation and provides summaries, discussions, contributions, and limitations of the study.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a literature review of this research. Section 2.1 reviews relevant assessments and learning theories. Section 2.2 reviews collaborative learning and collaborative assessment theories and research which are closely related to this study. The review is based on literatures in education, assessment, ALN (Asynchronous Learning Networks), and GSS (Group Support Systems) published in primarily U.S. books, journals and conference proceedings.

2.1 Assessment and Learning

The review starts with an overview of the assessment movement as the background of this research. Section 2.1.2 discusses different assessment methods categorized based on assessment objectives, format, scoring, and score interpretation. Section 2.1.3 reviews theoretical perspectives related to knowledge, learning, and assessment. Issues concerning validity and reliability of assessment are discussed in Section 2.1.4.

2.1.1 Broader Context

This research is situated within the broader context of assessment movement, which has it roots in the Total Quality Movement (TQM) (Deming, 1986) and its adoption in education: Continuous Quality Improvement (CQI).

2.1.1.1 Assessment Movement. While students have always been assessed in courses, the purpose of assessment has evolved dramatically in different phases of higher education history, from mainly evaluating students' learning, to providing accountability and funding purposes to external audiences, to being regarded as an integral part of learning. A brief review of the assessment movement history helps us to understand the necessity of this research in today's higher education environment (Huba and Freed, 1999).

- Changing resources and the seeds of reform

The post-war period of the 1950s and 1960s was a time of expansion in higher education. Between 1955 and 1970, the number of students pursuing academic degrees tripled and generous support from federal and state governments helped institutions keep place. During this time, the value of a college education was assumed, and universities functioned in a relatively autonomous fashion. There was little need to reveal to external audiences what was happening in college classrooms. However, by the 1970s, higher education was in a grave financial crisis. In addition, the populations of students attending college had become more diverse. Concerns that college graduates did not have the skills and abilities needed in the workplace surfaced. The public and the politicians who represented them began to question the value of higher education. A movement to bring about reform in higher education – and education at all levels – began. As a result, in 1984 and 1985 alone, four reports were issued addressing the need for reform on the college campus: Access to Quality Undergraduate Education (Southern Regional Education Board, 1985), Integrity in the College Curriculum (Association of American Colleges, 1985), Involvement in Learning (National Institute of Education, 1984), and To Reclaim a Legacy (Bennett, 1984). These reports received less attention than *A Nation at Risk* (National Commission on Excellence in Education, 1983), the report that triggered the reform movement in elementary and secondary schools. Their messages were clear and strong: instruction in higher education must become learner-centered, and learners, faculty, and institutions all need feedback in order to improve.

- Calls for accountability

In some states, politicians assumed the responsibility for initiating reform. A number of legislatures have implemented performance funding programs. In performance funding, some portion of the public money earmarked for higher education are allocated to institutions based on institutional ability to meet performance targets like retention rates, graduation rates, or demonstration of student learning. However, the test was not sensitive to institutional differences and needs, but rather was a common instrument for use at all state-funded institutions and was developed by faculty members from across the state.

In part to curtail the direct involvement of state legislatures in higher education, regional accreditation agencies – organizations comprised of institutions of higher education themselves – became involved. Accreditation agencies declared that they would require member institutions to conduct outcome assessment in order to maintain their status as accredited institutions. As time passed, specialized accrediting bodies-those that accredit professional programs rather than institutions (e.g. business, medicine, engineering, architecture) also began to adopt an outcomes approach to program evaluation.

- The Continuous Improvement Movement

In the late 1980s, higher education was influenced by the use of quality principles and practices in business to reduce costs, improve quality of services, and enhance learning. W. E. Deming is recognized as one of the founders of the quality improvement movement. He believed that continuous improvement is the path to improved quality, greater productivity, and reduced cost. Deming's Fourteen Points (1986), the most cited set of principles for continuous improvement, have been reframed for other settings, one of which is education. TQM cultures use feedback from customers, partners and employees to continuously improve products and processes. Some of Deming's fourteen points most relevant to educational assessment include:

- Cease dependence on inspection to achieve quality by building quality into the product in the first place;
- Eliminate numerical quotas and management by objective, substituting leadership instead;
- Promote education and self-improvement; and
- Involve everyone in accomplishing the transformation.

Considerable effort has gone into translating ideas generated by TQM to education (Weaver, 1992). TQM recognizes students as both customers and employees of the educational system (Olson, 1992), calls for changes in teachers' relationships with both students and administrators (Rhodes, 1992), and proposes to assess student progress regularly throughout the school year. Continuous Quality Improvement (CQI), as the application of TQM to education, has the goal of scaling education through attention to learning effectiveness, affordability for learners and providers, and faculty and student satisfaction. In Asynchronous Learning Networks, CQI has been used to design the quality framework of online learning, which is summarized into "Five Pillars of Quality Online Education" (Lorenzo and Moore, 2002), which includes learning effectiveness, cost effectiveness, faculty satisfaction, student satisfaction, and access. In assessment, CQI concepts mean assessment has evolved from providing accountability of students' learning to being increasingly regarded as an important part of the education process (Wright, 2003).

2.1.1.2 Definitions of Assessment. What is assessment? While some people equate assessment with tests, exams, and evaluations, the term assessment is used to reflect the education movement as discussed above. Here are two definitions provided by educational researchers:

> "Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance. When it is embedded effectively within larger institutional systems, assessment can help us focus our collective attention, examine our assumptions, and create a shared academic culture dedicated to assuring and improving the quality of higher education".
>
> (Angelo, 1995)

> "Assessment is defined as the systematic basis for making inferences about the learning and development of students. More specifically, assessment is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students' learning and development".
>
> (Erwin, 1991)

2.1.2 Types of Assessment

The traditional form of assessment in higher-education is instructor-centered, in-class assessment, where instructors provide questions, proctor exams, and grade answers while students answer questions in class. With the advance of educational theory and research, other forms of assessment evolved, such as Alternative Assessment (Rudner and Boston, 1994), Classroom Assessment (Angelo and Cross, 1993), Learner-centered Assessment (Huba and Freed, 1999), and Self and Peer Assessment (Falchikov and Goldfinch, 2000;

Sluijsmans, Brand-Gruwel et al., 2003). Based on the common aspects these assessments share, different types of assessment can be categorized based on:

1. Objectives: For what purposes is the assessment conducted? Does the assessment mainly serve the needs of institutions, instructors, or students?

2. Format: What is being assessed? Do we directly or indirectly assess what we ask students to achieve? Do we mainly assess students' learning or assess the effective of pedagogy and technology intervention?

3. Scoring: How do we judge students' performance? Do we judge the performance based on absolute right or wrong answers, or in some other ways?

4. Score interpreting and reporting: How do we use scores? How do we interpret the assessment results?

2.1.2.1 Objectives. As discussed in the review of the assessment movement above, the purpose or objective of assessment evolved from mainly serving institutional needs in order to provide accountability and funding purposes, to also taking instructors' and learners' needs into account. **Classroom assessment** (Angelo and Cross, 1993) focuses on providing constant feedback to instructors based on student's assessment in class to improve instructor's teaching. **Learner-centered assessment** (Huba and Freed, 1999) refers to the development and use of assessments like projects, papers, performances, portfolios, or exhibitions that evaluate higher-order thinking and require students to directly reveal the very abilities that professors desire. Sometimes these methods are referred to as authentic assessment because of their intrinsic value; at other times, they are referred to as performance assessments because they require students to demonstrate their learning; at still other times, they are termed qualitative assessments because they allow us to evaluate the nature and quality of students' work. Further, scoring in learner-centered assessment is based on subjective judgment using criteria rather than an answer key that allows objective summation of correct answers.

Another important type of assessment that mainly serves the purpose of students' learning is formative assessment. **Formative assessment** is often conducted at the beginning or during a program, thus providing the opportunity for immediate evidence for student learning in a particular course or at a particular point in a program. Classroom assessment is one of the most common formative assessment techniques (Angelo and Cross, 1993). The purpose is to improve quality of student learning and should not be evaluative or involve grading students (Boston, 2002). In contrast, summative assessment is comprehensive in nature, provides accountability and is used to check the level of learning at the end of the program. For example, if upon completion of a program students will have the knowledge to pass an accreditation test, taking the test would be summative in nature since it is based on the cumulative learning experience.

2.1.2.2 Format. The format of assessment ranges from the familiar multiple choice, true false questions, and written essays, to portfolios, capstone, performances, etc. (Wright, 2003). These different methods can be categorized based on whether students' learning is assessed directly or indirectly. In direct assessments, we ask students to demonstrate what they know or can do with their knowledge. Direct assessment may take a variety of forms: projects, products, papers or theses, exhibitions, performances, case studies, clinical evaluations, portfolios, interviews and oral exams. Indirect assessments of learning include self-report measures such as surveys distributed to students, which can be used both in courses and at the program and institutional levels.

Based on current best practice, Steinkuehler and Derry (2003) composed a list of strategies for assessing students' learning. These strategies include:

- Rubric-Scored Written Products

While "grading" is the most common form of assessing students' learning, rubric scoring is more fine-detailed characterization of students' products than simple grading. Rubric scores allow instructors and researchers to statistically analyze students' learning effectiveness. Rubrics typically consist of: (1) a set of categories – features or aspects of student work that are of interest, such as "use of course concept x" or "degree of reflection"; and (2) hierarchical levels of performance within each category, such as "0 – course concept x not used, 1 – course concept x inappropriately used, 2 – course concept x appropriately used but not justified, 3 – course concept x appropriately used and justified." Despite of its effectiveness in assessing students' learning, developing an adequate rubric for a given student product requires time and, often, multiple iterations of revision. Collaborative examinations use rubric scoring to grade answers to essay questions.

- Portfolios

Portfolios are student-prepared collections of documents that evidence understanding of important concepts or mastery of key skills. Portfolio assessments provide each student the opportunity to demonstrate their understanding of course material and, when used longitudinally, how that understanding changes over time. The instructor must communicate his or her expectations to students at the beginning of the course. To alleviate time-consuming nature of portfolio assessment, variations such as showcase portfolio or checklist portfolio can be used. Depending on whether the documents that students produce are direct reflections of their learning, portfolios can be used as either direct or indirect assessment.

- Argumentative Reasoning Assessments

Thinking as argument is implicated in all of the beliefs people hold, the judgments they make, and the conclusions they come to (Kuhn, 1991, p.3). Students' argumentative reasoning can be assessed using students' written products, group written products or, transcripts of group discussions. Argumentative reasoning assessments are only applicable to assess those activities and assignments in which students are, in fact, prompted to display the content and structure of their reasoning. When argumentative reasoning is part of the learning outcomes, argumentative reasoning is a direct assessment of students' learning.

- Mental model analysis

Human beings understand the world by constructing models of it in their minds, therefore, the analysis of students' mental models provides insight into the content and structure of the knowledge individuals taking a course construct. A variety of different data sources can be used as the basis from which to infer individual's mental models. Typically, student written text or transcripts are analyzed for their semantic content and diagrammed as networks. As a variation, *Concept maps* are mental model representations which students build themselves using paper or specialized software. Depending on its use, mental model analysis can be either direct or indirect assessments.

- Collaborative Practices Evaluation

Most closely related to this study, collaborative practices can be evaluated in terms of *quantity* (e.g., the distribution of posts within a given group), *quality* (e.g., application of a scoring rubric designed to assess group process), and *overall structure* (e.g., diagrams that represent the structure of the content of a group discussion). Studies focusing on collaborative practices typically focus on either **individual** as the unit of analysis, where the primary interest is in examining how and what the individual student learns from interaction within the collaborative setting, or **group** as the unit of analysis, where the group is treated as a complex system whose characteristics are of primary interest.

This study mainly adopted the individual level of analysis to study the impact of assessment methods on individual student's learning. Individual students' answers were graded using rubric scoring. Group processes were investigated using student surveys after the exam as well as exam logs and student interviews.

2.1.2.3 Scoring. Scoring of the responses can be automated and done by machine (e.g., multiple choice), by the student's teacher, or by an outside group of trained scorers or evaluators. Some argue machine scoring is the most economical and the most "objective" method (EdSource, 2003). New technologies may be added to the ways that computers can be used to score tests. This could broaden the types of test items that can be scored cost effectively for large-scale assessment. Firms such as Educational Testing Service (Okada, Tarumi et al.) are involved with the development and refinement of statistical and psychometric procedures, and the implementation of new technologies to support automated scoring procedures. Some of these measurement models are already used in a wide range of ETS assessments and products. Measurement models such as Item Response Theory models (IRT) are used in product development, improvement, and evaluation. IRT models support research, test development and statistical analyses in many of ETS's major testing programs such as the GRE®, GMAT®, TOEFL®, and CLEP®. IRT models underlie the development of computer-based tests (CBT).

Critics of objective scoring argue multiple choice and true false test items typically test only factual knowledge. It is possible to write multiple choice and true false items that go beyond checking recall of facts to measure higher-order thinking, and items that do so sometimes appear on standardized tests prepared by professional test developers at companies like ETS. However, when objectively scored items are written by individuals without professional training in test development, they tend to focus on factual knowledge. Another criticism of objectively scored test items is that they assess knowledge bit by bit, item by item, typically with no reference to any eventual real-world application. They are only indirect indicators of more complex abilities such as reasoning about cutting-edge issues or using information to solve important problems in a particular field. Furthermore, objectively scored tests always have a right answer. For these reasons, when we use them, we send students the message that it is important to master isolated facts and skills and to always know the right answers. However, the challenges faced by adults in general and by professionals in particular fields tend to be those that require the simultaneous coordination and integration of many aspects of knowledge and skill in situations with few right answers. As the educator and psychologist Howard Gardner pointed out (Gardner, 1991), the ability to take objectively scored tests successfully is a useless skill as soon as one graduates from college. The rest of one's life, he says, is a series of projects.

In contrast, self-assessment and peer-assessment let students grade their own work or peer's work. Studies have shown that self and peer assessment are reliable, valid, and efficient assessment methods, with significant impact on learner's learning (Topping, 1998; Falchikov and Goldfinch, 2000; McVarish and Solloway, 2002; Sluijsmans, Brand-Gruwel et al., 2003). Given this is one of the central aspects in collaborative assessment, more discussion on this topic will be presented in Section 2.2.3.4 – peer grading.

2.1.2.4 Score Interpretation and Reporting. How well a student performs on a test can be critical, particularly on a standardized test. A student's raw score on a test is just the first step. Next comes an evaluation of what that score means and how it should be interpreted. For a criterion-referenced (or standard-referenced) test, the results are reported based on a set of established expectations and performance standards, such as "basic, proficient, and advanced." For a norm-referenced test, results are reported relative to a comparison group. In the case of many ETS tests and the Stanford-9 portion of the California's STAR program (EdSource, 2003), for example, students are placed in a percentile rank, with the 99th percentile meaning that a student performed better than 99% of the comparison group.

2.1.3 Theoretical Perspectives on Knowledge, Learning, and Assessment

Education pioneers John Dewey (1897) and Howard Gardner (2003) have argued that different types of assessment reflect different underlying assumptions of learning. The existence of different assessment methods indicates different views on fundamental issues such as what is knowledge, how does knowledge transfer from one person to another, and ultimately how can we assess learning. While these three questions are closely related, our discussion below examines each of them in turn.

2.1.3.1 What is Knowledge. Conventional assessments adopt a repository view on knowledge and utilize a passive memory model. From this perspective, learning is the uncritical transfer of knowledge from the instructor to students, and assessment is the recall of correct knowledge from students' memories. However, knowledge is a broad and abstract notion that has defined epistemological debate in western and eastern philosophy since the Greek era (Bridgwater and Kurtz, 1963). While western philosophers have generally agreed that knowledge is *"justified true belief"*, and separated the subject who knows from the object that is known, eastern philosophy emphasizes oneness, e.g. oneness of mind and body, oneness of humanity and nature, and oneness of self and other (Nonaka and Takeuchi, 1995). The data-information-knowledge pyramid is commonly used to differentiate knowledge from other forms of information. Data is perceived as raw numbers and facts, information as data endowed with relevance and purpose, and knowledge as information made actionable, about beliefs and commitment (Drucker, 1998). Although illuminating to some extent in differentiating knowledge from other forms of information, it can be argued that as "one man's knowledge is another man's data" (Stewart, 1997), and this distinction is not truly valid (Alavi and Leider, 1999).

Different types of knowledge exist. For example, Polanyi (1967) distinguishes between tacit and explicit knowledge, which is now widely accepted (Nonaka and Takeuchi, 1995). Erhaut (1994) categorizes professional knowledge into **propositional knowledge, process knowledge, and personal knowledge**. Propositional knowledge is in the form of discipline-based concepts, generalizations and practical principles, and specific propositions about cases, decisions, and actions. Process knowledge is in the form of acquiring information, giving information, and controlling one's behavior. Personal knowledge is concerned with interpretation of experience and understanding of assumptions. While traditional higher education emphasizes the intrinsic value of learning and thus the centrality of propositional knowledge, professional education gives priority to operational outcomes and skill and thus personal and process knowledge.

Focusing on the design of assessment methods using Asynchronous Learning Networks, rather than epistemology, this research takes the view of knowledge as actively constructed. Such a view of knowledge has a long history in cognitive and social science, and recently in organizational and learning research. For example, Bannon and Kuutti (1996) critiqued the notion of memory as passive storage, and argued instead for an active, constructive view of "remembering" that has a long history in psychology and other fields. Bartlett's (1932) experiment of serial reproduction reveals that remembering is a constructive act, which not only requires active participation, but is also influenced by context. This view is further supported by the constructivism and collaborativism theories, as shown in the next section.

2.1.3.2 Knowledge Transfer and Learning Theories. Related to the discussion of knowledge, debates on the origin, creation, and transfer of knowledge have also been going on for hundreds of years. Two basic views exist on knowledge creation: Rationalism and Empiricism. Rationalism perceives knowledge can be obtained deductively from reasoning. Empiricism perceives knowledge can be obtained inductively from sensory experience (Bridgwater and Kurtz, 1963). Related to the transfer of knowledge are the learning theories. Leidner and Jarvenpaa (1995) provided an excellent review of learning theories and their implications in education. They summarized learning models as:

- Objectivist model of learning
- The constructivist model of learning
- The cooperative model of learning
- The cognitive information process model of learning
- The sociocultural model of learning

Each of the learning theories is discussed in turn with implications for assessments.

- Objectivist model of learning

The objectivist model of learning is based on Skinner's stimulus-response theory (Jonassen, 1993). The tenet of the model is that there is an objective reality and that the goal of learning is to understand this reality and modify behavior accordingly. The goal of teaching is to facilitate the transfer of knowledge from expert to the learner. The model assumes that: 1) there exists a reality that is agreed upon by individuals, 2) this reality can be represented and transferred to a learner, 3) the purpose of the mind is to act as a mirror of reality rather than as an interpreter of reality, and 4) all learners use essentially the same processes for representing and understanding the world.

The objectivist model advocates assessment strategies where the instructor is in control to test the recall of knowledge transferred from the instructor to the students. Errors in examinations are the results of imperfect or incomplete knowledge transfer. Drill exercises using true-false and multiple choice questions which require rote memorization reflect the pedagogical assumptions of the objectivist model of learning. Assessments based on the objectivist model may be appropriate in some contexts, e.g. in assessing propositional (factual) and process (procedural) knowledge (Erhaut, 1994).

- The constructivist model of learning

Based upon the work of Piaget (1969; 1970), constructivism argues that rather than being transmitted, knowledge is created, or constructed, by each learner. Early constructivism denies the existence of an external reality independent of each individual's mind. The mind is not a tool for reproducing the external reality, but rather the mind produces its own, unique conception of events (Jonassen, 1993). The constructivist model calls for learner-centered instruction. Class time might become a project-oriented session where the instructor provides tools for helping learners construct their own view of reality. Learning focuses on discovering conceptual relationships, exploring multiple representations or perspectives of issues, or immersing the learner in the real-world context in which learning is relevant (Jonassen, 1993).

In assessment, constructivism advocates learner-centered, non-criterion forms of assessment. The role of the instructor shifts from controlling the whole examination process, to support, guide, and facilitate students' activities. Portfolio assessments are examples of assessment reflecting constructivism, where the students organize, synthesize, and communicate their achievements throughout the semester using documents. Other non-criterion forms of performance assessment such as student learning journals are also examples in this category (Hawkins, May 1993).

Critics of constructivism argue that there is little benefit in having learners construct preordained knowledge; it is only when learners are allowed to construct new meaning, such as in high-order learning, that the goals of constructivism are truly achieved. Indeed, one of the main goals of this research is to examine whether collaborative examinations encourage the use of deep learning approaches to achieve more meaningful learning than conventional examinations.

- The cooperative model of learning

An offspring of the constructivist model is the cooperative, or collaborative, learning model. Whereas in constructivism learning is assumed to occur as an individual interacts with objects, in collaborativism, learning emerges through interaction of individuals with other individuals (Slavin, 1990). In addition, collaboratists also assume that knowledge is created as it is shared, and the more it is shared, the more it is learned. Studies have demonstrated that cooperative learning is superior to individualistic instruction in a wide array of content areas in terms of increases in individual environment, positive changes in social attitudes, and general enhancement of motivation to learn (Flynn, 1992). Learners tend to generate high-level reasoning strategies, a greater diversity of ideas and procedures, more critical thinking, and more creative responses when they are actively learning in cooperative groups than when they are learning individually or competitively (Schlechter, 1990).

The cooperative learning model advocates cooperative assessment strategies. The traditional competitive assessment strategies may disable learning as a learner may be motivated to withhold knowledge to achieve higher grades (Ledlow, 2003). In collaborative or cooperative assessments, the instructor's role is to provide feedback although feedback from the peers is similarly important. Group projects where students are graded based on the quality of group work rather than competing with each other are examples of assessments that reflect the cooperative learning model. While conventional assessments mainly assess factual and procedural knowledge, collaborative and cooperative assessments promote group skills, communication skills, higher-order learning, and long-term retention of knowledge.

- The cognitive information process model of learning

The cognitive information processing model is another extension of the constructivist model and focuses on cognitive processes used in learning. Learning involves processing instructional input to develop, test, and refine mental models in long-term memory until they are effective and reliable enough in problem-solving situations (Shuell, 1986). Major assumptions of the model are that learners differ in terms of their preferred learning style, and attention is limited. Instructional methods that match an individual's learning style will be the most effective. This suggests the need for individualized instruction.

The implication of the cognitive information processing model in assessment is the adaptive or individualized tests that test the cognitive abilities of examinees adaptively. While traditional exams have a fixed length and present the same number of questions to every examinee without considering how well each individual is doing in the exam, a computerized adaptive test discovers the level of difficulty at and below which an examinee can successfully answer exam questions. Examinees at different levels of ability will then see quite different sets of questions: the low-ability examinee will be presented relatively easy questions, the high-ability examinee will be presented relatively difficult questions. Both individuals may answer the same percentage of questions correctly, but because the high-ability person can answer more difficult questions correctly, he or she will receive a higher score. The exam ends when either the accuracy of the examinee ability estimate reaches a statistically acceptable level or when the maximum number of items has been presented. Given the limited attention of examinees, adaptive exams are more likely to test the examinee's ability more efficiently.

- The sociocultural model of learning

Whereas the cooperative learning model and the cognitive information process model are extensions of constructivism, the sociocultural model is both an extension and a reaction against some assumptions of constructivism. In particular, socioculturalists disagree with Piaget's view that the goal of learning is the formation of abstract concepts to represent reality. Rather, knowledge cannot be divorced from the historical and cultural background of the learner (O' Loughlin, 1992). The more meaningful, the more deeply processed, the more situated in context, and the more rooted in cultural background, meta-cognition, and personal knowledge an event is, the more readily it is learned. When applied to assessment, the sociocultural model of learning implies assessment methods should situate in everyday social and cultural context, and test the examinee's ability to critique and change the society.

Table 2.1 summarizes the learning theories (Leidner and Jarvenpaa, 1995) with assessment strategies advocated by them:

Table 2.1 Learning Theories and Preferred Assessment Strategies

Model	Basic premise	Preferred Assessment Strategies
Objectivism	Learning is the uncritical absorption of objective knowledge	Instructor in control of the assessment to test recall of knowledge
Constructivism	Learning is a process of constructing knowledge by an individual	Learner-centered, non-criterion assessments (e.g., student learning journal)
Collaborativism	Learning emerges through shared understanding of more that one learner	Cooperative assessment that encourage idea sharing (e.g. group project)
Cognitive Processing	Learning is the processing and transfer of new knowledge into long-term memory	Assessment adapts to each examinee's knowledge level
Socioculturism	Learning is subjective and individualistic	Assessment situates in the everyday cultural and social context

2.1.3.3 What Should be Assessed and How. While diverse views exist on knowledge and learning, various assessment theories exist in terms of "what should be assessed" and "how can we assess learning". Wright (2003) indicated four main theories exist on the issue of "what should be assessed": Benjamin Bloom's taxonomy of cognitive processes (1956), the distinction between "surface" and "deep" learning (Biggs, 1994; Entwistle, 2000; Bradford, 2004); Perry's Scheme of Intellectual Development (1970); and Lauren Resnick's list of "higher order learning" (1987). Each is discussed in detail below.

Bloom's taxonomy of educational objectives was developed initially as a common frame of reference to address the difficulties of cooperating and communicating about work on educational evaluation. A group of psychologists approached this task by identifying the desired outcomes of education, and specified in operational terms the actions, feelings, and thoughts students are expected to develop as a result of instructional process. The result is a threefold division of educational objectives that not only gives direction to determine the nature of the evidence to be used in appraising the effects of learning experiences, but also gives direction to the understanding of the learning process by providing an order among the outcomes. The three domains are: Cognitive, Affective, and Psychomotor, which are related to the thinking, feeling and acting in learning.

- Cognitive objectives emphasize remembering or reproducing something and solving of some intellective tasks to demonstrate abilities such as comprehension, analysis, synthesis, and evaluation. Two classes of objectives are included in this domain: the "knowledge" objective which involves recall of specifics and universals, methods and processes, pattern, structure, and other knowledge; the "intellectual abilities and skills" objectives which involves organizing and reorganizing material to achieve a particular purpose. The principle of complexity was used to order objectives in the cognitive domain. The largest proportion of educational objectives falls into this domain.

- Affective objectives emphasize a feeling tone, an emotion, or a degree of acceptance or rejections. Affective objectives vary from simple attention to selected phenomena to complex but internally consistent qualities of character and conscience. A large number of such objectives are expressed as interests, attitudes, appreciations, values, and emotional sets or biases.

- Psychomotor objectives emphasize some muscular or motor skills, which are frequently found in physical education, trade and technical courses. This dimension does not apply to this research because the study is not in the above areas. Thus this domain will not be discussed further.

The evaluation of the outcomes of learning involves different techniques to appraise thinking, feeling, and acting. In this study, Bloom's taxonomy was used in the development of the student questionnaires to measure learning approach (using the affective learning dimension), and perceived learning (using the cognitive learning dimension). In addition, the taxonomy was incorporated in the design of question design guidelines and grading guidelines.

In addition the Bloom's taxonomy, this study also investigated exam learning using students' *approach to learning*. Studies have shown the main advantages of collaborative learning and peer assessment include critical thinking, deeper approach to learning, motivation, and team-skills (Schlechter, 1990; Flynn, 1992). In contemporary educational theory, one influential group of researchers has identified students' approaches to be either surface level or deep level (Marton and Saljo, 1976; Biggs, 1987a; Entwistle, 2000; Gordon and Debus, 2002). Approaches to learning are a direct description of the learning process used by students. The categories of learning approaches were developed from interviews and observations of students performing normal learning tasks such as reading articles. Marton and Saljo (1976) identified two discrete approaches to reading articles. Students who used a *surface approach* concentrated on surface features of the learning task, such as key words or phrases. Their strategy was to memorize and reproduce elements that seemed appropriate. Students adopting a *deep approach* concentrated on the underlying meaning of an article. The intention was to understand the real message of a piece of writing or the underlying purpose of an academic task. A deep learning approach is consistent with a search for knowledge and understanding, whereas a surface learner is concerned only with passing exams by memorizing facts.

Learning approaches have a motivation and a strategy element, which are related (Biggs, 1987a; Biggs, 1987b; Entwistle, 2000). Students attempt to *understand* a topic if it is of real interest to them or if they can see its relevance to their current or future professional roles. On the other hand a surface approach is associated with limited interest in a task or an extrinsic motivation. Learning approaches are not stable psychological traits. It is true that students normally have a predisposition to either deep or surface approaches in general. However, this preferred approach can be modified by the teaching context or learning environment for individual course or particular learning tasks. Measures of approaches to learning can, therefore, be related to aspects of the teaching and learning environment. In the AAHE's (American Association for Higher Education) conference on assessment in 2000 and 1999, researchers pointed out the importance of developing assessment methods to promote deep learning (Suskie, 2000). Suggestions were proposed by researchers (Entwistle, 2000) based on the findings of years of research to create curricula, pedagogies, and assessments that promote deep learning.

While the original research which characterized deep and surface approaches used qualitative methods, questionnaires have since been developed (details in Section 4.6 construct measurement). Richardson (1994) published an extensive review of this literature, and he concluded that both qualitative and quantitative research procedures have produced evidence from a reasonable variety of national systems of higher education for the broad distinction between two fundamental approaches to studying: first, an orientation towards comprehending the meaning of the materials to be learned; and second, an orientation towards merely being able to reproduce those materials for the purposes of academic assessment.

This study investigated students' adoption of deep vs. surface learning approaches, as well as individual vs. collaborative learning strategies, when faced with different assessments. Details are discussed in Chapter 3.

In terms of how learning effectiveness can be assessed, in addition to the different types of assessments we discussed in Section 1.2, Steinkuehler and Derry (2003) also offered the following strategies for investigating effectiveness of pedagogy or technology interventions:

- Attitudinal survey

Attitudinal surveys can provide valuable information on students' perceptions of the course, the discipline, the materials and tools used, and their individual progress. This strategy is highly useful for identifying which elements of the pedagogy and/or technology students' feel are most easily mastered, most comfortable, and/or most personally rewarding. Students' responses should be *anonymous*. Data collected through such surveys can be both valid and reliable (Hinton, 1993), but the development of adequate instruments is a nontrivial task. One limitation of this method is that data generated through such surveys is *self-reported* rather than direct observation. Variations to this method include pretest/posttest comparison, epistemological survey etc.

- Pretest/posttest comparison

Pretest-posttest comparisons, when used in a true experimental design, allow relatively straightforward assessment of a pedagogical or technological intervention by detecting differences in learning outcomes between two points in time – before and after. In true experimental design, researchers randomly assign students to one or two groups (e.g. an experiment group and a control group). Students are given a pretest, then the treatment groups are given the treatment, and finally all students are given a posttest.

- Structured interviews

They can provide valuable insight into students' understanding (or misunderstanding) of the course content, document students' course-related knowledge and skills either at specific points in time (e.g., before and after instruction) or longitudinally, or provide information on students' perceptions of the pedagogical or technological methods used.

In this dissertation study, all three methods discussed above were used to investigate the effect of online collaborative examinations. The study design is presented in Chapter 4.

In summary, theories and practices abound in examining knowledge, learning, and assessment. Adapting the four dimensions of learning theories (Leidner and Jarvenpaa, 1995), the four dimensions of assessment are proposed, each addressing one of the following aspects of assessment:

- Who is served: assessment is conducted mainly to serve institutional, instructor or learners' interests
- Realism of context: low vs. high realism of context
- Knowledge: abstract or personally experienced
- What is assessed: levels of skills being assessed.

Based on the diagram developed by Leidner and Jarvenpaa (1995), the different assessment methods discussed in the previous section are populated onto the four-dimensional assessment space as shown in Figure 2.1.

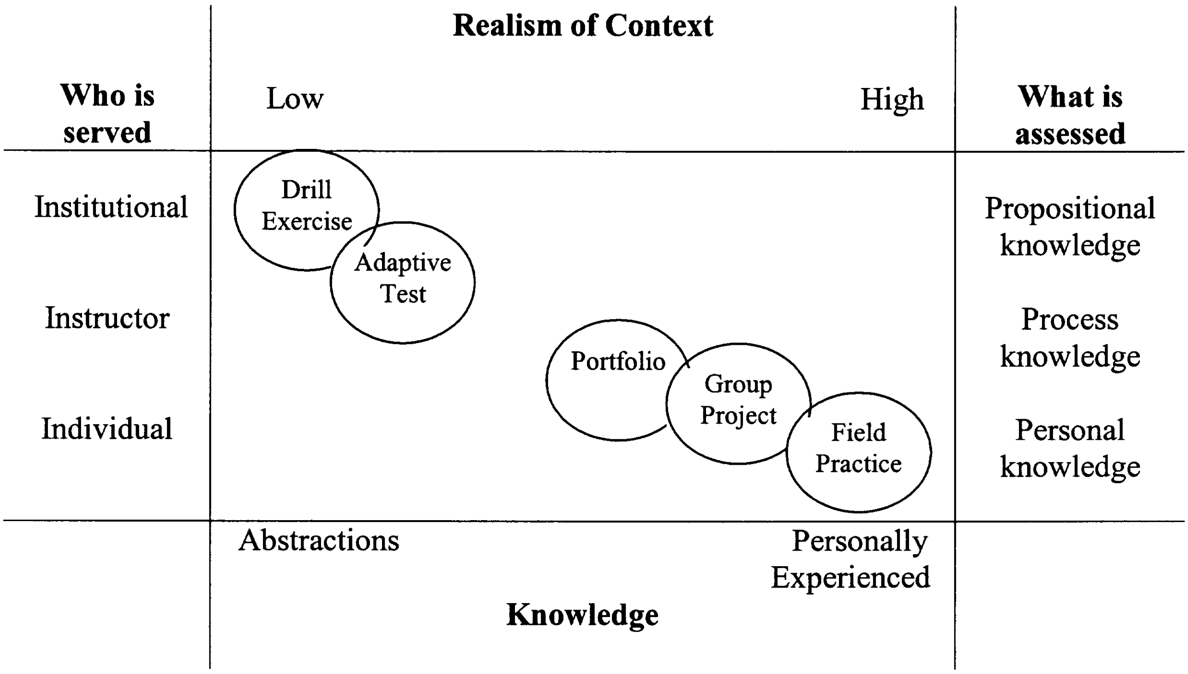


Figure 2.1 Four dimensions of assessment methods.

2.1.4 Soundness of Assessment

Regardless of different types of assessment, one main purpose for any assessment is to provide information on how well students master what they are supposed to learn. But how do we know whether the assessment is designed properly to provide us with this information? When tests are used for high-stakes decision about students and schools, testing experts stress that the testing instrument should be of high quality and validated for the intended purpose (EdSource). Traditionally, the determination of a test's quality revolves around three questions:

- Is the test valid? This is the overarching concern, and it involves asking whether a test provides accurate information for the purposes for which it is being used. If a test is used to determine how well students master standards, does it do a good job of covering those standards?
- Is the test fair? Is it free of built-in biases that create advantages or disadvantages based on individual student characteristics such as racial background? Have students had an opportunity to learn what is being tested?

- Is the test reliable? If a student took the test on two different occasions, would that student's scores remain fairly consistent both times? And do similar students yield the same results time after time?

The soundness of assessment framework (Shindler and Yang, April, 2000) provides four dimensions to evaluate assessments. Within this framework, any assessment practice can be considered sound to the degree that it possesses **validity, reliability, efficiency, and has a positive effect on its users**. Validity is defined by the degree to which a method measures the most important concepts, matches the content covered, and is the best-suited form of methodology to capture the desired learning. Reliability could be characterized by the degree to which a method can obtain an accurate representation of the learning, both among raters (or hypothetical rates) and across multiple performances. Efficiency deals with how "doable" an assessment method is, and how well it can be performed without either taking time away from other teaching and/or other learning. The area related to the effect on the learner could also be considered what has been termed "consequential validity", which includes the motivational, psychological and epistemological affects the assessment has on any learner and/or the class as a whole. The working definition of assessment soundness include:

- Validity:
  - Assessment measures what it intends to measure
  - Assessment measures the most relevant learning from course/assignment content
  - Assessment method is well matched to the assessment target

- Reliability:
  - Assessment device could be used reliably by two different individuals
  - Assessment device could be used reliably for repeated trials/performances

- An appropriate sample of performances is collected to represent a true representation of performance/ability
- Performance criteria is described in measurable, specific, concrete, objective outcome terms

- Efficiency:
  - Assessment data can be collected in an efficient, timely, doable manner
  - Assessment does not unnecessarily interfere with teaching or learning tasks

- Influence on Student Affect:
  - Assessment procedure has an overall positive affect on the student-teacher relationship
  - Assessment has an overall positive affect on the student's motivation level
  - Assessment promotes a sense of competence by providing +/- performance feedback
  - Assessment creates a sense of internal locus of control by providing a clear and attainable target and path to attaining it.
  - Assessment creates a greater sense of belonging and cooperation among the members of the class.

Applying his own framework, Shindler (2003) studied two collaborative assessment practices in teacher education courses. He believed that if teacher education programs are to promote the value of collaboration within their candidates, they must teach and model collaborative pedagogy within their programs. His study is a qualitative examination of the soundness of two forms of collaborative assessment within teacher education courses. The forms of assessment being investigated are 1) collaborative or group exams, and 2) a system of collaborative, interactive roundtable presentations. Qualitative data were collected using participant surveys, focus group interviews, and instructor participation observation. Using the dimensions of soundness defined above, the results of the study suggest that these collaborative assessment methods compared favorably on all four dimensions of soundness. Participants experienced a greater degree of critical thinking, motivation to prepare, enjoyment of the assessment process, and relationship with classmates.

While online collaborative examinations are designed to mainly promote learning through assessment, and thus the study has its main focus on students' learning, the validity and efficiency dimensions will also be addressed. Validity will be examined by comparing peer grading with the instructor's grading. Efficiency will be examined using student and instructor log data and interviews. Reliability of peer grading has been examined in other studies, and is not investigated in this research for practical reasons. Details will be discussed in Chapter 3 and 4.

2.1.5 Computer Technologies for Assessment

With the advance of computer technologies, more and more educational activities are facilitated using computers and network technologies. However, compared with the use of IT in other areas of education such as learning (Alavi, 1994) and online or classroom teaching (Hiltz, 1986; Coppola, Hiltz et al., 2001), the adoption of IT in assessment is still at its early stage. Bennett (2003) categorizes the use of computer technology in educational assessment into three generations. First-generation computer-based tests take limited advantage of technology and resemble the paper-pen type of traditional exams. Computers are used to deliver exams in large-scale adaptively. These tests measure traditional skills and use test designs and item formats closely resembling paper-based tests, with the exception that tests are given adaptively. For example, current tests such as GRE®, GMAT®, and SAT® are administered using computers selecting questions based in part on previous response and tailoring individual skill levels. The next-generation electronic tests use new item formats (including multimedia and constructed response), automatic item generation, automatic scoring, and electronic networks to make performance assessment an integral program component. This type of text allows customers to interact with testing companies entirely electronically. In the third generation, testing will assess new skills and serve both individual growth as well as accountability function normally fulfilled by large-scale test. Tests will be integrated with instructions via electronic tools so that performance is sampled repeatedly over time. In addition, tests use complex simulations, including virtual reality, that model real environments and allow more natural interaction with computers. In summary:

- First-generation computer-based tests:
  1. Primarily serve institutional needs.
  2. Measure traditional skills and use test designs and item formats closely resembling paper-based tests, with the exception that tests are given adaptively.
  3. Administered in dedicated test centers as a "one-time" measurement
  4. Take limited advantage of technology. Resemble the paper-pen type of traditional exams.

- Next-generation electronic tests:
  1. Primarily serve institutional needs
  2. Use new item formats (including multimedia and constructed response), automatic item generation, automatic scoring, and electronic networks to make performance assessment an integral program component; measure some new constructs.
  3. Administered in dedicated test centers as a "one-time" measurement
  4. Allow customers to interact with testing companies entirely electronically

- Generation "R":
  1. Serve both institutional and individual purposes
  2. Integrated with instruction via electronic tools so that performance is sampled repeatedly over time; designed according to cognitive principles

3. Use complex simulations, including virtual reality, that model real environments and allow more natural interaction with computers.
4. Administered at a distance
5. Assess new skills.

The above classification, though informative, reflects the view on assessment as mainly serving evaluation needs for individuals and institutions. Learning through assessment is not given consideration. As this study takes the view of assessments as learning as well as evaluation processes, Collaborative Examinations adopt a different type of the computer technologies where collaborative learning is facilitated: Group Support Systems. Section 2.2 focuses on collaborative learning and computer technologies used in supporting collaborative learning.

2.2 Collaborative Assessment and Collaborative Learning

This section provides a more detailed review of previous research on collaborative assessment and its theoretical foundation - collaborative learning. Section 2.2.1 reviews social learning theories related to collaborative learning, including social presence studies in ALN. Section 2.2.2 reviews computer mediated collaborative learning studies, and analyzes the capacity of Group Support Systems (GSS) in supporting collaborative learning activities. Depending on the stage where student collaboration occurs in the exam process, Section 2.2.3 summarizes previous research on collaborative assessment into designing grading criteria, exam questions, answering, and grading. Section 2.2.4 presents previous studies on online examinations conducted at NJIT, which this study is closely related to this research.

2.2.1 Collaborative Learning

While individual learning is based on the theories of constructivist and active learning (Piaget, 1969; Piaget, 1970), collaborative learning is based on the social learning theories. Collaborative learning is evolved from the work of psychologists such as Johnson and Johnson (1996) and Slavin (1990). It involves social (interpersonal) processes by which a small group of students work together (i.e. collaborate and work as a team) to complete an academic problem-solving task designed to promote learning (i.e. get actively involved and participate in problem solving).

Social learning theories argue that learning is fundamentally social in nature. Such theories are not new. John Dewey (1963) argued strongly for a social view of learning, so did Lev Vygotsky (1962), whose theories underlie much of the current increasing emphasis on the social dimensions of learning in virtually all areas of educational research. Many theories that are distinctively social have been advanced. **Situated Learning** (Suchman, 1987; Brown, Collins et al., 1989; Artman and Warn, 1995; Bardram, 1997) refers to the belief that all learning is situated in the particular physical and social context in which it takes place, and thus recommends pedagogical approaches that embed learning in meaningful activities that make deliberate use of their social and physical contexts. **Distributed Cognition** (Cole, 1991; Hutchins, 1995; Hollan, Hutchins et al., 1999) focuses on learning interactions and cognitive tools and contend that knowledge is distributed across the individual, others, and artifacts. Distributed cognition suggests that our understandings develop not in isolation but through our interaction with other people and the cognitive tools that support interactions. **Learning communities** is another set of theories rooted in the observation that knowledge and learning are a natural part of the life of communities that share values, beliefs, languages, and ways of doing thing (Brown and Duguid, 1991). Etienne Wenger (1998), for example, speaks of learning communities in terms of "Communities of Practice." An important notion of communities of practice is the idea that all learning is situated in practice and that all practice is essentially social in nature.

Applying social theories of learning into the online environment, online educators are concerned about the capacity of online environment to support social activities and interaction. These kinds of questions have typically been explored in what has been called "**Social Presence**" research, and recently "**Virtual Learning Community**" (Swan and Shea, 2004). While Social Presence Theory (Short, Williams et al., 1976) and Media Richness Theory (Daft and Lengel, 1986) argue that low bandwidth media have low social presence, and so cannot convey the social support necessary to sustain learning, ALN researchers contest the view that ALN is lacking richness or social presence (Rice, Hiltz et al., 2004). What is important, they argue, is not media capabilities, but rather **personal perceptions of presence** (Gunawardena and Zittle, 1997; Richardson and Swan, 2003). As the root of the virtual learning community research, research on immediacy in traditional classrooms has focused on teacher immediacy behavior (Weiner and Mehrabian, 1968). Research on social presence in online environment, however, has expanded its scope to the immediacy behavior of all discussion participants. For example, Gunawardena and Zittle (1997) developed a survey to explore student perceptions of social presence in computer-mediated course discussions. The researchers found that course participants created social presence by projecting their identities through the use of affective textual device to build a discourse community among themselves. Richardson and Swan's (2003) research replicated and extended these findings. They found that students' overall perception of social presence was a predictor of their perceived learning in seventeen different online courses. Picciano (2002) reported similar findings.

Indeed, online education is regarded as well constructed to support social learning theories because of the unique nature of asynchronous course discussions, including creating a culture of mindfulness and reflection (Hiltz, 1986), and democratic atmosphere (Harasim, 1990). Recent research suggests the emergence of a virtual learning network (Swan and Shea, 2004). The extent of the emergence of the learning network is dependent on the amount and type of online activities, and perceived presence, which includes teaching, social, and cognitive (Garrison, Anderson et al., 2000; Coppola, Hiltz et al., 2001). Such learning network, in turn, improves student satisfaction and learning in ALNs.

As can be seen from the above, a few key factors can be identified from learning theories and empirical ALN research that are critical to successful online learning, such as active participation, student collaboration, small group interaction, and online discussion. Applied to this study, the collaborative learning theories indicate that the collaborative examination is expected to enhance students' learning through the active participation and collaboration with peers in exam processes. Also, the ALN environment is expected to be able to support such interactions. Details are described in the next section.

2.2.2 Computer Mediated Collaborative Learning

Previous research suggests that collaborative learning is an effective, viable, and perhaps preferred alternative to individual and lecture-oriented teaching methods in higher education. A number of studies using Group Support Systems (GSS) indicate that the effectiveness of collaborative learning in terms of learning outcomes and student affective reactions may be further enhanced through the use of GSS capabilities (Alavi, 1994; Sloffer, Dueber et al., 1999; Vogel, Wagner et al., 1999; Alavi and Dufner, 2004).

Alavi (1994) conducted an experiment with three classes: two with GDSS, one without GDSS, in FTF lectures. Lectures consisted of 40-minute lecture and a group-learning task assigned to groups of four students. Functions of the GDSS tool include: Brainstorming; comment card; compactor; point allocation; ranking; rating; scoring; sub-group selection; and voting. Dependent measures included students' perceptions of their learning and the evaluation of their classroom experience (three factors in learning and two factors in evaluation). She hypothesized that GDSS enhances the effectiveness of collaborative learning by increasing group process gains and decreasing group process losses. More specifically she hypothesized that: 1) GDSS increases effectives of collaborative learning by increasing student participation and active involvement in knowledge construction by facilitating generation, exchange, and analysis of information during the learning group interactions; 2) GDSS increases effectives of collaborative learning by supporting cooperation and teamwork among the students through facilitating information sharing and group process support and process structuring; 3) GDSS increases effectives of collaborative learning by facilitating evaluation and modification of student's mental models through exposure to alternative perspectives and increased and rapid feedback from group members. Results showed GDSS groups have higher scores than groups without GDSS. Participants' midterm and final exam scores were available, and there was no difference in midterm scores. However, GDSS groups obtained higher final grades than groups without GDSS. This study indicated that the effectiveness of collaborative learning in terms of learning outcomes and student affective reactions may be further enhanced through the use of GDSS capabilities.

Leidner and Fuller (1997) examined whether collaborative learning using electronic discussions for case analysis is superior to individual constructive learning involving individual case analysis. The study was conducted in the same time/same place classroom context with the electronic discussions implemented using GroupSystems developed by the University of Arizona. The research employed a quasi-experimental factorial design to examine whether technology-enabled collaborative learning involving case analysis creates higher interests, higher perceived learning, and greater performance than learning in an individual constructive environment for the same task. The study found that: 1) students involved in collaborative learning using GSS showed higher levels of interests in learning than those who worked individually, 2) students involved in individual learning outperformed students in either small (five students per group) or large collaborative groups (40 student per group) which used GSS, 3) students perceived their learning to be higher after they had first interacted in a small discussion than when they worked entirely individually, 4) there were no differences in either interest or performance levels for students in small versus large group discussions. The findings suggest that collaborative learning exceeds individual constructive learning in terms of creating interests. Small group discussions also increased perceived learning, yet students performed better when they work alone.

Vogel et al. (1999) reported the experience of using of an Asynchronous Learning Environment in Hong Kong. GroupSystems was used with the following tools: brainstorming and discussion tools; Categorizing, Prioritizing and voting tools. The Groupware was used in conducting educational activities such as: remote viewing of live lecture video files using RealVideo combined with chat software, desktop videoconferencing with the classroom, including live response, offline dialog with instructors and fellow students through email and conference board, and office watching of past course lectures. Vogel reported favorable results in students' learning and faculty satisfaction in adopting the ALN into classrooms in Hong Kong.

Sloffer et al. (1999) studied the use of asynchronous conferencing to promote critical thinking. He argued that systems should match the tasks they are trying to support. A tool, ACT, was designed to make cognitive processes visible and encourage reflection in students as they engage in critical thinking activities. Features of ACT include a structuring tool, a label tool, and a post-before-read tool to help promote critical thinking skills. The system was used in two courses, a graduate-level seminar and an upper-level undergraduate sociology course. The graduate-level course used the system in a debate task, while the undergraduate-level course used the system for students to analyze cases and problems and to facilitate peer critiquing. In the undergraduate course, students developed an individual position on the problem, provide their rationale, and post a document explaining these in ACT. At least two other students then offered two critiques of each position; the instructor also offered a critique. Finally the students revised their position and defense based on the critiques and submitted a final position paper. The structure tool was used to help students manage information. The post-before-read tool was used to require a student state her position before seeing others. The instructor set up a set of labels, and students were asked to identify labels for their critiques. Examples of labels include grammar/syntax, rethink this, balanced justification, aspects, not a clear position, understanding aspects, new criteria. Students can also use "new criterion" label and petition the instructor to make new labels. Results show Grammar/syntax, which is considered the easiest category for students and least threatening for them to apply, is the most frequently used label. "Rethink This", which is considered the most globally critical category, was used least frequent.

As shown from the above studies, advantages of implementing collaborative learning with GSS include:

- Increased student involvement with the course material and each other
- Promote problem solving and critical thinking skills; higher levels of critical thinking and lower levels of rote memorization
- Promote student learning and academic achievement
- Enhance student satisfaction

The next section reviews research effort in incorporating collaborative learning model into assessment in higher education.

2.2.3 Past Research in Collaborative Assessment

A cursory review of literature shows that student involvement in assessment appears to have been increasing in recent years. This increase appears across the spectrum of discipline areas including science and engineering, arts and humanities, mathematics and education, and social science and business studies, and across a very wide range of student experiences from pre-course to advanced stages and in post-course professional practice. Furthermore, the burgeoning research literature on peer assessment suggests that student involvement is a world-wide phenomenon (Topping, 1998; Sunny San-Ju Lin Liu, 2001; Liu, Liu et al., 2002; Sluijsmans, Brand-Gruwel et al., 2003; Rushton, Sum 1993).

With the recognition of the collaborative nature of ALN, a few studies have been conducted to incorporate student active participation and collaboration into 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 grading schemes (Kwok and Ma, 1999), question composition (Wilson, 2004), collaborative question answering (Shindler, 2003) and peer and self-grading (Topping, 1998; Falchikov and Goldfinch, 2000; Sluijsmans, Brand-Gruwel et al., 2003). One study that features student active participation in all phases of exams was conducted at NJIT (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001; Shen, Cheng et al., 2004; Wu, Bieber et al., 2004). This section reviews these studies.

2.2.3.1 Collaborative Criteria. Kwob and Ma (1999) designed a collaborative assessment task, which involved the development of an evaluation scheme for a semester long project in a 2nd year course of Distributed Information Systems. The collaborative assessment task was carried out in the four stages. First, students were provided with a description of basic set of assessment criteria, which were designed and drafted by the lecturers of the course at the beginning of the semester. They were then asked to discuss these criteria with their lecturers with respect to the course objectives while taking into account their interests and abilities, the learning resources and assessment policy of the institution. As a result, some new criteria were added and some existing ones were modified. Second, Delphi method was used in the experiment where students, together with the lecturers, were required to select six criteria to be used in the formulation of the evaluation scheme through an iterative process of voting, discussion of the outcome and re-voting until they reached a consensus. The number of criteria was six so as to allow students to focus attention on the most important aspect of learning. Third, the students had to assign weights reflecting the relative importance of the selected criteria based on an iterative process similar to the one used in the second stage. Fourth, the collaborative assessment task lasted for the whole semester of 14 weeks. Throughout the semester, the lectures could give students feedback on their learning processes and outcomes according to the evaluation schemes agreed at beginning of the semester. Students could self-assess on their own performance and peer-assess on other students.

A special GSS was designed and implemented to support the group process for the collaborative assessment task. The system provides the following functions: brainstorming, voting and weighting. The system also supported anonymity in order to enable the students voice their opinions freely without fearing a direct confrontation with other students or lecturers.

The study postulates that GSS enhances effectiveness of collaborative assessment by increasing group process gains and reducing group process losses. Two hypotheses are formulated: 1) subjects in a GSS-supported collaborative assessment environment are higher on deep approach to learning than subjects in a collaborative assessment environment that is not supported by GSS; 2) subjects in a GSS-supported collaborative assessment environment achieve a higher level of project grades than subjects in a collaborative assessment environment that is not supported by GSS.

The results support both hypotheses. In this study, online exams were implemented using Webboard and WebCT. While Kwob and Ma's study focused on system, this study focuses on the students' learning. Although Webboard and WebCT are not specialized GSS tools, but general ALN tools, both systems can support brainstorming, voting, and other activities through the discussion forums once proper processes and forums are setup by the instructor.

2.2.3.2 Question Composing. Wilson (2004) conducted a study where students developed questions for midterm and final exams throughout the 15-week semester. Students used ExamNet, an Intranet based software customized for this research. Teams of approximately three students each were assigned to research and contribute six exam questions in each of three weeks during the semester, i.e. each team enters a total of 18 questions during the semester. Individuals could receive a portion of the 10% course participation credit by writing challenges to posted exam questions. Teams were expected to consider and reply to challenges, though changes to exam questions or answers were not required. Students were not required to read questions in ExamNet, but they were promised that at least 50% of exam questions would be drawn directly from the ExamNet database. Research questions include: what types of dialogue arise in ExamNet (helpful vs. confrontational); what motivates students to challenge questions (driven by rewards or altruistic in nature); what do students like least about ExamNet and the collaboration process applied in this research; and what would students like to change about he design of ExamNet (e.g. accountability and anonymity).

Entries of exam questions, challenges, and replies to challenges were logged each time they used ExamNet. On-online questionnaire was conducted prior to the final exam, which asked about students' attitudes and activities. Overall, the findings corroborate prior studies that used ALNs to support student-developed exams. ExamNet's approach to exam question development extends the design reported by Corbitt et al. (1999) by including the ability for students to critique and revise questions. Similar to finding of Corbitt et al. (1999), students liked the interactivity of participating in question development, although in this study they proved to be less keen on the feedback they received from their peers. As found by Shen et al. (2001), students in Wilson's study considered the ALN to be an important part of learning in the course.

In this study, collaborative learning was implemented not only in question design phase, but also the grading phase. Therefore, the question design phase adopts a simpler procedure than Wilson's design.

2.2.3.3 Answering. Shindler (2003) studied two forms of collaborative assessment in teacher education courses: cooperative group exams, and roundtable interactive peer feedback presentation assessment. In the cooperative group exams, students are allowed to form their own groups and work together to develop their response to written exam prompts. Students have done a great deal of cooperative class work in the courses before the exams. Prompts consist of items that require an extensive amount of course content synthesis and application. Two conditions were designed in this process. In one condition each student's exam is evaluated individually; whereas in another condition groups submit only one set of responses as a collective, and each receives the same grade. Data consisted of participant surveys, focus group interviews, and instructor participation observation. Using the dimensions of soundness discussed in the previous chapter, the results of the study suggest that these collaborative assessment methods compared favorably on all four dimensions of soundness. Participants experienced a greater degree of critical thinking, motivation to prepare, enjoyment of the assessment process, and relationship with classmates.

In this study, students answer questions individually in the online exams. Different from the teacher education course, the study is conducted with students. The answering phase is designed as an individual activity to ensure that the exam can assess individual student's mastery of knowledge.

2.2.3.4 Peer Grading. Student involvement in assessment typically takes the form of peer assessment or self assessment. In both of these activities, students are engaged with criteria and standards, and applying them to make judgments. In **self-assessment**, students judge their own work, while in **peer assessment** they judge the work of their peers. An analysis of 62 studies showed that self, peer and co-assessment are effective tools for developing competencies required in professional organizations (Sluijsmans, Dochy et al., 1999; Sluijsmans, Brand-Gruwel et al., 2003).

Falchikov and Goldfinch (2000) conducted a meta-analysis of the **validity** of peer marking by validating students' ratings against those of teachers as a standard (as opposed to the **reliability** of peer assessment which studies the agreement between peer ratings). The study selected 48 quantitative peer assessment studies from more than 100 papers on peer assessment/evaluation, spanning the period 1959 to 1999. The selection criteria were that each study must be situated within higher education and that it must contain correlation coefficients or proportions of cases.

In the meta-analysis, the value of correlation coefficient *r* varied from 0.14 to 0.99. The mean overall value was *r* = 0.69, indicating definite evidence of agreement between peer and teacher marks on average. The study also analyzes the following variables and their relation to the correlation of student/instructor marking:

- Subject area difference (science and engineering vs. social science and arts)
- Advanced level courses vs. introductory courses
- Number of students involved in each peer assessment
- Explicit and student owned criteria vs. other criteria
- Nature of assessment task (traditional academic assessment e.g. essays, tests, presentations vs. professional practice e.g. intern performance, teaching practice, counseling skills)
- Study quality (better vs. poorly designed)
- Number of dimensions used in rating (large vs. small number of dimensions)

The following variables were identified as likely to be influential in terms of improving agreement between faculty and peer assessment, shown below in decreasing order of importance:

- Peer assessments which require marking of several **individual dimensions** appear to be less valid than peer assessment which requires a **global judgment** based on well understood criteria. This is because of the tendency to avoid extreme ratings in each dimensions, which add up to a large error overall.

- Peer assessment of **academic products** and processes seems to correspond more closely to faculty ratings than peer assessment in the context of professional practice.

- **Studies that are well designed** appear to give rise to better peer-teacher agreements than those with poor experimental designs.

- No evidence to support the superiority of **multiple peer ratings** over ratings by **single student**. Ratings by very large numbers of peers (20+) appear to lead to poorer agreement.

- No clear differences in validity of peer assessments in terms of the **subject area** in which they take place, but peers in medically related subjects have a tendency to agree less well in some cases.

- **Student familiarity with, and ownership of, criteria** tends to enhance peer assessment validity.

- Peer assessment carried out on advanced level courses is no more valid than that conducted on introductory courses, in general.

Based on these findings, Falchikov and Goldfinch (2000) provided the following recommendations for implementing peer assessment:

- Avoiding using very large numbers of peers per assessment group

- Conduct peer assessment studies in traditional academic settings and involve students in peer assessment of academic products and processes.

- Do not expect student assessors to rate many individual dimensions. It is better to use an overall global mark with well-understood criteria.

- Involve students in discussions about criteria.

- Pay great attention to the design, implementation and reporting of the study.

- Peer assessment can be successful in any discipline area and at any level.

In this study, students in the online exams grade answers to the questions they composed. The instructor reviews all grades and provides final grades. Students are given the opportunity to contest their grades to the instructor. See Chapter 3 for details.

2.2.4 Previous Studies of NJIT Online Examinations

At NJIT, an online assessment process was designed and studied through student surveys. The exam was termed collaborative exam in previous publications, and is termed the **NJIT online exam** in this document to prevent confusion. The uniqueness of the NJIT online exam, as compared with other forms of collaborative assessments, is the student involvement in each stage of the exam process (except the grading criteria stage). The basic procedures of the NJIT online exam take three to four weeks with the following steps:

- Each student creates exam problems
- The instructor edits the problem 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.

The NJIT online exams were conducted in a graduate-level course for both masters and Ph.D. level students in Information Systems for several semesters starting in 1999. The exam was conducted on EIES™ and WebBoard™, asynchronous conferencing tools used for online learning at NJIT. Student surveys were conducted with results reported in fall 1999, spring 2000, and fall 2000 (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001; Shen, Cheng et al., 2004), and later in spring and summer 2002 (Wu, Bieber et al., 2004). A total of 281 students voluntarily participated in these surveys, resulting in a high total return rate of 93%. Overall positive attitudes towards the new exam process were reported in student surveys. Students reported high perceived learning effects, satisfaction, and recommended the exam for future courses. From answers to the open-ended questions (Shen, Cheng et al., 2000), students reported favorable features of the exam process such as flexibility in time and organizing resources, promoting critical thinking, experiencing less pressure, opportunity to learn from peers, providing a true distance learning experience without having to travel to exam centers for exams, etc. However, students also reported dissatisfaction with issues such as students' ability to design questions and grading, instructor's role in the exam process, question routing, exam efficiency, etc.

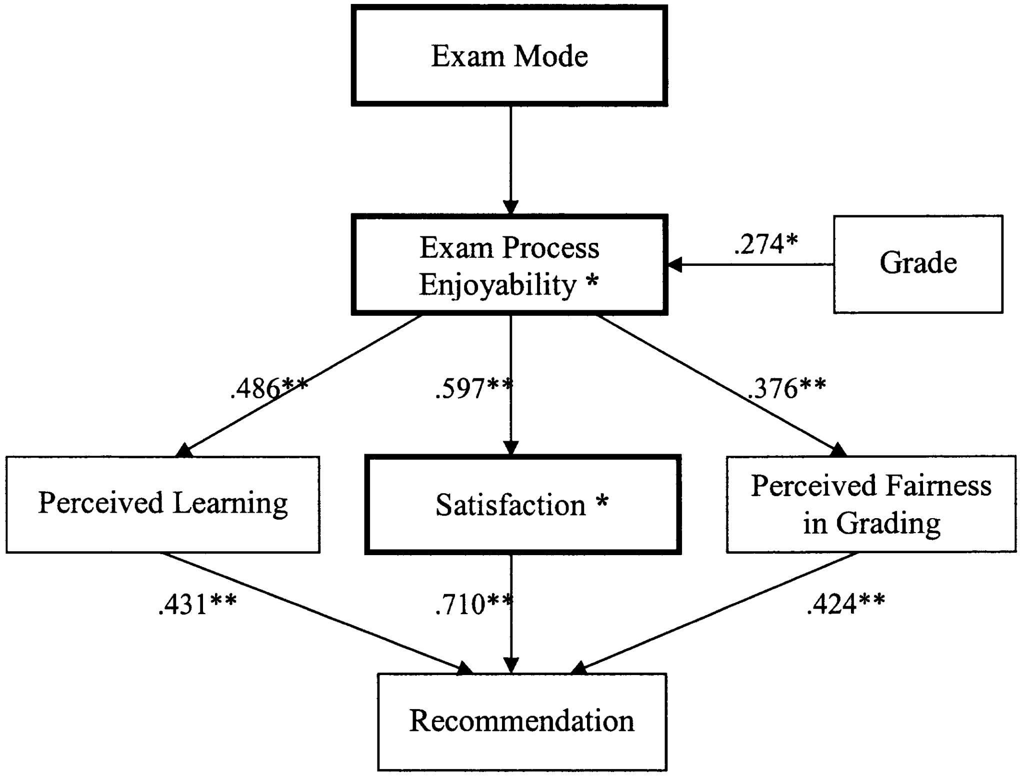
To further investigate the relative merit of the NJIT online exam with the traditional examination, a field experiment was conducted in fall 2000 where half of the total 114 students enrolled in CIS 677 took the online exam and the other half took the traditional exam. A research model was designed based on significant factors found in previous studies (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001; Wu, Bieber et al., 2004). Two questionnaires were designed with common items to allow comparison of students' attitudes in the two exam modes. Results in Table 2.2 show that, compared with traditional exam, students in the 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, $p<.05$). Students in the online exam condition appear to have lower perception of the fairness in grading (mean at 3.62 and 3.40), but this result is not significant. In addition, students in the online exam condition did not perceive a higher level of learning (mean at 3.81 and 3.65).

Table 2.2 Comparison Between Traditional Exam and NJIT Online Exam

Factor	Exam Type	N	Mean	S. D.	Sig.
Enjoyability of the Exam Process	Traditional	41	3.11	.70	*
	Online	52	3.46	.81	
Perceived Learning	Traditional	41	3.81	.71	
	Online	52	3.65	.68	
Perceived Fairness in Grading	Traditional	41	3.62	.79	
	Online	52	3.40	.97	
Overall Satisfaction	Traditional	41	2.78	1.37	*
	Online	52	3.33	1.26	

*: Significant at p<.05 level
Source: Shen, Cheng, Bieber and Hiltz, 2004

Figure 2.2 shows the research model with results of correlations between factors and significant differences indicated in bold (Shen, Cheng, Bieber and Hiltz, 2004a).



*: significant at p<.05 level
**: significant at p<.01 level
Outline box in bold indicates significant differences are found between the two exams

Figure 2.2 Research model in the Fall 2000 study.

More details of the NJIT online examinations can be found in relevant publications (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001; Shen, Cheng et al., 2004; Wu, Bieber et al., 2004).

While generally favorable attitudes towards the NJIT online examination were found in previous NJIT studies, they did not answer what actually happened in students' exam studying process. Did the online exam encourage students to adopt deep learning strategies rather than surface learning? Did the online exam encourage students to learn from peers? These questions are investigated in this study.

CHAPTER 3

RESEARCH FRAMEWORK AND HYPOTHESES

The literature review in previous chapters indicates the importance of collaborative learning in enhancing students' learning, especially in the online environment. This study further investigates the effect of collaborative learning and its impact on student learning, group process, and exam outcomes. Section 3.1 provides research questions. Section 3.2 presents the research framework and defines key variables. Section 3.3 discusses the hypotheses.

3.1 Research Questions

The online exam process described in previous NJIT studies allowed students to be actively engaged in each phase of the exam process. The online exam procedure that was slightly revised from the previous online exam process, which still features students' involvement in designing questions, answering, and grading, is termed **Participatory Exam** in this document. Although previous studies revealed an overall positive attitude, the participatory exam does not directly encourage students to truly interact with peers. Each step in the online exam process, i.e. preparing questions, answering, grading, contesting, is an individual process. Students participate in each step individually without specific requirements for interaction with peers.

Incorporating collaborativism learning theory, a new exam procedure is designed which incorporates small group activities into the online exam process. To differentiate from the participatory exam, the newly designed exam process with small group activities is termed **Collaborative Exam** in this document. Similar to the participatory exam, students also actively participate in each phase of the exam, including designing questions, answering, and grading. The collaborative exam is different from the participatory exam in that small group activities are incorporated into the question design and the grading phases. Students in three to five people a group collaboratively design exam questions, and grade exam answers. The first research question is concerned with comparing the participatory exam with the collaborative exam:

Q1. What are the differences between the collaborative examination and the participatory examination?

This study also investigates the overall merits of the online exam in comparison with the traditional in-class exam. Based on constructivism theories and the work of Piaget (1969; 1970), knowledge is actively constructed by each learner rather than being transmitted (Yarusso, 1992). Used in the education domain, the constructivist model calls for learner-centered instruction (O' Loughlin, 1992) and assessment (Huba and Freed, 1999; McLoughlin and Luca, 2002). Adopting constructivism theories, both the participatory and the collaborative online exams allow students to actively participate in the exam process. The second research question is concerned about comparing the collaborative exam with the traditional exam.

Q2. What are the differences between the collaborative exam and the traditional in-class exam?

Given the collaborative exam is the main innovation of this study, and it has not been conducted before, the research questions are concerned about comparing the collaborative exam with the participatory exam (Q1), and comparing the collaborative exam with the traditional exam (Q2). The participatory exam is also compared with the traditional exam through post-host analysis in the ANOVA tests, and significant results, if any, are reported. See Chapter 6 for details.

As discussed in previous chapters, students dynamically formulate learning strategies in exam study. Students' exam strategies are not only related to the exam type, but also students' predispositions on learning approach (Biggs, 1987a; Biggs, 1987b; Entwistle, 2000). The third research question investigates whether students' predispositions on learning affect their exam study strategies and learning outcomes:

Q3. Is there a relationship between students' predispositions in learning and their perceptions of the exams?

Similarly, students' perceptions of the collaborative exam may also be affected by course characteristics (Boverie, Nagel et al., 1998), such as undergraduate vs. graduate courses, distance learning vs. in-class courses, and classes of different sizes. The fourth research question investigates whether course characteristics affect students learning and exam outcomes in the collaborative exam:

Q4. Do course characteristics affect students' perceptions of the collaborative exam?

One important outcome of the exam is grades. The ability to critically evaluate other's work is essential in people's professional life, and should be fostered in school education. While the validity and reliability of peer grading have been established in meta analysis of peer grading (Topping, 1998; Falchikov and Goldfinch, 2000), students in previous collaborative exam studies were concerned about the fairness of peer grading (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001; Shen, Cheng et al., 2004). The fifth research question addresses grades:

Q5: Is there a relationship between grades and students' perceptions of the exam?

In addition, a number of factors are expected to correlate with each other in predicting the exam outcomes. The sixth research question addresses the relationship among the factors:

Q6: Are there relationships among exam mode, students' strategies in exam studying process, and perceptions of the exam?

With the rapid adoption of wireless and mobile computing technologies in various aspects of our work and life, the educational environment is also increasingly mobilized (Weiser, 1998). Campuses and classrooms are equipped with wireless computing networks such as 802.11b networks, and an increasing number of students are equipped with mobile computing devices such as Notebook computers, mobile phones, PDAs, and tablet PCs. Some researchers have started investigating the new phenomenon of sharing knowledge using mobile devices (Abowd, 1999; Fagrell, Ljungberg et al., 1999). Although the Internet is the main medium through which the online exams are conducted, the use of mobile devices in exam study is also investigated in this research. The seventh research question explores whether the increasing adoption of pervasive computing devices has an impact on students' exam studying process:

Q7. How does the use of mobile computing devices affect students' learning in the exam studying process?

The study also investigates the efficiency of the online exams. As Schindler and Yang (April, 2000) pointed out, efficiency is one important dimension in judging the soundness of assessment. While traditional assessments lack the characteristics advocated by constructivism and collaborativism theories, they seem to take less time to complete. Indeed, "time consuming" is one of the issues students pointed out in previous surveys (Shen, Bieber et al., 2004b). While a traditional exam usually takes a few hours to complete, the online exams take a couple of weeks to be completed through a series of tasks. The last two research questions are concerned about the efficiency of the exams:

Q8. Do students spend more or less time in preparing for, studying, and taking the online exam than the traditional exam?

Q9. Do instructors spend more or less time in preparing for and conducting the online exam than the traditional exam?

3.2 Research Framework and Variables

To investigate the research questions, this study adopts a research framework based on the Online Interaction Learning Model developed by Hiltz, Benbunan-Fich, and Harasim (Benbunan-Fich, Hiltz et al., 2004). The Online Interaction Learning Model is an integrated theoretical framework for learning networks research. The model identifies *online learning interactivity* as a critical factor in determining students' experiences in ALN courses. The model indicates that the extent of students' perceptions of their experiences in an ALN as motivating, actively involving, socially rich, and collaborative mediate the learning outcomes.

Based on the online interaction learning model, this research adopts a three-level input-process-output model. Figure 3.1 shows the research framework.

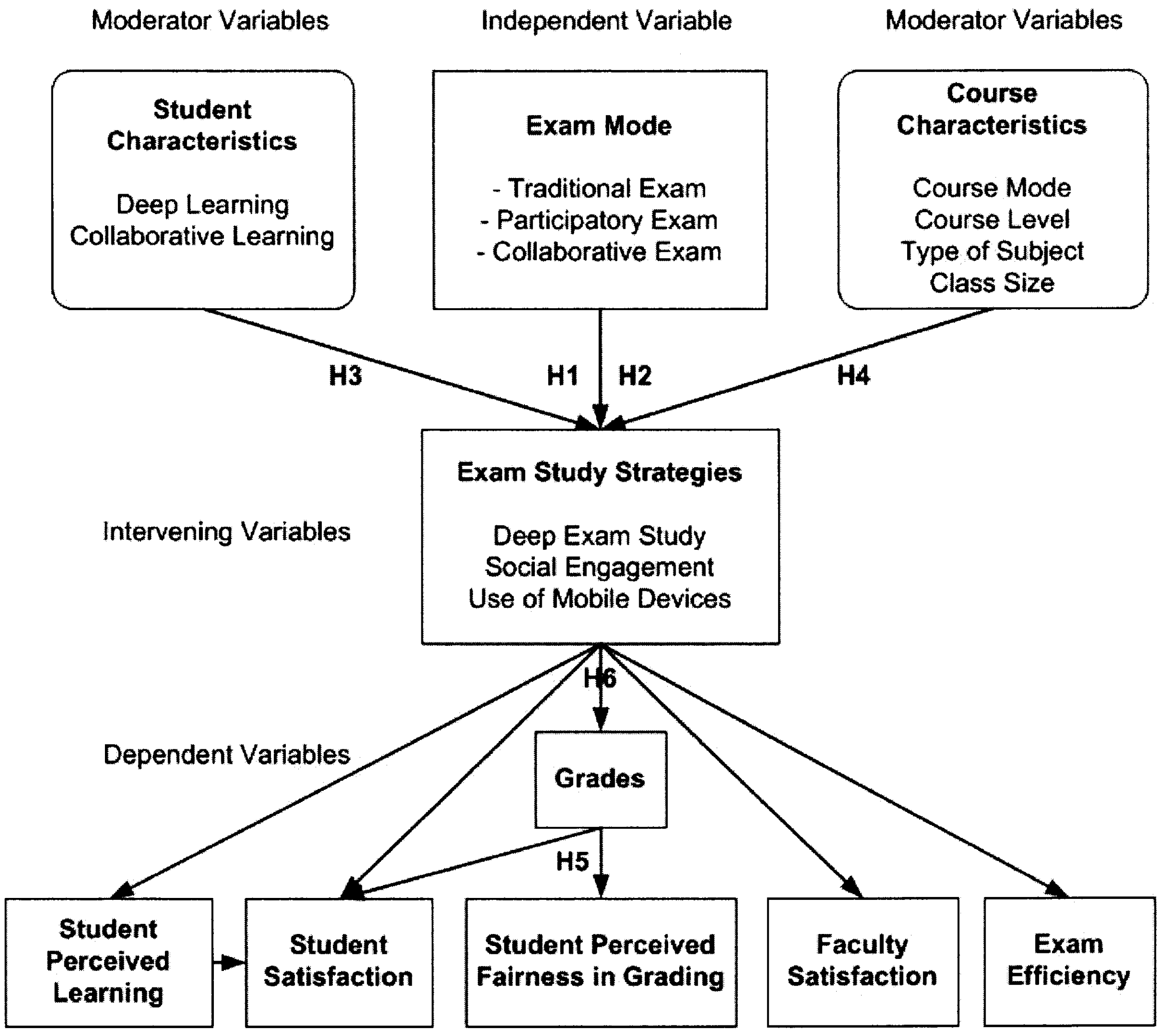


Figure 3.1 Collaborative examinations research framework.

Each variable in the research framework is discussed in detail below.

3.2.1 Independent Variable - Exam Mode

Exam mode is the independent variable of the research framework. Three exam modes are designed: the traditional exam, the participatory exam, and the collaborative exam. The traditional exam is a modified version of the typical in-class, proctored exam. Modifications are incorporated to make the traditional exam a baseline condition for this experiment, i.e., make the three exam processes as similar as possible except for the controlled differences. Participatory and collaborative exams are online exams featuring student's involvement in the question design, answering, and grading phases. Students participate in each phase of the participatory exam individually. Students participate in the question design and grading phases of the collaborative exam in small groups of three to five students a group. Details of each exam mode are discussed in the next chapter.

3.2.2 Moderator Variables – Course & Student

Course characteristics and student characteristics are the moderator variables in the research framework. Many studies in ALN research have indicated that course characteristics (Fjermestad, Hiltz et al., 2004) and student characteristics (Hiltz and Shea, 2004), among other contextual factors (Arbaugh and Benbunan-Fich, 2004), are likely to influence the ALN learning outcome.

As part of student characteristics, students' learning predisposition is expected to moderate the effect of exam modes. In particular, the following two aspects of students' learning predispositions are examined in this study:

- Deep Learning
- Collaborative Learning

Deep learning is the student's approach to learning as directly described based on the learning process used by the student. A deep learning approach is consistent with a search for knowledge and understanding, whereas a surface learner is concerned only with passing exams by memorizing facts (Marton and Saljo, 1976; Biggs, 1987a; Entwistle, 2000; Gordon and Debus, 2002). Previous studies have shown that students normally have a predisposition to either deep or surface learning. In this study, students' deep learning predispositions are measured before the exam as general questions of students' approaches in learning.

Collaborative learning is developed in this study to describe students' orientations in using collaborative learning strategies and learning from other students. A collaborative learning approach is consistent with learning from interacting with other students and enjoying working with other students. Similar to deep learning, students' collaborative learning predispositions are measured before the exam as general questions of their learning approaches.

In this study, the following four aspects of course characteristics are examined:

- Course Mode: DL vs. Blended
- Course Level: Graduate vs. Undergraduate
- Type of subject: Discussion-oriented vs. Technical
- Class Size: Small vs. Large

DL (Distance Learning) are the courses that are offered completely online without face-to-face lectures and meeting times. Blended are the face-to-face courses with significant online components. Since all face-to-face courses that participated in this study had substantial online components, such as weekly discussions, group assignments, etc, "blended" is used in this document instead of face-to-face courses. Based on the

Teaching Goals Inventory and Self-Scorable Worksheet (Angelo and Cross, 1993), the type of course subjects is categorized into technical course, whose primary focus is to teach students technical facts and principles of the subject matter (e.g., programming, coding, technical analysis), or discussion-oriented course, whose primary focus is to help students develop thinking skills (e.g., evaluations of issues, analysis of the impact of technologies).

3.2.3 Intervening Variables - Exam Study Strategies

The intervening variables are the strategies students adopt in the exam study process. Exam study process is where the learning activities occur in preparing for and participating in an exam. In the traditional exam, the exam study process includes preparation before the exam, learning while answering questions during an exam, and reviewing the instructor's feedbacks after the exam. In the online exams, the exam study process includes not only the activities discussed in the traditional exam, but also additional activities such as the question design phase, the grading phase, and reviewing other's postings in the ALN environment.

Three aspects of the exam study strategies are investigated in this study:

- Deep Exam Study
- Social Engagement
- Use of Mobile Devices

While deep learning is associated with students' general learning approach, **deep exam study** is associated with the specific learning strategies students adopt in exam study. Deep exam study is defined in this study as the extent of the student's search for knowledge and understanding in the exam studying process. Low adoption of deep exam study (surface exam study) is associated with passing exams by memorizing facts, while high adoption of deep exam study is associated with seeking knowledge and understanding. Deep exam study is reflected in a number of strategies the student uses in exam studying, such as taking professional perspectives by putting oneself in the position of a professional to understand course materials, adopting academic perspectives to relate course's material to other subject areas, and spending extra time to obtain more information on interesting topics (Tang, 1994, 1991). In this study, students' adoptions of deep exam study are measured after the exam as specific approaches students adopt in exam studying.

Similarly, while collaborative learning is associated with students' general learning orientation, **social engagement** is associated with the specific learning strategies students adopt in exam study. As reviewed in Chapter 2, the term "social presence" has been widely used by online educators who are concerned about the capacity of online environment to support social activities and interaction (Short, Williams et al., 1976; Swan, 2004). While "social presence" has advanced from being regarded as a media capability (Daft and Lengel, 1986) to being considered as personal perceptions of presence of other people in the online environment (Gunawardena and Zittle, 1997; Richardson and Swan, 2003), the term has been used mainly in mediated environments. In this study, social engagement is defined as the extent of student's active involvement in learning from other students through the exam study process, and the sense of forming a learning community through the exam process. High adoption of social engagement in exam studying is reflected in a number of strategies the student uses in exam studying, such as getting to know other students better through the exam process, the perception of enhancing one's understanding of course material by interacting with others in the exam studying process, and acknowledging learning from others. In this study, students' adoptions of social engagement are measured after the exam as specific approaches students adopt in exam studying.

The use of mobile devices is also investigated as an intervening variable in this study.

3.2.4 Dependent Variables - Exam Outcomes

The exam outcomes are the dependent variables in the research framework. The variables are identified corresponding to the Soundness of Assessment framework (Shindler and Yang, April, 2000), which indicates that any assessment practice can be considered sound to the degree that it possesses validity, reliability, efficiency, and has a positive effect on its users. The following aspects of exam outcomes are examined:

- Grades
- Student Perceived Learning
- Student Satisfaction & Perceived Fairness in Grading
- Faculty Satisfaction and Exam Efficiency

The reliability or validity of peer grading have been studied by others (Topping, 1998; Falchikov and Goldfinch, 2000), and is not the main focus of this study. For investigation purposes, assessment validity is explored by comparing instructor's grades with students' grades. Student perceived learning is measured using post exam surveys. Actual learning is measured in some courses, where the experiment is conducted during the midterm and a conventional exam is given as the final exam. Faculty satisfaction and exam efficiency are examined using log data kept by instructors and students. Details are discussed in Chapter 4.

3.3 Hypotheses

Research questions 1-6 are investigated by testing hypotheses sets 1-6. Details of the hypotheses are presented below.

3.3.1 Collaborative vs. Participatory Exam

Corresponding with the first research question, the first set of hypotheses compares the collaborative exam with the participatory exam. Collaborative learning theories suggest that knowledge is created as it is shared, and learning emerges through interaction of an individual with other individuals (Slavin, 1990). The more knowledge is shared, the more it is learned. In particular, it has been demonstrated that small group activities are essential for effective collaborative learning (Harasim, 1990; Bruffee, 1999).

Based on the collaborative learning theories, it is hypothesized that the collaborative exam will encourage higher levels of deep exam study and social engagement in the exam study process (H1.1 and H1.2). Students' perceptions of the exam will be higher in the collaborative exam compared with the participatory exam (H1.3, H1.4, H1.6). Because better learning is hypothesized to occur in the collaborative exam, the grade of the exam is also expected to be higher in the collaborative exam than the participatory exam (H1.5). In terms of question quality, it is hypothesized that students in the collaborative exam will design better quality questions than the participatory exam (H1.7). Question quality is measured by the grade given by the instructor based on question quality. In terms of grading validity, it is hypothesized that students in the collaborative exam mode will be able to provide more valid grades than students in the participatory exam (H1.8). Validity of grade is judged by comparing student's grade with the instructor's grade. Although meta studies on peer grading show that there is no evidence to support the superiority of multiple peer ratings over ratings by a single student (Falchikov and Goldfinch, 2000), the statement refers to multiple students individually grading, not in small groups. It is postulated that small group interaction will allow groups to provide more valid grades than single students.

H1.1: Students taking the collaborative examination will have higher social engagement in the exam studying process than students taking the participatory exam.

H1.2: Students taking the collaborative examination will have higher adoption of deep learning in the exam studying process than students taking the participatory exam.

H1.3: Students taking the collaborative examination will have higher perception of learning than students taking the participatory exam.

H1.4: Students taking the collaborative examination will have higher satisfaction of the exam than students taking the participatory exam.

H1.5: Students taking the collaborative examination will have higher grades than students taking the participatory exam.

H1.6: Students taking the collaborative examination will have higher perception of fairness in grading than students taking the participatory exam.

H1.7: Students' questions will have higher quality (judged by the question quality grade given by the instructor) than those from the participatory exam.

H1.8: Students' grading will have higher correlation with the instructor's grading in the collaborative examination than the participatory exam.

3.3.2 Collaborative vs. Traditional Exam

Corresponding with the second research question, the second set of hypotheses compares the collaborative exam with the traditional in-class exam. It is hypothesized that students will have higher perceptions of the collaborative exam than the traditional exam (H2.1-H2.5), except for the perception of grading fairness (H2.6). In previous collaborative exam studies, especially the third study comparing the traditional exam with the NJIT online exam, students reported the traditional exam grading as more fair than the online exam (Shen, Cheng et al., 2004). Therefore, in H2.6, it is hypothesized that students will perceive grading in the traditional mode as more fair than the collaborative online exam.

H2.1: Students taking the collaborative examination will have higher adoption of deep learning in the exam studying process than students taking the traditional exam.

H2.2: Students taking the collaborative examination will have higher social engagement in the exam studying process than students taking the traditional exam.

H2.3: Students taking the collaborative examination will have higher perception of learning than students taking the traditional exam.

H2.4: Students taking the collaborative examination will have higher satisfaction of the exam than students taking the traditional exam.

H2.5: Students taking the collaborative examination will achieve higher grades than students taking the traditional exam.

H2.6: Students taking the collaborative examination will have lower perception of fairness in grading than students taking the traditional exam.

3.3.3 Student Learning Characteristics

Corresponding with the third research question, the third set of hypotheses addresses students' learning predispositions and the impact on their adoptions of exam study strategies. As discussed in previous chapters, learning approaches have a motivation and a strategy element, which are related (Kember, Charlesworth et al., 1997). Students dynamically form exam study strategies according to different assessment methods, while their predisposition towards a surface or deep learning approach also impacts their exam studying strategy (Tang, 1991; Tang, 1994). Therefore, it is hypothesized that students' predispositions on deep learning will affect their adoption of deep learning in the exam studying (H3.1), and students' predispositions on collaborative learning will affect their level of social engagement in exam studying (H3.2).

The interaction effects of the exam mode and learning predispositions are also hypothesized. Studies have shown that the student's perception of social presence is a predictor of his/her perceived learning in online courses (Swan and Shea, 2004). Students are expected to be more likely to keep the level of deep learning/social engagement in the exam studying process in the collaborative exam, while the likelihood of such will be less in the traditional exam (H3.3 and H3.4).

H3.1: Students' adoption of deep learning in the exam studying process will be positively related to their predispositions in deep learning.

H3.2: Students' level of social engagement in the exam studying process will be positively related to their predispositions in collaborative learning.

H3.3: The correlation between students' adoptions of deep learning in the exam studying process and their predispositions in deep learning will be higher in the collaborative exam than the traditional exam.

H3.4: The correlation between students' level of social engagement in the exam studying process and their predispositions in collaborative learning will be higher in the collaborative exam than the traditional exam.

3.3.4 Course Characteristics

Corresponding with the fourth research question, the fourth set of hypotheses addresses course differences and their impacts on students' perceptions of the collaborative exam, including perceived learning, satisfaction, and perceived fairness in grading. Students pointed out in previous surveys that the NJIT online exam enables "a true distance learning experience" (Shen, Bieber et al., 2004b). It is therefore hypothesized that DL students will have higher perceptions of the collaborative exam than blended students (H4.1). Studies (Boverie, Nagel et al., 1998) have shown that there are significant differences between undergraduate and graduate students in online courses in terms of intrinsic goals, peer learning, and help seeking. Also significant differences are found between DL students who met once every week and students who met only once during the semester in terms of intrinsic goals, control beliefs, elaboration, critical thinking, and help seeking. It is therefore hypothesized that undergraduate students will have lower perceptions of the collaborative exam than graduate students (H4.2). Based on the categorization method discussed earlier, technical courses focus more on teaching students technical facts and principles of the subject matter (e.g., programming, coding, technical analysis), and discussion-oriented courses focus more on helping students develop thinking skills (e.g., evaluations of issues, analysis of the impact of technologies). Considering that collaborative exam allows students to have access to multiple recourses and the freedom to synthesize materials, it is hypothesized that students taking the technical courses will have lower perceptions than students taking the discussion-oriented courses (H4.3). Given the studies on online class size and students' learning and satisfaction, it is hypothesized that classes of smaller sizes will result in higher perceptions of the collaborative exam than classes of large sizes (H4.4). Each of the hypotheses has three sub-hypotheses based on three outcome variables: perceived learning, satisfaction, and perceived fairness in grading.

H4.1: DL students will have higher perceptions of the collaborative exam than blended students.
(H4.1.1: perceived learning; H4.1.2: satisfaction; H4.1.3: perceived fairness in grading)

H4.2: Undergraduate students will have lower perceptions of the collaborative exam than graduate students.
(H4.2.1: perceived learning; H4.2.2: satisfaction; H4.2.3: perceived fairness in grading)

H4.3: Students taking the technical courses will have lower perceptions of the collaborative exam than students taking the discussion-oriented courses.
(H4.3.1: perceived learning; H4.3.2: satisfaction; H4.3.3: perceived fairness in grading)

H4.4: Students in small classes will have higher perceptions of the collaborative exam than students in large classes.
(H4.4.1: perceived learning; H4.4.2: satisfaction; H4.4.3: perceived fairness in grading)

3.3.5 Exam Outcomes

Corresponding with the fifth research question, the fifth set of hypotheses addresses the relationship of grades with the exam outcomes. As important outcomes of the exam, grades are hypothesized to be positively correlated with students' satisfaction (H5.1), and the perceived fairness in grading (H5.2).

H5.1: Students' satisfaction will be positively related to students' grades.

H5.2: Students' perceived fairness in grading will be positively related to students' grades.

Corresponding with the sixth research question, the sixth set of hypotheses addresses the relationship among the exam mode, exam studying process variables (deep learning, social engagement), and exam outcomes. Studies have shown that higher perceived learning leads to higher satisfaction, and sometimes perceived learning and satisfaction merge as one factor (Boverie, Nagel et al., 1998). Therefore it is hypothesized that students' satisfaction with the exam will be positively related to students' perceived learning (H6.1). The interaction effects of exam mode and deep learning/social engagement on satisfaction/perceived learning are hypothesized (H6.2-H6.5).

H6.1: Students' satisfaction with the exam will be positively related to students' perceived learning.

H6.2: The correlation between students' adoption of deep learning in the exam studying process and students' satisfaction with the exam will be higher in the collaborative exam than the traditional exam.

H6.3: The correlation between students' level of social engagement in the exam studying process and students' satisfaction with the exam will be higher in the collaborative exam than the traditional exam.

H6.4: The correlation between students' adoption of deep learning in the exam studying process and students' perceived learning will be higher in the collaborative exam than the traditional exam.

H6.5: The correlation between students' level of social engagement in the exam studying process and students' perceived learning will be higher in the collaborative exam than the traditional exam.

The rest of the research questions (7-9) are explored qualitatively and no specific hypotheses are made.

CHAPTER 4

RESEARCH METHODOLOGIES

This chapter starts with an overview of the research methodology and data collection methods. Section 4.2 provides details on the field experiment design. Section 4.3 and 4.4 describe subjects and tasks. Section 4.5 summarizes research instruments. Section 4.6 and 4.7 discuss details of the online exam procedures, and the systems and tools used to conduct the exams. Section 4.8 provides details of construct measurement. Section 4.9 describes the pilot study.

4.1 Overview

The main research methodology used in this study was field experimentation. Data were collected through qualitative and quantitative methods. While each research question was examined with one main resource of data, multiple methods were used to collect data from various sources to allow triangulation (Creswell, 1994). Triangulation as a data collection method in research is recommended, especially in educational settings where contextual factors are hard to control. Quantitative data were collected through pre and post exam student surveys, and students and faculty exam logs. Qualitative data were collected through semi-structured interviews with students and faculty. Specifically, the following research questions listed in Chapter 3 were mainly tested quantitatively with corresponding hypotheses 1-6, and were cross-examined with exam logs and interview results:

Q1. What are the differences between the collaborative examination and the participatory examination?

Q2. What are the differences between the collaborative exam and the traditional in-class exam?

Q3. Is there a relationship between students' predispositions in learning and their perceptions of the exams?

Q4. Do course characteristics affect students' perceptions of the collaborative exam?

Q5: Is there a relationship between grades and students' perceptions of the exam?

Q6: Are there relationships among exam mode, students' strategies in exam studying process, perceived learning, and satisfaction?

The following questions were mainly explored qualitatively through the semi-structured interviews, and were cross-examined with the exam logs:

Q7. How does the use of mobile computing devices affect students' learning in the exam studying process?

Q8. Do students spend more or less time in preparing for, studying, and taking the online exam than the traditional exam?

Q9. Do instructors spend more or less time in preparing for and conducting the online exam than the traditional exam?

Table 4.1 provides an overview of the variables in the research framework, quantitative and qualitative data collection methods, and the corresponding appendix number. As shown, each aspect of the research framework was examined through multiple data collection methods.

Table 4.1 Overview of Data Collection Methods

Variables		**Data Collection Methods**	**Appendix**
Moderator	Student characteristics	Pre-exam student survey	B.1
	Course characteristics	Pre-exam instructor interview	C.1
Intervening	Deep Exam Study	Post-exam student survey & Student interview	B.3 B.4
	Social Engagement	Post-exam student survey & Student interview	B.3 B.4
	Use of Mobile Devices	Student exam log & Student interview	B.2 B.4
Dependent	Validity	Comparing peer grading with instructor grading	None
	Student perceived learning	Post-exam survey & Student interview	B.3 B.4
	Student satisfaction	Post-exam survey & Student interview	B.3 B.4
	Faculty satisfaction	Instructor interview	C.3
	Exam Efficiency	Student exam log & Instructor exam log	B.2 C.2

4.2 Field Experiment Design

To investigate the research model and test hypotheses discussed in Chapter 3, a field experiment was designed (Table 4.2). The field experiment was a single factor (exam mode) design with three levels, including a traditional exam as the control condition. The experiment was a between group design with pre and post surveys.

Table 4.2 1* 3 Field Experiment Design with Pre and Post Surveys

	Pre-exam Survey	**Post-exam Survey**
Traditional Exam		
Participatory Exam		
Collaborative Exam		

Figure 4.1 compares and contrasts the three exam processes.

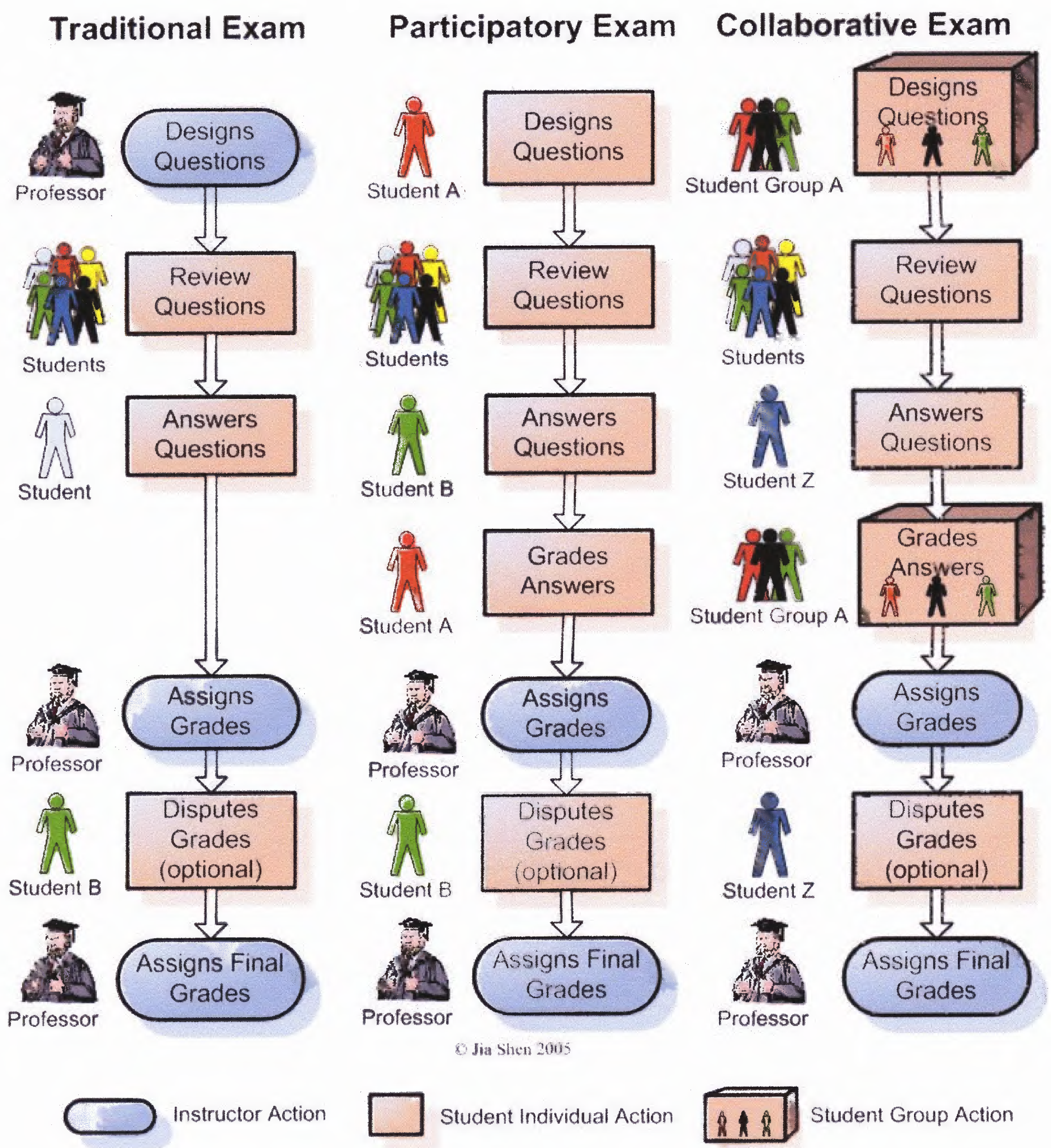


Figure 4.1 Comparison of the three exam modes.

The participatory exam condition and the collaborative exam condition both conduct exams online in the ALN environment. The two conditions share the following process: students design questions, answer questions others designed, and grade the answers to the questions they designed. Optionally, students may contest instructor's grades, and the instructor resolves disputes by assigning final grades. The two conditions differ in that, in the participatory exam condition, students participate in the question design and grading processes individually, and in the collaborative exam condition students participate in the two processes in small groups of three to five students a group. To avoid free-riding in group work (Nunamaker, Dennis et al., 1991), students in the collaborative exam condition first design questions and graded answers individually, and then discuss and collaboratively enhance the qualities of questions and grading. To maintain the purpose of the exam testing individual student's knowledge, the answering phase remains individual in both the participatory and collaborative exam conditions.

Students in the traditional exam condition take the exam in class, or through proctors in remote sites in the case of distance learning students. The instructor provides the questions and grade students' answers, and students answer the questions. In addition, the traditional exam process was modified in this experiment to make it the baseline condition in this experiment. The modifications include the addition of the question review phase and the optional grade contesting phase. First, since students in the online exams can see all the questions posted by other students during the question design phase in the ALN conferences, and possibly review the questions before the answering phase, students in the modified traditional exam are provided with some sample exam questions to review before they take the in-class exam. Instructors are instructed to provide sample questions from previous exams or some guidance to help students in their exam preparation. Second, given the students in the online exam have the option to dispute their grades, students in the modified traditional exam are also provided with the option to dispute their grades if they disagree with the instructor's grading. The instructor resolves the dispute by providing a final grade.

To make sure the three conditions were similar except for the controlled differences, a number of issues were considered in this study. For example, students in the three exam modes were notified of the study at around the same time period during the semesters. Regarding questions, the exam for the same course in different exam modes had equal number of questions whenever possible. For example, if two multiple choice and two essay questions were required in the participatory online exam, then two multiple choice and two essay questions were required in the collaborative online exam for the exam course. The exam for the same course in all conditions also shared similar question format. For example, if essay questions were solicited from students in the collaborative exam mode, then essay questions were used in traditional exam as well. The same course in different exam modes also shared similar grading guidelines, whenever possible. Ideally, exams in all three conditions would have the same access to resources. This means the traditional exam mode should allow students to have access to computers, the Internet, course notes, books and other references. While such equality was not possible for the traditional exams that were conducted in class, efforts were made to make sure students in the traditional mode had adequate access to references. For example, course notes, books, or other references were allowed in traditional exam mode with the instructor's permission.

4.3 Subjects

Courses participating in the study were drawn from the College of Computing Sciences, College of Engineering, and School of Management from NJIT. Mainly upper lever (350 level) undergraduate courses and lower level (600 level) graduate courses participated in the study. Some of these courses had both distance learning (DL) and blended sections using the same syllabi. Usually, the same instructor taught both the DL and FTF sections. Table 4.3 shows the ten courses that participated in this study.

Table 4.3 Courses in the Experiment

Course ID	School	Level	Course
CIS365	College of Computing Science	Undergraduate	File Structure and Management (COBOL Programming)
CIS390	College of Computing Science	Undergraduate	System Analysis and Design
CIS431	College of Computing Science	Undergraduate	Database System Design and Management
CIS433	College of Computing Science	Undergraduate	Electronic Commerce Requirements and Design
CIS490	College of Computing Science	Undergraduate	Software Development Methodologies
CIS663	College of Computing Science	Graduate	Advanced System Analysis and Design
CIS675	College of Computing Science	Graduate	Information Systems Research Methodologies
CIS684	College of Computing Science	Graduate	Business Process Innovation
IE685	College of Engineering	Graduate	System Safety
MIS635	School of Management	Graduate	Management of Business Telecommunications

Given the experiments were conducted in real field settings (i.e., online and traditional courses), total randomization in subject assignment was not feasible. In fact, random assignment of students in every class to one of the three exam modes may seem desirable, but not practical. First, the reactivity effect may confound the results of the experiment. Reactivity effect refers to subjects' awareness of the fact that they are participating in a study, which may create the observed behavior in the experiment (Stevens, 1951; Rosenthal and Rosnow, 1991). If every class had three exams and students were randomly assigned to one of them, students were likely to react by challenging the assignment, requesting to switch to other exam modes, etc. Second, the workload for the instructor to conduct three exams in one course would be overwhelming. It was especially the concern when it was the first time for the instructors to conduct the online collaborative or participatory exams. Most instructors were already having a heavy teaching load after all.

Considering these issues, students participated in this study by the courses they were taking. In all but very few cases, each course section was assigned to only one exam mode. In this way, the workload was much more manageable for the instructor. Furthermore, students were less likely to react to the experiment since all students in the same class were in the same exam condition. However, the disadvantage of this assignment method was the threat of selection effect in confounding the experiment results. Selection effect refers to the pre-existing differences between conditions that may account for the differences in the dependent variables (Stevens, 1951; Rosenthal and Rosnow, 1991).

To reduce the possible pre-existing differences in students, the following measures were taken. First, when the same course or course section participated in the study multiple times (i.e., in summer, spring, and/or fall semesters), efforts were made to assign the course section to a different exam mode than its previous assignment to help counterbalance the differences in different exam modes. Second, in assigning courses to conditions, efforts were made to assign courses with similar characteristics to different exam modes in order to counterbalance the pre-existing differences in students. For example, when there were three graduate-level sections, each of them was randomly assigned to one of the exam modes. Third, in data analysis, pre-survey data collected before the exam on subjects' learning predispositions were used to examine whether there were significant pre-existing differences between courses and sections. See Chapter 5 for more details on this in data analysis.

Twenty-two sections of the ten courses shown in Table 4.3 signed up for the experiment in spring, summer, and fall semesters with a total of 586 students. It is necessary to indicate that all the courses in the summer experiment were the ten-week long regular courses, not the five-week short courses offered in the summer.

Using the balancing technique described above, the assignment of courses to the experiment conditions is shown in Table 4.4. Seven course sections participated in the traditional exam, six in the participatory exam, and eight in the collaborative exam. Superscripted letters are used in the table to denote courses that were used as equivalents in order to counter balance the pre-existing differences. Courses with the same letter were considered inter-changeable in condition assignment. For example, CIS 675-BLD, CIS 684-BLD, and CIS 675-BLD with letter G were interchangeable. (BLD: for courses offered in a blended mode. DL: for totally online distance learning courses).

Table 4.4 Course Sections in Experiment Conditions

		Exam Mode			Total
		Traditional	Participatory	Collaborative	
2004	Spring	[C] MIS635-BLD (26) [A] CIS675-DL (13)	[B] CIS490-BLD (34) [A] CIS675- BLD (26)	[B] CIS490-BLD (23) [C] MIS635-DL (27)	**149**
	Summer		[C] MIS635-BLD (7) [C] MIS635-DL (23)	[D] CIS390-DL (25) [A] CIS675-DL (18)	**73**
	Fall	[D] CIS390-BLD (46) [E] CIS431-BLD (76) [B] CIS490-BLD (48) [F] CIS663-BLD (13) [G] CIS675-BLD (13)	[E] CIS433-BLD (34) [F] MIS635-BLD (12) [G] CIS684-BLD (16)	[E] CIS365-BLD (59) [F] MIS635-DL (19) [G] CIS675- DL (19) IE685-BLD (9)	**364**
Total		**235**	**152**	**199**	**586**

The superscripted letters denote courses of similar nature that were used to counterbalance the pre-existing differences in courses and students.
Number is parentheses ( ) indicates the number of students in each course section.

4.4 Tasks

The tasks of the experiment had two parts: the exam itself, and the tasks related to data collection. As part of the course requirement and grading scheme, the exam was not optional. The exam was considered as instructor's pedagogy in assessment design. Depending on the course, the exam was conducted either as a midterm or a final exam in the spring and summer experiments, and all exams were conducted as the midterm in the fall experiment. See Section 4.6 for more details.

The second part of the experiment tasks were related to data collection. This included the pre and post exam surveys, exam logs, and the post exam interviews. These tasks were set up as optional, and with the instructors' permissions, students were awarded with extra credits if they chose to participate. Alternatives to receive extra credit were provided to students who chose not to participate in the data collection activities.

4.5 Research Instruments

4.5.1 Overview

A number of research instruments were designed for this study. Figure 4.2 shows the research instruments for both instructors and students and their sequence when being used before, during and after the exam.

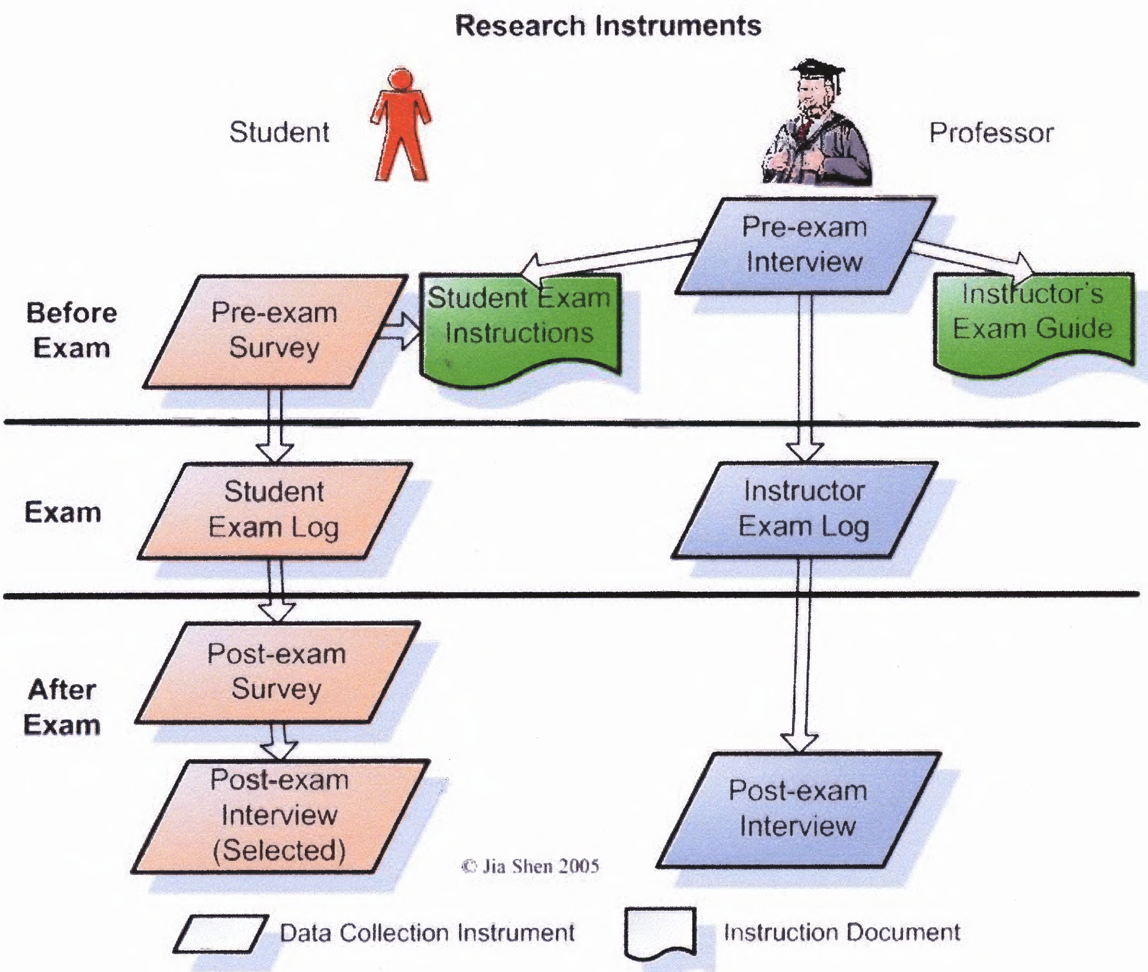


Figure 4.2 Research instruments and timeline.

The student pre-exam survey, exam log, and post-exam surveys were implemented and conducted online. The instruments for students include:

Pre-exam Survey and Consent Form: Before the exam started, students were given a short announcement regarding the study and schedule. They were then given a consent form (Appendix A) which is a standard form used at NJIT to inform subjects of the study procedure, benefits, and risks of the participation in the study. After signing the consent form (by clicking on the "I agree" button in the web browser), students filled out the pre-exam survey (Appendix B.1). The pre-exam survey gathered information regarding the subjects' demographic background, previous experience with online exams, and their learning predispositions.

Student Exam Log: Students were asked to keep the exam log daily during the exam period to record information about their exam study (Appendix B.2). Information collected through the log include type of activity (individual vs. group), and for each type, exam study length, use of materials, and use of medium (Internet, mobile, etc.). Students were asked to start recording their study activities in the log in their exam preparation phase, during the exam period (for the online exams), till the exam was completed after the dispute phase.

Post-exam Survey: The post-exam survey (Appendix B.3) collects information on student exam study strategies and exam outcomes. Exam study strategies include their use of deep exam learning, social engagement, and their use of mobile devices. Exam outcomes include satisfaction, perceived learning, and perceived fairness in grading. The survey also included a few open-ended questions to gather students' comments about the exam.

Post-exam Student Interview: The post-exam student interview (Appendix B.4) further explores students' experiences in the online exams through semi-structured interviews.

Students were asked about their learning strategies, group process, satisfaction, and learning outcomes in detail regarding the online exams they participated.

The instructors who participated in this study were interviewed before the exam started, kept an exam log during the exam process, and were interviewed again after the exam completed. The research instruments for instructors include:

Pre-exam Interview: Before the exam started, the instructors were interviewed (Appendix C.1). The instructor was asked about the nature of course, assignments, and formats of the exams usually conducted in the course. The instructor then worked with the researcher to decide details of the exam for this study, including question design criteria, question format, number of questions each student needed to design and answer, access to resources, etc. The decisions were then incorporated into the student exam instructions as well as the instructor's guide documents.

Instructor Exam Log: Instructors who conducted the online exams were asked to keep a daily log during the exam period to record their activities in conducting the exam (Appendix C.2). Information collected through the log include the type of activity (e.g., communicating with students, grading), and the time spent on each activity.

Post-exam Interview: Instructors who conducted the online exams were interviewed after the exam completed (Appendix C.3). The semi-structured interview explored faculty's experiences in conducting the online exams, including their impressions of the online exam quality, workload, learning outcomes for students and faculty, and exam efficiency.

In addition to the instruments described above to collect data from students and faculty, instruction documents were designed to guide students and faculty through the exam process.

Student Exam Instructions: The student exam instructions (Appendix D) are the detailed guidelines that provide step-by-step instructions for students' participation in this study. Three versions of the instructions were designed and distributed to students before their participation in one of the three exam modes: traditional exam (Appendix D.1), participatory online exam (Appendix D.2), and collaborative online exam (Appendix D.3). The documents were customized for each course incorporating the suggestions made by the instructors in the pre-exam interviews. As part of the exam instructions, detailed question design guidelines and grading guidelines were provided.

Instructor's Exam Guide: The instructor's exam guides (Appendix E) are the detailed guidelines that provide step-by-step instructions for instructor's participation in this study. Three versions of the guides were designed and distributed to instructors before they start conducting one of the exams: traditional exam (Appendix E.1), participatory online exam (Appendix E.2), and collaborative online exam (Appendix E.3). The instructor's exam log was provided as an appendix for easy access. Samples of Excel templates for grading were also provided in the document. The question design guidelines and grading guidelines matching those in the student exam instructions were also included.

Two unique features of the online exams are the question design phase and the answer grading phase. Instruments were designed for each of the phases respectively and are discussed in more detail below.

4.5.2 Question Design Guidelines

To guide question design, six categories of question objectives were created based on the Bloom's taxonomy (Bloom, 1956), and provided as part of the question design guidelines:

1. Knowledge of specifics- knowing major concepts, methods, and theories
2. Comprehension- understanding of major concepts, methods, and theories
3. Application - using theories, methods, and concepts in new context
4. Analysis – analyzing and solving problems
5. Synthesis – relating materials from several areas to make argument
6. Evaluation – judging the value of ideas and assessing quality of arguments

Students were instructed to design questions at the appropriate levels as specified by the instructor in the exam instructions. For example, essay questions may be required at level 3 (application) or up. Students were asked to identify the level of question objectives in the question posting. The categories were set up not only to help students think about the different types of questions that can be designed, but also to help enhance the equality of questions. When assigning questions to students for answering, the instructors were also told to balance the questions on different levels.

In addition to identifying question objectives, students were encouraged to use scenarios from the real world as the question context. Since multiple questions were required, they were also asked to address different aspects of course materials. Lastly, students were asked to ensure that their questions were different from the other ones already posted online. See Appendix D and E for the question design guidelines used in this study.

4.5.3 Grading Guidelines

In designing the grading guidelines, a number of categories were used to guide students to judge the quality of peer's work. Meta analysis of the validity and reliability of peer assessment showed that the most important factor that affects the validity of peer grading is the number of sub dimensions used (Falchikov and Goldfinch, 2000). Peer assessments which require marking of several **individual dimensions** appear to be less valid than peer assessment which requires a **global judgment** based on well understood criteria. This is because of the tendency to avoid extreme ratings in each dimension, which adds up to a large error overall. The grading guideline used in this study has four main categories: correctness and completeness, clarity of writing, following editing guidelines, and submission time and format. Students graded answers by providing a grade for each of the four categories, and the total answer grade was calculated by summing the grades of the sub categories.

In addition to the answer grade, the question quality and the grading quality in the online exams were also evaluated and graded. As shown in studies on collaborative learning and collaborative assessment, students are most motivated when each and every part of their effort is recognized (Bruffee, 1999; Corbitt, Wright et al., 1999). To best motivate students' participation in all the steps in the online exam modes, students were graded not only on their answers, but the quality of their questions, and the grades they provide. In most cases the percentage for the three grades in the online exam modes was 15% for question design (question grade), 70% for answer (answer grade), and 15% for quality of grade (grading grade). To improve exam efficiency, second-level grading was removed in this study in the participatory and collaborative online exams compared with previous studies. See Appendix D and E for the grading guidelines used in this study.

4.6 Detailed Procedures and Issues in Implementations

This section provides details of the exam procedures used in this study. Section 4.6.1 reports the standard exam procedures in detail, and Section 4.6.2 discusses issues in implementing the standard procedures in the field experiment and some deviations in the study.

4.6.1 Detailed Procedures

The traditional exams were conducted in class in one to three hours. The online exams were typically conducted in a 2.5 week timeframe. Table 4.5 summarizes the timeline and procedures of the online exam for students and instructors.

Table 4.5 Detailed Procedures for Online Exams

		Student	Instructor
Exam Week 0 (Before the exam)			Post instructions to students; Set up online conferences; Assign IDs (C: assign groups)
Exam Week 1	**Day 1**	Question Design (P: Individual C: Small Group)	
	2		
	3		
	4	Review questions	Review & revise questions; Assign questions for answering
	5		
	6		
	7		
Exam Week 2	**Day 1**	Answering	
	2		
	3		
	4	Grading (P: Individual C: Small Group)	
	5		
	6		
	7		Assign answer grades
Exam Week 3	**Day 1**		
	2		
	3	Dispute grades to instructor	
	4		Resolve disputes

P: Participatory Exam; **C**: Collaborative Exam. Note: Cells in gray denote no activity

Before the Exam: As shown, before the exam started, the instructor posted announcements and detailed instructions to students. The instructor determined the number and type of questions students needed to design before the exam started. Typical question design requirements were two essay-type questions per student (or six questions per three-student groups in the collaborative exam). In some technical courses where programming skills were assessed, students also designed coding and debugging questions according to the guidelines provided by the instructor.

The instructor then set up the online exam conferences and provided each student with question IDs. Students participated anonymously in the online exams, and all questions and answers were posted in the exam conferences using the question IDs provided by the instructor. Instructors conducting the collaborative exam assigned students to groups of three or four students per group. Existing groups that had worked together on previous assignments or projects (termed assignment groups) were preferred to new groups. While group arrangements are likely to influence outcomes in terms of group process and group outcomes, different theories suggest different arrangements. While constructivism theories based on the work of Piaget (Piaget, 1970) suggest grouping students of *like* ability, collaborative learning theories based on the work of Vygotsky (Vygotsky, 1962) suggest group students of *varying* ability. Assignment groups were used to utilize the cohesion and familiarity group members had already established with each other (Mennecke, Hoffer et al., 1995), which was especially important considering the short length of the exam process. Internal anonymity in small group discussions in the collaborative exam was not required, as most of the group members already knew each other.

Exam Week 1 (Day 1- 3): Students in the online exams were given three days to design questions. Students in the participatory online exam designed questions individually. Students in the collaborative online exam designed questions in groups, where students were instructed to first individually design questions and then discuss and improve them with the group. Questions were designed using the question design guidelines discussed earlier.

Exam Week 1 (Day 4-7): After designing the questions, students were given two to four days to review questions while the instructor reviewed and revised question when necessary to ensure the quality of questions. The instructor then assigned questions to students other than the author of the question, and in the case of the collaborative exam, to students not from the group who authored the question, for answering. Students were not allowed to select questions themselves to reduce cheating.

Exam Week 2 (Day 1-3): Students were given three days to answer the questions assigned by the instructor. The answering phase was an individual process for both the participatory and the collaborative online exams, to maintain the exam's purpose in testing individual student's knowledge.

Exam Week 2 (Day 4-6): After answering the questions, students or student groups who authored the questions graded the answers to their questions. The grading phase was also three-day long. Students in the collaborative online exam graded answers in groups, where students were instructed to first individually graded answers and then discuss and improve them with the group. Answers were graded using the answer grading guidelines discussed above.

Exam Week 2 (Day 7) – Exam Week 3 (Day 1-2): Following students' grading, the instructor reviewed and assigned an answer grade to each answer.

Exam Week 3 (Day 3): Students were given one to two days to dispute the instructor's answer grades if they disagreed with the instructor. Dispute instructions were provided, including providing justifications for the dispute and re-grading of the answer using the answer guidelines.

Exam Week 3 (Day 4): The instructor then resolved the dispute by re-evaluating the question, answer, answer grading, and student dispute, and provided final answer grades.

4.6.2 Issues in Implementations

While the online exams were conducted following the general exam procedures and the timeline as discussed above, there were variations and deviations to the procedures in some courses in the field experiment. The deviations occurred in order to accommodate the specific requirements of each course and the requests of the instructors. First, while all the courses except for one (CIS390) in the spring and summer semesters participated in the experiment as the final exam, all the courses in the fall semester participated in the experiment as the midterm exam. This change was made due to the fact that when the experiments were first conducted in the spring and summer semesters, they were initially set up as final exams. With the experiences gained from the first two semesters, both the instructors and the experimenter agreed that it would be more appropriate to have the online exams during the midterm exam period so that the two-week exam procedures would not coincide with the busy schedule in most courses at the end of the semester. See Section 9.2 for instructor's comments on this issue in more detail.

Second, while some courses devoted a specific period of time for the online exams without having other assignments or projects, some courses had assignments going on in parallel when the online exam was conducted. In a couple of cases, there were assignments or projects due during the online exam process. From the students' and the instructor's interviews, it seems having other assignments during the online exam period, especially having those assignments due during the exam, interfered with the students' exam study and their participation in the online exam. While this issue is hard to control in the field experiment, it is discussed in the limitations Section 9.4.

Third, while both the participatory and the collaborative exams followed the 2.5 weeks, or 18 days, timeline as discussed above, in some courses the actual length varied from 15 days to 24 days. The variations typically were caused by the time the instructor needed to complete their activities. For example, it took some instructors only one or two days to review, revise, and assign the questions to students to answer, and the same task may have taken other instructors three to five days, especially when they had a large class or other personal time constraints. Similarly, the time needed for the instructor to assign answer grades also varied from a couple of days to up to four or five days. In some cases the instructor went away on short trips during the online exam process, and the schedule had to be postponed for a few days to accommodate the instructor's schedule. In contrast, in all but very few cases, the lengths of students' activities were the same in all the courses that participated in the online exams. That is, three days were spent for each of the three main student activities, including designing questions, answering, and grading. The only variation was made by one instructor who felt the deadline was on a weekend (e.g., students' questions due on a Sunday night), and extended it to the following weekday (e.g., Monday night) in response to the requests of her part-time graduate students. In general, there was very little variation of the exam schedule from the students' perspective.

Fourth, the types and the number of questions elicited from students in the online exams also differed in different courses. The specific requirements for each course were discussed with the instructors before the exam and were decided to best fit the course and the experiment needs. In most of the graduate discussion-oriented courses, two essay-type questions were required from each student (e.g., CIS675, MIS635). In undergraduate technical courses, three questions were required from each student (e.g., CIS365, CIS433), including one definition question, one coding question, and one debugging question. The question design and grading guidelines were also slightly modified to address different types of questions.

Fifth, while the majority of the groups in the collaborative exam were previous assignment groups as discussed in Section 4.6.1, there were a number of cases where the students had to be regrouped or formed into new groups for the collaborative exam. This occurred primarily when the assignment groups had too few or too many students to satisfy the ideal three to four students per group requirement in the collaborative exam, or when there were no such groups before the exam in the course. For example, in one course there were a few two-person assignment groups, and these assignment groups had to be combined into four-person groups for the collaborative exam. In other cases, when there were five or more people in an assignment group, students had to be randomly selected out of the group and be formed into new groups. Despite of the effort to form all groups into three to four students per group, in some cases students dropped out of the course and thus left the exam groups with two or even just one person in a group. As a result, the actual group size varied from two to five students per group in the experiment, and in a very few cases, just one student. This issue is discussed in the limitations Section 9.4.

Last, some variations occurred in the traditional exams conducted as the baseline condition in the experiment. While all the traditional exams were conducted as in-class, proctored exams, the length of the exam varied from one hour (e.g., in an undergraduate course) to three hours (e.g., in a graduate course). Second, while the instructors followed the instructions to provide sample questions or guidance to help the students in their preparation for the traditional exam, the types of sample questions and guidance varied in different courses. Some instructors compiled a complete list of the questions designed by students in another section participating in the online exams, and gave the full list as sample questions to the students in the traditional exam, while other instructors selected questions from the list and gave a shortened list to the students. One professor used previous weekly quiz questions as sample questions, and another one provided the overall exam structure and one sample question as the exam guidance. Third, while the instructors were asked to make the traditional exam open sources, the actual instructions varied from allowing students to take all the materials to the exam including the textbooks, lecture slides, articles, and personal notes, to limited materials, such as one article without markup on it. Lastly, while efforts were made to make students in the different sections of the same course (thus different exam conditions) have the same type and number of questions (e.g., if two essay questions in the participatory exam condition, then two essay questions in the traditional exam condition), in some courses the types and the number of questions the students faced were different. This is due to the fact that some instructors felt it was not appropriate, for example, to have just two essay questions for a two-hour in-class exam. In general, however, all the exams in the traditional condition followed the same general procedures in terms of: 1) providing sample questions, 2) having the instructor design or select the questions, administer the exam in-class, and grade answers, and 3) providing students with the optional grade dispute phase. Despite the variations, the traditional exams were adequate in serving the purpose of this experiment.

4.7 Systems and Tools

The two web-based education systems used at NJIT when this study was conducted in 2004 were used in this study: Webboard® and WebCT®. Depending on the instructor's choice, some courses were offered on Webboard and others on WebCT. WebCT is an integrated environment that can support a wide variety of educational activities (such as grade book, quiz tools etc.), and Webboard offers only the discussion forum. The study was conducted on the existing systems the instructors chose for their courses. Both systems have the threaded discussion forum component, which are similar in terms of the interface and functionalities. The online exams were conducted mainly using the discussion forum component on these two systems.

In the discussion forum on Webboard and WebCT, conferences can be created for discussions on separate topics. Webboard and WebCT allow configuration of conferences to control accessibility to either public or private. Public conferences are accessible to all users in the course, while private conferences are accessible to limited number of users who have been granted access to the conference. In addition, conferences can be configured as anonymous, which makes all the messages in the conference shown as anonymous postings.

Three types of exam conferences were created with different configurations and purposes, as shown in Table 4.6. The exam announcements conference was created as a public conference. The purpose was for the instructor to post exam announcements, instructions, and in the case of traditional exam, sample questions for students to review. The exam announcements conference was created in all three exam modes. The online exam main conference was a public conference created for students in the two online exams to post exam questions, answers, and grades. The online exam main conference was configured as anonymous to allow students post messages anonymously with the question IDs they were assigned. The online exam private group conference was the private conference created for each exam group in the collaborative exam. Group members discussed questions and grading during the question design phase and grading phase before posting their results in the online exam main conference.

Table 4.6 Exam Conferences

Exam Conference	Configuration	Purpose	Used in Exam Mode		
			T	P	C
Exam Announcements	Public	Post exam instructions, announcements, sample questions for traditional exam	Yes	Yes	Yes
Online Exam Main Conference	Public & Anonymous	Post exam questions, answers, grading	No	Yes	Yes
Online Exam Private Group Conference	Private	Group discussions on question design and grading	No	No	Yes

T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam

As examples, Figures 4.3 and 4.4 are the screenshots of the conferences on WebCT and Webboard set up for the online exams. Figure 4.3 shows the conferences in WebCT set up for the participatory online exam. Final Exam Announcement is the exam announcements conference, where the instructor can post announcements and answer students' questions, and it is configured as public. Final Exam Conference is the online exam main conference, where students post questions, answers, and grading, and the instructor posts final grades. It is configured as public and anonymous to allow anonymous postings.

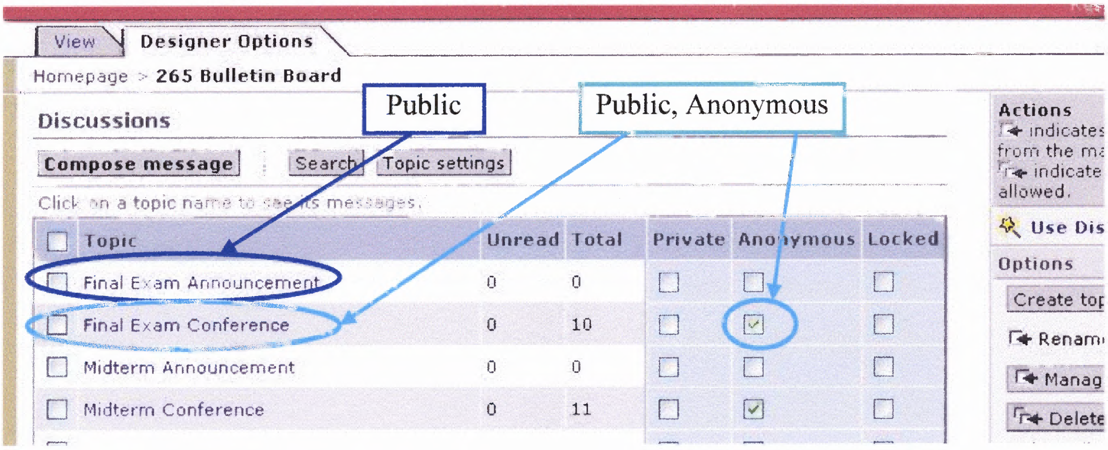


Figure 4.3 Example of WebCT conferences in the participatory exam.

Figure 4.4 shows the conferences on Webboard set up for the collaborative exam. In addition to the Final Exam Announcement and Final Exam Conference similar to those shown in Figure 4.3, each exam group has a private group conference (G1 – G4 in the example) for private group discussions. The instructor gives access to the group conference only to students belonging to the group.

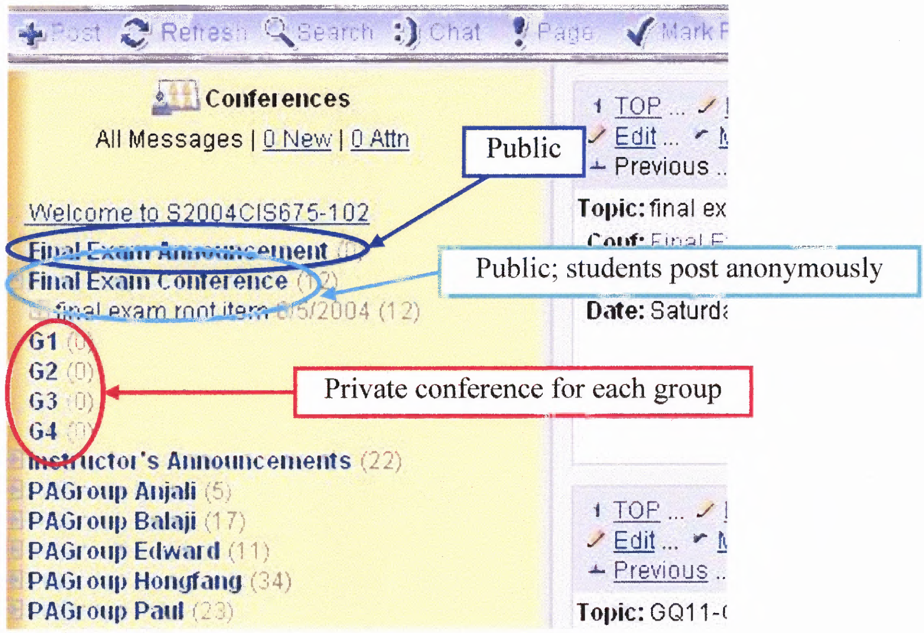


Figure 4.4 Example of Webboard conferences in the collaborative exam.

Students in the online exams were instructed to post questions anonymously using the question IDs assigned to them as the "title" of their message in the online exam main conference. Answers and grades were instructed to be posted as replies to the corresponding questions and answers. Similarly, the instructor was asked to post revised questions and answer grades as replies to the corresponding questions and answers. In the end, each question and the corresponding answer, student grade, instructor grade, dispute, and dispute resolve should appear as one thread of a message.

Figures 4.5 and 4.6 show examples of the online exam conference at the end of the exam process on Webboard and WebCT. Figure 4.5 shows the postings in the exam conference of the participatory exam on WebCT. Figure 4.6 shows the postings in the exam conference of the collaborative exam on Webboard. The two types of online exam share a similar threaded message structure at the end of the exam, which includes a message for each of the following: question, revised question (optional), answer, answer grading, final answer grading, dispute (optional), and resolution (optional).

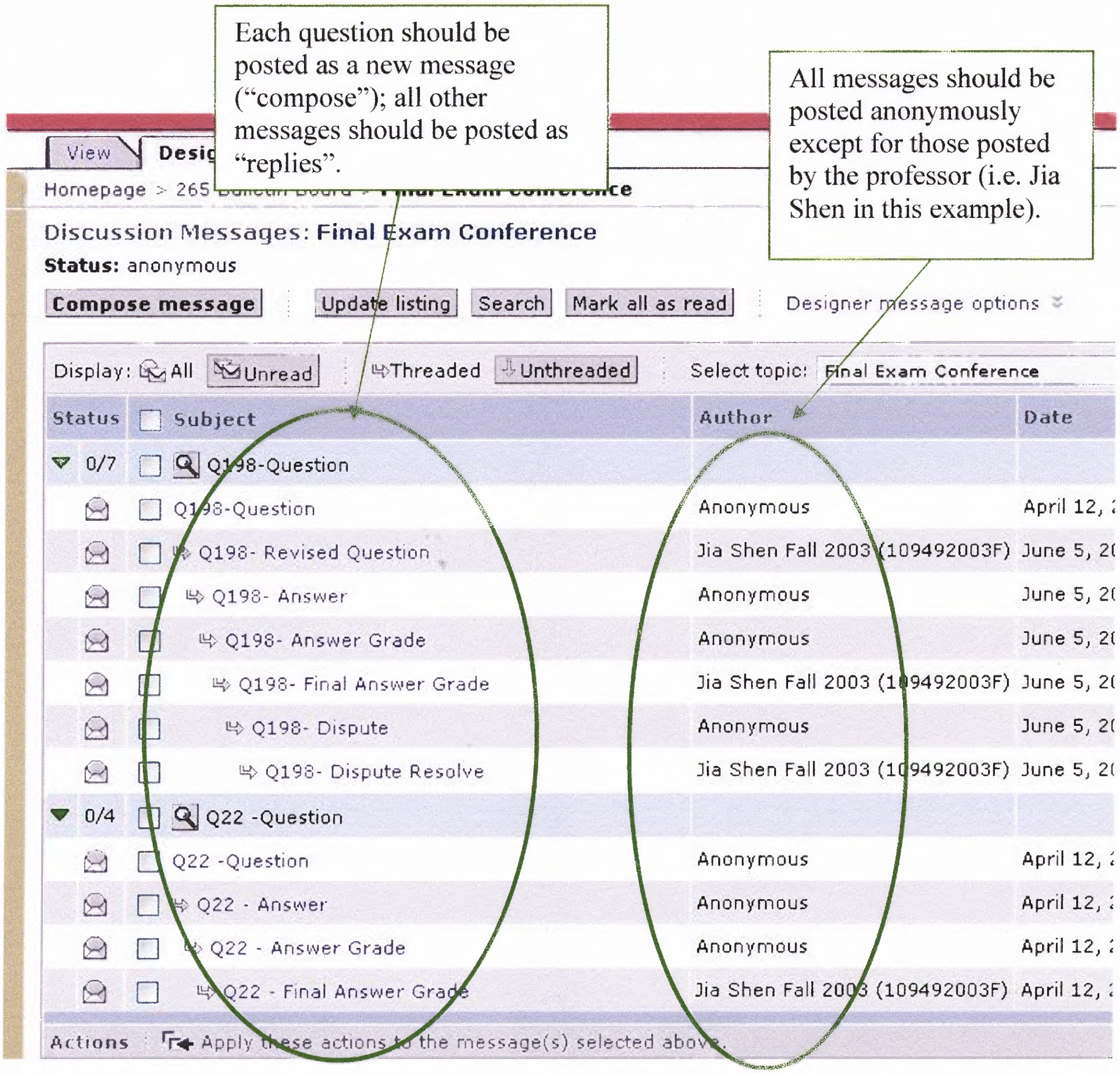


Figure 4.5 Example of messages on WebCT in the participatory exam.

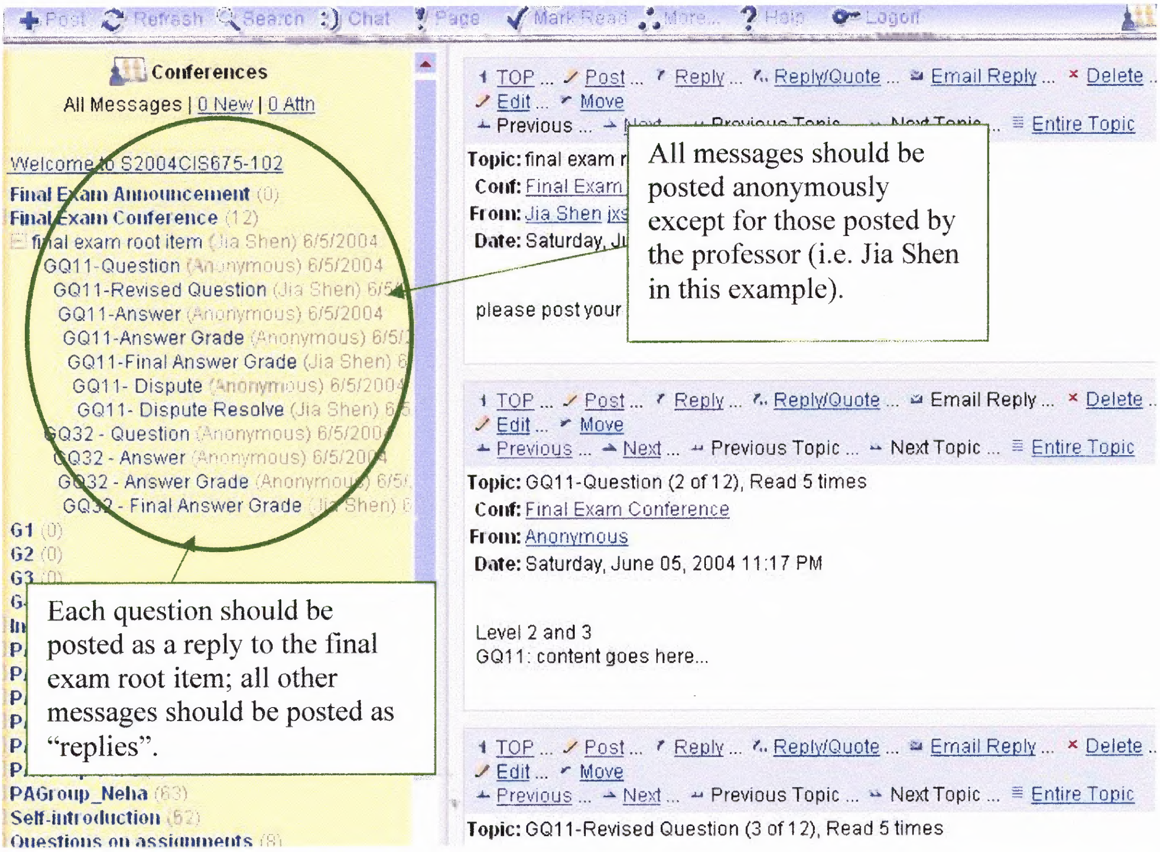


Figure 4.6 Example of messages on Webboard in the collaborative exam.

In addition to the discussion forum on Webboard and WebCT, Turnitin® were used in preventing and detecting plagiarism in the exam. Turnitin offers web-based services (www.turnitin.com) where instructors can set up courses and assignments, and students can submit assignments electronically. The system compares the assignment submission with documents from a wide variety of sources, including students assignments from previous semesters, documents on the web, academic publications etc. A submission index score is given for each submitted report indicating the level of similarity between the submission and all the sources in the Turnitin database. During this study, NJIT had subscription to the basic service, and it was available to all instructors. However, the "peer review" service was not subscribed. This prevented students seeing other students' reports or submissions index results, therefore students were not able to use Turnitin to detect plagiarism in answer grading. To solve this problem, instead of having students check turnitin results, instructors were asked to review the submission results on Turnitin when providing the answer grades. See the instructor's guide in the appendix for more details.

4.8 Construct Measurement

The major constructs in the research model were measured quantitatively using the pre and post exam surveys. Table 4.7 lists the major constructs, sources, levels of adaptation from the source, and number of items.

Table 4.7 Construct Measurement

Level	Construct	Source	Adaptation	No. of Items
Attribute Variables	Deep/Surface Learning Predisposition	R-SPQ-2F (motive subscale)	Minor change	10
	Collaborative Learning Predisposition	VU	Minor change	10
Intervening Variables	Deep/Surface Exam Study	R-SPQ-2F (strategy subscale) & APS	Major change	12
	Social Engagement	VU	Minor change	6 (10)
Dependent Variables	Perceived Learning	Bloom's Taxonomy & CEQ	Major change	12 (4)
	Satisfaction	CEQ	Minor change	6 (4)
	Perceived Fairness in Grading	CEQ	Minor change	3 (4)

Note: the number in parentheses ( ) denotes the number of items for the collaborative and/or participatory exams only.

As shown, the pre and post exam student questionnaires adapted items from the Revised Two Factor Study Process Questionnaire (R-SPQ-2F) developed by John Biggs and David Kember (Biggs, Kember et al., 2001), the Virtual University (VU) course questionnaire developed by Starr Roxanne Hiltz (Hiltz, 1996), the Assessment Preparation Strategies (APS) Questionnaire developed by Catherine Tang (Tang, 1991), and the Collaborative Exam Questionnaire (CEQ) used in earlier NJIT studies developed by Jia Shen, Starr Roxanne Hiltz and Michael Bieber (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001). These measurements are described in more detail below.

4.8.1 Constructs Measured in Pre-exam Questionnaire

The attribute variables were measured through the pre-exam questionnaire, including deep learning predisposition and collaborative learning predisposition.

Deep Learning Predisposition: Ten items were adapted from the Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) developed by John Biggs (Biggs, Kember et al., 2001) to measure this construct. Scales were developed by other researchers to measure deep and surface approaches in learning, such as the Approaches to Studying Inventory (ASI) (Ramsden and Entwistle, 1981; Entwistle and Ramsden, 1983), the Study Process Questionnaire (SPQ) and the Learning Process Questionnaire (LPQ) (Biggs, 1987b), and the R-SPQ-2F (Biggs, Kember et al., 2001). The R-SPQ-2F was selected for this research because of its sensitivity to learning context and education innovations. While student's learning styles developed in cognitive psychology and learning are regarded as an attribute of the student and are not likely to change, the SPQ is based on a Presage- Process- Product (3P) model of teaching and learning, which regards students' learning approaches as continuously changing based on the interactions of student factors (prior knowledge, ability), teaching context (objectives, assessment, etc.), activities, and learning outcomes. The SPQ and The revised SPQ (R-SPQ-2F) questionnaire have been used in studies in before-after modes to assess the impact of innovations (Tang, 1994; Kember, Charlesworth et al., 1997).

The R-SPQ-2F has 20-items to measure deep vs. surface learning approach, each with motive and strategy subscales. The scale has been tested and shows good Cronbach's alpha value for reliability and a good fit to the intended two-factor structuring. In this study, ten items of the motive subscale from the R-SPQ-2F were adapted to measure deep vs. surface learning approach at the "presage" level. Minor changes were made to adapt the items to the examination scenario and the course settings.

Collaborative Learning Predisposition: Ten items were adapted from the individual/collaborative learning subscale in the Virtual University (VU) course questionnaire developed by Starr Roxanne Hiltz (Hiltz, 1996) to measure this construct. Minor changes to the original items were applied to adapt the questions to the examination scenario and different course environments.

4.8.2 Constructs Measured in Post-exam Questionnaire

The intervening and dependent variables were measured through the post-exam questionnaire, including deep/surface exam study, social engagement, perceived learning, satisfaction, and perceived fairness in grading.

Deep Exam Study: 12 items were developed to measure the level of deep learning in exam study. The 12 items were designed based on the strategy subscale of R-SPQ-2F, as described earlier, and the Assessment Preparation Strategies (APS) questionnaire developed by Catherine Tang (Tang, 1991). APS was designed based on SPQ to explore the strategies that the students employed when they were preparing for different assessment methods. The original APS questionnaire consisted of 30 items constituting factors such as learning materials and processing, narrow focusing and cue seeking, relating study material to professional context, relating study material to other subjects, and students' organization of work and study time. Although the APS questionnaire was not publicly accessible, some information was provided in the publications with some example items (Tang, 1994; Tiwari and Tang, 2001). Changes to the items were applied to adapt the questions to the examination scenario and different course environments.

Social Engagement: Six items were adapted from the collaborative learning subscale in the Virtual University (VU) course questionnaire developed by Starr Roxanne Hiltz (Hiltz, 1996) to measure social engagement in exam study. Minor changes to the original items were applied to adapt the questions to the examination scenario and different course environments. Additional ten items were used in the collaborative exam post survey on the group process.

Perceived Learning: 12 items were developed to measure perceived learning as one of the exam outcomes. Among the 12 items, ten were newly designed based on Bloom's taxonomy (Bloom, 1956) to measure perceived learning from lower levels such as understanding the materials to higher levels such as comprehension, understanding, application, analysis, synthesizing, and evaluation. The other two items were adapted from the Collaborative Exam Questionnaire (CEQ) used in NJIT online exams research conducted by Jia Shen, Starr Roxanne Hiltz and Michael Bieber (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001). Additional four items were used in the two online exam questionnaires regarding learning through the online exams.

Satisfaction: Six items were developed to measure satisfaction as one of the exam outcomes. The items were adapted from the Collaborative Exam Questionnaire (CEQ) used in earlier research (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001). The scale was tested with three student surveys with a good Cronbach's alpha value. Additional four items were used in the two online exam questionnaires.

Perceived Fairness in Grading: Three items were developed to measure perceived fairness in grading as one of the exam outcomes. The items were adapted from the Collaborative Exam Questionnaire (CEQ) used in earlier research (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001). The scale was tested in previous three student surveys with a good Cronbach's alpha value. Additional four items were used in the two online exam questionnaires.

4.9 Pilot Study

The spring 2004 study was initially conducted as a pilot study in three courses at NJIT: CIS675, MIS635, and CIS490. A total of 149 students enrolled in these classes participated in the study. A few issues were discovered through the pilot study, and were addressed in the experiments in the summer and fall semesters:

First, in the pilot study, two questions were solicited from each student in the online exams. Using the question objectives, one question was required on level one, and another one on level two and up. Through instructors' feedback and students' comments, the question on level one was considered too simple and not appropriate for the open-book format in the online exams. To resolve this issue, subsequent studies required all questions to be designed on level 2 and up.

Second, the answers in the pilot study had a maximum length requirement, but no minimum length requirement. The issue was brought up by the instructors in the interviews. In subsequent studies, a minimum length requirement was added in addition to the maximum requirement.

Third, students and faculty commented that the student exam instructions were too long. To resolve this issue, a one-page quick overview page was added to help provide an overview of the online exams. Numbers were provided in the overview page as index to the detailed procedures in the document. In addition, screenshots from WebCT/Webboard were added to better illustrate the use of the ALN systems.

Fourth, given the small number of faculty who conducted the online exams in the pilot study (two besides the researcher), the instructors were guided through face-to-face meetings and via email during the exam process. In subsequent semesters, the instructor's guide was designed and provided to the instructors to make sure the exams were conducted following the same procedures. An Excel example was provided in the instructor's log.

Fifth, the research instruments, including the student exam instructions, the pre and post exam surveys, and the exam log, were tested through the pilot study and proved to be effective. Initial factor analysis using the data from pilot study was conducted, and the results suggested the surveys were designed properly except for two questions in the pre-exam questionnaire which did not converge to the factors as designed. These two questions were eliminated in the subsequent studies. One main revision to the exam instruction document was the addition in grading guidelines to deal with plagiarism. While this issue was not specifically included in the grading guidelines in the pilot study, it was added to the grading guideline after it was brought up by the instructors through the pilot study. The use of Turnitin was enforced starting from the summer semester too. See Appendix D and E for details of the grading guidelines and exam procedures.

Most of issues discovered above were minor. There were no major changes to the experiment procedures or research instruments. Therefore the data from the spring semester were included as part of the main study, and were analyzed together with the summer and fall data in the following chapters.

CHAPTER 5

DESCRIPTIVE STATISTICS AND INDEX VALIDATION

This chapter starts the presentation of study results. Section 5.1 provides the descriptive statistics of the courses and the students in the study. Section 5.2 describes the procedure used to validate the indexes using principle component factor analysis. Section 5.3 provides the normalization tests and data transformation results.

5.1 Descriptive Statistics

Table 5.1 shows the total number of students who returned the exam surveys by semester and exam mode. While 586 students participated in the exams (shown in Table 4.4), 485 students participated in the pre exam and/or post exam surveys. Among them, 173 students participated in the traditional exam condition, 137 in the participatory exam condition, and 175 in the collaborative exam condition. The data analysis are conducted using the survey results from the 485 students.

Table 5.1 Number of Subjects Completed the Surveys

		Exam Mode			**Total**
		Traditional	Participatory	Collaborative	
2004	Spring	38	53	47	138
	Summer	0	30	38	68
	Fall	135	54	90	279
	Total	**173**	**137**	**175**	**485**

Table 5.2 shows the number of students who returned the exam surveys in each course by semester and the exam mode. The twenty-two courses participated in the study were evenly distributed among the three conditions, including seven courses in the

traditional exam mode, seven courses in the participatory exam mode, and eight courses in the collaborative exam mode.

Table 5.2 Number of Subjects Completed the Surveys by Course

		Exam Mode			Total
		Traditional	Participatory	Collaborative	
2004	Spring	MIS635-BLD (25) CIS675-DL (13)	CIS490-BLD (29) CIS675- BLD (24)	CIS490-BLD (21) MIS635-DL (26)	**138**
	Summer		MIS635-BLD (7) MIS635-DL (23)	CIS390-DL (21) CIS675-DL (17)	**68**
	Fall	CIS390-BLD (30) CIS431-BLD (56) CIS490-BLD (26) CIS663-BLD (10) CIS675-BLD (13)	CIS433-BLD (29) MIS635-BLD (11) CIS684-BLD (14)	CIS365-BLD (44) MIS635-DL (18) CIS675- DL (19) IE685-BLD (9)	**279**
Total		**173**	**137**	**175**	**485**

The number in parentheses () indicates the number of subjects in each course.
Bld: Blended mode; **DL**: Distance Learning
CIS: Computer and Information Systems (note: Computer Science and Information Systems are two separate departments now as of spring 2005);
MIS: Management Information Systems; **IE**: Industrial Engineering

Next, descriptive statistics are presented in two sections. Section 5.1.1 describes course characteristics, and Section 5.1.2 describes student characteristics.

5.1.1 Course Characteristics

As discussed in the research model, four aspects of course characteristics are of interest in this study: course mode (blended vs. distance learning), course level (undergraduate vs. graduate), course subject (technical vs. discussion-oriented), and class size.

Table 5.3 shows the number and percentage of students in blended vs. distance-learning courses in each exam mode. Overall, 72% of the courses were blended courses that had a combination of face-to-face meetings and online activities, and 28% of the courses were complete online courses.

Table 5.3 Course Mode by Exam Mode

		Course Mode		Total
		Blended	Distance Learning	
Exam Mode	Traditional	160 92.5%	13 7.5%	173 100%
	Participatory	114 83.2%	23 16.8%	137 100%
	Collaborative	74 42.3%	101 57.7%	175 100%
Total		**348** **71.8%**	**137** **28.2%**	**485** **100%**

Table 5.4 shows the number and percentage of students in undergraduate vs. graduate courses in each exam mode. Overall, about half (52.8%) of the students were enrolled in undergraduate courses, and the other half (47.2%) in graduate courses. For each exam mode, the distribution was about even except for the traditional exam, where a slightly higher percentage of students were enrolled in undergraduate courses (64.7%).

Table 5.4 Course Level by Exam Mode

		Course Level		Total
		Undergraduate	Graduate	
Exam Mode	Traditional	112 64.7%	61 35.3%	173 100%
	Participatory	58 42.3%	79 57.7%	137 100%
	Collaborative	86 49.1%	89 50.9%	175 100%
Total		**256** **52.8%**	**229** **47.2%**	**485** **100**

Table 5.5 shows the number and percentage of students in technical vs. discussion-oriented courses in each exam mode. Each course was evaluated to determine whether the course primary focus is technical, i.e., to teach technical facts and principles

of the subject matter (e.g., programming, coding, technical analysis), or discussion-oriented, i.e., to help students develop thinking skills. Relatively speaking, the following courses are classified as technical courses: CIS365, CIS431, CIS433, CIS490, and the following courses are classified as discussion-oriented courses: CIS390, MIS635, CIS663, CIS675, CIS684, IE685. Overall, about half (42.3%) of the courses in the study were technical-oriented, and the other half (57.7%) were discussion-oriented. For each exam mode, there was also an even combination of the two types of courses.

Table 5.5 Course Subject by Exam Mode

		Course Subject		**Total**
		Technical-oriented	Discussion-oriented	
Exam Mode	Traditional	82 47.4%	91 52.6%	173 100%
	Participatory	58 42.3%	79 57.7%	137 100%
	Collaborative	65 37.1%	110 62.9%	175 100%
Total		**205** **42.3%**	**280** **57.7%**	**485** **100%**

The classes that participated in the study varied in class size from seven students per class to 76 students (when two classes taught by the same professor participated in the same exam mode). The class size was categorized into three levels: low (1-19 students per class), medium (20-29 students per class), and high (above 30 students per class). Table 5.6 shows the number and percentage of students in the three class sizes in each exam mode. Overall, 27% of the classes participated in the study were of small size, 28% of middle size, and 44% of large size. For each exam mode, there was also an even combination of classes of the three sizes.

Table 5.6 Class Size by Exam Mode

		Class Size			Total
		Low (1-19)	Medium (20-29)	High (30+)	
Exam Mode	Traditional	36 20.8%	25 14.5%	112 64.7%	173 100%
	Participatory	32 23.4%	47 34.3%	58 42.3%	137 100%
	Collaborative	63 36.0%	68 38.9%	44 25.1%	175 100%
Total		**131** **27.0%**	**140** **28.9%**	**214** **44.1%**	**485** **100%**

5.1.2 Student Characteristics – Demographic Background

Table 5.7 provides the details of the demographic information of the students who participated in the study. Among the 485 subjects, 457 of them provided valid background information and 28 did not. Therefore the total number of responses is 457 in Table 5.7.

Table 5.7 Student Demographic Background

Student Characteristics		N	Percentage
Gender	Male	328	71.8%
	Female	129	28.2%
English as Native Language	No	252	55.1%
	Yes	205	44.9%
Program	Undergraduate	244	53.4%
	Graduate	213	46.6%
Work Experience	None	205	44.9%
	< 1 year	99	21.7%
	1-3 years	74	16.2%
	4-9 years	52	11.4%
	10-15 years	17	3.7%
	>15 years	10	2.2%
WebCT/Webboard Experience	None	48	10.5%
	One other course	61	13.3%
	2-4 other courses	214	46.8%
	5 or more other courses	134	29.3%
Expected Exam Grade	100-96	116	25.4%
	95-91	169	37.0%
	90-86	87	19.0%
	85-81	41	9.0%
	80-76	27	5.9%
	75-71	11	2.4%
	<70-66	2	0.4%
	65-61	2	0.4%
	< 60	2	0.4%
Previous Online Exam Experience	Never heard of it before	307	67.2%
	Heard from other students but have not experienced myself	109	23.9%
	Participated in it myself	32	7.0%
	Partially participated where I only contributed questions	9	2.0%

As shown, the majority of the students in the study are male (71.8%), which is close to the overall population in the school in this study. Most students are not native English speakers (55.1%). The percentage of undergraduate vs. graduate students in this study is relatively balanced (53.4% vs. 46.6%). The majorities of the subjects either do not have any course-related work experience (44.9%), or have less than one year work experience related to the course (21.7%). Nearly 90% of the students in the study had used WebCT or Webboard in at least one other course before the study, which suggests the proliferation of the use of online learning systems at NJIT. Over 60% of the students expected their exam grades to be higher than 90 points, and few students expected exam grades to be lower than 80 points (9.5%).

Students were also asked about their previous experiences with the NJIT online exam, which was used experimentally in a few classes in the IS department during the 1999-2001 period. The majority of students (67.2%) had never heard about the online exam and its concept before. About 24% students heard about the online exam and its procedures from other students, but they had not experienced it themselves. 7% of students participated in the online exam themselves, some of whom were participants in this experiment in earlier semesters (i.e., spring or summer 04 semester). Some professors partially adopted the online exam process by using only the question design phase, and students were asked whether they participated in such exams before. 2% reported that they participated in such partial online exams, where they contributed exam questions only.

5.2 Index Validation

This section presents the validation of the constructs in the pre-exam and post-exam surveys. Section 5.2.1 discusses the constructs in the pre-exam survey, and Section 5.2.2 discusses the constructs in the post-exam survey. Section 5.2.3 discusses the normalization tests, and data transformation procedures and results. Question items are identified using question IDs in this section, and the corresponding items can be found in Section 5.4.

5.2.1 Pre-exam Questionnaire Scale Validation

In addition to the subject's demographic information discussed in the previous section, the pre-exam questionnaire also consists of 20 questions (items) that measure students' learning predispositions, including deep learning and collaborative learning. This section validates these constructs from the pre-exam questionnaire using principal component factor analysis in SPSS. Before the analysis, negative items were converted into positive ones, and are denoted with an "r" after the item number (e.g. preq11r).

Initial factor analysis was conducted after the pilot study, and two items were eliminated (PREQ111, PREQ112R) as they did not converge with other items as expected. The remaining 18 items were included in the final analysis, and the principal component factor analysis with PROMAX rotation was conducted. PROMAX is an oblique rotation method which is used when there is no absolute theoretical foundation that there is no correlation among the constructs. The test yielded three factors, as shown in the rotated component matrix in Table 5.8. Since item PREQ120R had a negative loading onto factor 2, which reduced the interpretability of the factor, this item was eliminated from further analysis.

Table 5.8 Initial Factor Loadings of 18 Items in the Pre-exam Survey

	Factor		
	1	2	3
PREQ115	.819		
PREQ113	.774		
PREQ114R	.774		
PREQ118R	.760		
PREQ119	.755		
PREQ116R	.681		
PREQ117	.604		
PREQ14		.729	
PREQ12		.711	
PREQ16		.709	
PREQ18		.692	
PREQ110		.664	
PREQ120R- Eliminated		-.618	
PREQ15R			.743
PREQ17R			.699
PREQ11R			.697
PREQ19R			.686
PREQ13R			.605

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

The remaining 17 items were included in the final factor analysis, and the results are shown in Table 5.9 - Table 5.12. Two important outputs are the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and Bartlett's test of sphericity. The KMO statistic varies between 0 and 1, with a value close to 1 indicating that patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors. Values between 0.7 and 0.8 are considered good, values between 0.8 and 0.9 are great, and values above 0.9 are superb. A significant Bartlett's test shows that the R-matrix is not an identity matrix, and therefore factor analysis is appropriate. As shown in Table 5.9, the KMO result of .899 and Bartlett's result of .000 indicate that the sample is adequate and factor analysis is appropriate for these data.

Table 5.9 KMO and Bartlett's Test of the Pre-exam Survey

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.899
Bartlett's Test of Sphericity	Approx. Chi-Square	3161.836
	df	136
	Sig.	.000

Three factors were extracted from the 17 finalized pre-exam questionnaire items. The names given to the factors and the loadings of the items on them are shown in Table 5.10. As shown, the items measuring the deep learning construct split into two factors. The positive items converged into one factor, which is termed **deep learning**. The negative items converged into another factor, which is termed **surface learning**. The items measuring collaborative learning converged into one factor as expected.

Table 5.10 Final Factor Loading of the 17 Items in the Pre-exam Survey

	Factor		
	Pre-F1 Collaborative Learning	**Pre-F2 Deep Learning**	**Pre-F3 Surface Learning**
PREQ115	.829		
PREQ113	.780		
PREQ119	.767		
PREQ114R	.765		
PREQ118R	.761		
PREQ116R	.673		
PREQ117	.617		
PREQ12		.779	
PREQ14		.773	
PREQ18		.754	
PREQ16		.749	
PREQ110		.721	
PREQ15R			.832
PREQ11R			.671
PREQ17R			.644
PREQ19R			.631
PREQ13R			.494

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

Table 5.11 shows the variance explained by the three factors.

Table 5.11 Variance Explained by Pre-exam Factors

Factor	**Variance Explained (Before Rotation)**			**Variance Explained (After Rotation)** [a]
	Total	% of Variance	Cumulative %	Total
1. Collaborative Learning	6.167	36.279	36.279	4.695
2. Deep Learning	2.569	15.110	51.390	4.802
3. Surface Learning	1.296	7.624	59.013	3.866

Extraction Method: Principal Component Analysis.
[a] When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 5.12 is the component correlation matrix of the pre-exam factors. As shown, the three factors are correlated with each other. The table is symmetrical along the diagonal and thus only the upper half of cells is filled (i.e., the lower half is the same). The correlations confirm the use of PROMAX rotation method as described before.

Table 5.12 Component Correlation Matrix of Pre-exam Factors

	Pre-F1	**Pre-F2**	**Pre-F3**
Pre-F1	1.000	.342	.317
Pre-F2		1.000	.506
Pre-F3			1.000

The levels of internal consistency of the factors were measured using Cronbach's alpha. All three factors achieved alpha levels above .70. Please see Section 5.3 for details.

5.2.2 Post-exam Questionnaire Scale Validation

This section discusses the validation of scales from the post-exam questionnaire using factor analysis in SPSS. Items related to the following constructs in the research model were included in the post-exam questionnaire, including deep exam study; social engagement, perceived learning, satisfaction, and perceived fairness in grading. Before the analysis, negative items were converted into positive ones, and are denoted with an "r" after the item number (e.g. postq11r).

In the three post exam questionnaires for the three exam modes, 39 questions were common items. There were 22 additional question items in the questionnaires for the two online exam modes. Given the focus on the comparison of the different exam modes using the common questions, the additional questions are not included in this chapter, but they are analyzed and discussed in later chapters. After the pilot study, factor analysis was conducted and three questions (POSTQ15R, POSTQ19R, POSTQ111R) were eliminated due to poor convergence with other items. The remaining 36 items were used in the final analysis, and the principal component factor analysis with PROMAX rotation was conducted on the data. Table 5.13 shows the results after the initial test.

Table 5.13 Initial Factor Loadings of 36 Items in the Post-Exam Survey

	Factor							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
POSTQ35	.875							
POSTQ319	.862							
POSTQ320	.840							
POSTQ315	.832							
POSTQ317	.827							
POSTQ33	.819							
POSTQ39	.758							
POST311R	.752							
POSTQ313	.739							
POST37R	.723							
POSTQ318	.635							
POST31R	.556							
POSTQ321	.505			.463				
POST314R-Eliminated	.496							
POSTQ18		.804						
POSTQ14		.755						
POSTQ12		.677						
POSTQ112		.601						
POSTQ110		.597						
POSTQ16		.501						
POSTQ26			.789					
POSTQ21			.754					
POSTQ25			.746					
POSTQ22			.713					
POSTQ312				.904				
POST38R				.846				
POSTQ32				.543				
POST13R					.852			
POST11R					.820			
POST316R						.924		
POSTQ34						.774		.442
POSTQ23R-Eliminated								
POSTQ17R-Eliminated							.774	
POST24R- Eliminated							.560	
POSTQ310- Eliminated								.831
POSTQ36 - Eliminated								.406

Extraction Method: Principal Component Analysis. Rotation Method: Promax with Kaiser Normalization.

The initial test yielded eight factors, as shown in the rotated component matrix in Table 5.13. Item POSTQ23R was eliminated, which was not loaded to any factors. Items POSTQ17R, POSTQ24R, POSTQ310R, POSTQ314R, and POSTQ36 were eliminated to reduce the number of factors and to increase the interpretability of the constructs. The remaining 30 items were included in the final analysis, and the results are shown in Table 5.14 - Table 5.17. As show in Table 5.14, the KMO result of .934 and Bartlett's result of .000 indicate that the sample is adequate and factor analysis is appropriate for these data.

Table 5.14 KMO and Bartlett's Test of the Post-exam Survey

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.934
Bartlett's Test of Sphericity	Approx. Chi-Square	5965.426
	df	435
	Sig.	.000

Six factors were extracted from the 30 finalized post-exam questionnaire items. The names given to the factors and the loadings of the items on them are shown in Table 5.15. As shown, the items measuring the deep exam study construct split into two factors. The positive items converged into one factor, which is termed **deep exam study**. The two negative items converged into another factor, and is termed **surface exam study**. The other items converged into the corresponding constructs as expected.

Table 5.15 Final Factor Loadings of the 30 Items in the Post-exam Survey

	Factor					
	Post-F1 Perceived Learning	**Post-F2 Deep Exam Study**	**Post-F3 Social Engagement**	**Post-F4 Satisfaction**	**Post-F5 Surface Exam Study**	**Post-F6 Fairness in Grading**
POSTQ319	.881					
POSTQ35	.856					
POSTQ317	.855					
POSTQ315	.841					
POSTQ320	.830					
POSTQ33	.817					
POSTQ39	.780					
POSTQ313	.772					
POST311R	.708					
POST37R	.702					
POSTQ318	.659					
POST31R	.535					
POSTQ112		.786				
POSTQ18		.727				
POSTQ110		.707				
POSTQ12		.699				
POSTQ16		.624				
POSTQ14		.608				
POSTQ26			.802			
POSTQ21			.796			
POSTQ25			.782			
POSTQ22			.758			
POSTQ312				.856		
POST38R				.841		
POSTQ32				.539		
POSTQ321				.462		
POST13R					.858	
POST11R					.849	
POST316R						.899
POSTQ34						.809

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

Table 5.16 shows the variance explained by the six factors.

Table 5.16 Variance Explained by Post-exam Factors

Factor	Variance Explained (Before Rotation)			Variance Explained (After Rotation) [a]
	Total	% of Variance	Cumulative %	Total
1. Perceived Learning	10.645	35.484	35.484	10.088
2. Deep Exam Study	2.483	8.275	43.759	5.729
3. Social Engagement	1.816	6.054	49.813	5.069
4. Satisfaction	1.567	5.223	55.036	4.611
5. Surface Exam Study	1.377	4.590	59.626	1.756
6. Fairness in Grading	1.137	3.790	63.416	2.603

Extraction Method: Principal Component Analysis.
[a] When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 5.17 is the component correlation matrix of the post-exam factors. The table is symmetrical along the diagonal and thus only the upper half of cells is filled (i.e., the lower half is the same). As shown, the six factors are correlated with each other. The correlations confirm the use of PROMAX rotation method as described before.

Table 5.17 Component Correlation Matrix of Post-exam Factors

	Post-F1	**Post-F2**	**Post-F3**	**Post-F4**	**Post-F5**	**Post-F6**
Post-F1	1.000	.513	.486	.497	-.008	.319
Post-F2		1.000	.412	.233	.037	.022
Post-F3			1.000	.246	-.022	.024
Post-F4				1.000	-.045	.253
Post-F5					1.000	.080
Post-F6						1.000

The levels of internal consistency reliability of the factors were measured using Cronbach's alpha. All six factors achieved alpha levels above .70. Please see Section 5.4 for details.

5.3 Normal Distribution Test and Data Transformation

Before further data analysis, normal distribution tests were conducted which is the assumption of most parametric tests. Section 5.3.1 reports the normality test and transformation results on the factors validated in the previous section. Section 5.3.2 reports the normality test and transformation results of the grades data.

5.3.1 Normality Test and Transformation of Scales

The mean scores were calculated for all the factors, and the data were tested for normal distribution. Each factor was tested for Skewness (the tilt, or lack or it, in a distribution) and Kurtosis (the peakedness of a distribution). A common rule-of-thumb for normality is that the skewness and kurtosis scores should be within the range of +2.5 to –2.5 times their standard errors. A Kolmogorov-Smirnov (K-S) test was also used and the significance level p=0.01 was adopted. Table 5.18 shows the results of the normal distribution test of the sample.

Table 5.18 Normal Distribution Test of the Sample

Factor		Skewness		Kurtosis		Kolmogorov -Smirnov Z	P Value	H_0 Rejected
		Statistic	S. E.	Statistic	S. E.			
Pre-F1	**Pre-exam Collaborative Learning**	-0.371	0.119	-0.567	0.238	1.46	0.028	Not Rejected
Pre-F2	**Pre-exam Deep Learning**	-0.384	0.119	-0.289	0.238	1.25	0.090	Not Rejected
Pre-F3	**Pre-exam Surface Learning**	-1.142	0.119	1.631	0.238	2.63	0.000	Rejected
Post-F1	**Perceived Learning**	-0.788	0.125	1.099	0.250	1.85	0.002	Rejected
Post-F2	**Deep Exam Study**	-0.470	0.124	0.158	0.248	1.33	0.058	Not Rejected
Post-F3	**Surface Exam Study**	-0.067	0.124	-1.008	0.248	2.01	0.001	Rejected
Post-F4	**Social Engagement**	0.062	0.124	-0.931	0.248	1.35	0.053	Not Rejected
Post-F5	**Satisfaction**	-0.500	0.125	-0.396	0.250	1.59	0.013	Not Rejected
Post-F6	**Fairness in Grading**	-0.643	0.125	-0.175	0.250	2.83	0.000	Rejected

H_0: Sample Distribution fit into normal distribution at significant level p=.01

Data transformations were conducted for the factors (Pre-F3, Post-F1, Post-F3, Post-F6) that failed the K-S test (rejected the null hypothesis). The transformation methods work by compressing the right side of the distribution more than the left side (i.e., on positively skewed data), but the sample was negatively skewed (skewness statistics <0) in all but one factor. Therefore the data was first reflected and a constant was added to bring the minimum to 1 (i.e., 8-Average). Transformations were then applied, and finally the data was reflected again to restore the original order. The above procedure with Log_e (Ln) transformation was successful in bringing Pre-F3 to normal distribution. Post-F1 was transformed by squaring the mean score (i.e., mean score * mean score). Post-F3 and Post-F6 failed to be transformed into normal distribution using the simple transformation methods. Table 5.19 shows the results of the normal distribution test of F3 and F4 after the transformation.

Table 5.19 Normal Distribution Test of the Transformed Measures

Factor		Skewness		Kurtosis		Kolmogorov-Smirnov Z	P Value	H_0 Rejected
		Statistic	S. E.	Statistic	S. E.			
Pre-F3	**Pre-exam Surface Learning: (2.95-Ln(8-AVG))**	-0.058	0.119	-0.459	0.238	1.28	0.077	Not Rejected
Post-F1	**Perceived Learning: (AVG*AVG)**	0.150	0.125	-0.033	0.250	1.18	0.121	Not Rejected

H_0: Sample Distribution fit into normal distribution at significant level p=.01

5.3.2 Normality Test and Transformation of Grades

Students' exam grades (exam total) from the experiment in which they participated were collected. For the online exams, additional data were collected including question quality grade, grading quality grade, and answer grades (both student grader's and the instructor's). When the class has another exam (e.g., midterm or final) or term project, the grades were collected as well (termed other grade). All the grades were tested for skewness and kurtosis. A Kolmogorov-Smirnov (K-S) test was also used and the significance level p=0.01 was adopted. Table 5.20 shows the results of the normal distribution test of the sample.

Table 5.20 Normal Distribution Test of Grades

Grade		Skewness		Kurtosis		Kolmogorov -Smirnov Z	P Value	H_0 Rejected
		Statistic	S. E.	Statistic	S. E.			
P+C	Question Quality	-1.584	0.143	2.451	0.285	3.108	.000	Rejected
P+C	Grading Quality	-2.578	0.145	7.03	0.289	4.248	.000	Rejected
T+P+C	Exam Total	-1.731	0.115	3.526	0.229	3.301	.000	Rejected
T+P+C	Other Exam Grade	-0.93	0.134	1.282	0.268	1.426	.034	Not Rejected

H_0: Sample Distribution fit into normal distribution at significant level p=.01
P+C: Data were available only in the Participatory and the Collaborative Exams
T+P+C: Data were available in all three exam modes

Data transformations were conducted for question quality, grading quality, and exam total grades, which failed the K-S test (rejected the null hypothesis). The Log_e (Ln) transformation was successful in bringing exam total grade to normal distribution. Question quality and grading quality grades failed to be transformed into normal distribution using the simple transformation methods. Table 5.21 shows the results of the normal distribution test of exam total after the transformation.

Table 5.21 Normal Distribution Test of the Transformed Grades

Grades		Skewness		Kurtosis		Kolmogorov -Smirnov Z	P Value	H_0 Rejected
		Statistic	S. E.	Statistic	S. E.			
T+P+C	Exam Total: (5.33-Ln(101-Original))	0.481	0.115	0.521	0.23	1.174	.127	Not Rejected

H_0: Sample Distribution fit into normal distribution at significant level p=.01
T+P+C: Data were available in all three exam modes

5.4 Summary of Scales

Table 5.22 summarizes all the constructs (factors) found in the pre and post-exam surveys. The reliability of the constructs was measured using Cronbach's alpha, and all alpha values are greater than .70, indicating that all nine constructs are reliable. In addition to the Cronbach's alpha, the table also lists the items for each construct. Items are listed in decreasing order of their loading values to each factor. All items were measured using seven-point scale from Strongly Disagree (1) to Strongly Agree (7), except for item POSTQ21R as noted in the table. In addition, the following items show questions for the course CIS675 as an example. The course names were automatically customized in the online surveys to match with each course in the survey.

Table 5.22 Summary of Factors and Reliability

Factor	Item No.	Item [a, b]
Pre-exam Survey		
Collaborative Learning (Alpha=.87)	PREQ115	I have learned a lot from other students in our group assignments/projects in CIS675. Strongly Disagree: 1-2-3-4-5-6-7: Strongly Agree
	PREQ113	I enjoyed working with other students on group assignments/projects in CIS675.
	PREQ119	Interacting with other students has played an important role in my learning in CIS675.
	PREQ114R	I would rather have done the group assignments/projects in CIS675 individually.
	PREQ118R	Contact with other students in CIS675 has NOT played an important part in my learning.
	PREQ116R	The help I got from other students in this course was useless or misleading.
	PREQ117	Students in my CIS675 class tend to be very cooperative in sharing knowledge and learning together.
Deep Learning (Alpha=.83)	PREQ12	I find that at times studying CIS675 gives me a feeling of deep personal satisfaction.
	PREQ14	I feel that virtually any topic in CIS675 can be highly interesting once I get into it.
	PREQ18	I work hard at CIS675 because I find the material interesting.
	PREQ16	I find that studying CIS675 can at times be as exciting as a good novel or movie.
	PREQ110	I come to most CIS675 classes with questions in mind that I want answered.

Table 5.22 Summary of Factors and Reliability (Continued)

Factor	Item No.	Item [a, b]
Surface Learning (Alpha=.78)	PREQ15R	I find I can get by in most exams at NJIT so far by memorizing key sections rather than trying to understand them.
	PREQ11R	My aim is to pass CIS675 while doing as little work as possible.
	PREQ17R	I find it is NOT helpful to study CIS675 topics in depth because it wastes time, when all you need is a passing acquaintance with topics.
	PREQ19R	I see no point in learning material that is not likely to be in the CIS675 examination.
	PREQ13R	I keep my work to the minimum because I do not find CIS675 very interesting.
Post-exam Survey		
Social Engagement (Alpha=.80)	POSTQ26	I was able to get to know some students better during the exam process.
	POSTQ21	The exam process was mainly a(n): Individual experience: 1-2-3-4-5-6-7: Group Experience [b]
	POSTQ25	The CIS 675 exam allowed me to form a kind of learning community with other students.
	POSTQ22	My understanding of course material was enhanced by interacting with other students in the class through the exam process.
Deep Exam Study (Alpha=.80)	POSTQ112	When I was studying for the CIS675 exam, I found most topics interesting and spent extra time trying to obtain more information about them.
	POSTQ18	I found that I had to do enough work on a topic so that I could form my own conclusions before I was satisfied in studying for the CIS675 exam.
	POSTQ110	I tested myself on important CIS675 topics until I understood them completely while studying for exam.
	POSTQ12	When I was studying for the exam, I put myself in the position of "an Information Systems Evaluation researcher"/ "a system analyst and designer"/ "an internal telecommunications manager and/or business professional using telecommunications" to try to understand his/her role in organizations. [c]
	POSTQ16	I spent a lot of time finding out more about interesting topics which have been discussed in CIS675 when I was studying for the exam.
	POSTQ14	When I was studying for the exam, I found that I could relate CIS 675's material to other subject areas.
Surface Exam Study (Alpha=.70)	POST13R	I found the best way to pass the CIS675 exam I participated in is to try to remember answers to likely questions.
	POST11R	In studying for the CIS675 exam, I learned most of the things by rote, going over and over them until I knew them by heart even if I did not understand them.

Table 5.22 Summary of Factors and Reliability (Continued)

Factor	**Item No.**	**Item** [a, b]
Perceived Learning (Alpha=.93)	POSTQ319	My ability to **judge** the value of ideas and **assess** the quality of arguments has been improved through the CIS675 exam.
	POSTQ35	My ability to **use** methods, concepts, and theories I learned in CIS 675 in new situations has been improved through the exam.
	POSTQ317	My skill to **compare and discriminate** between ideas has been improved through the CIS675 exam.
	POSTQ315	My ability to use course material to **make generalizations or predictions** has been improved through the CIS 675 exam.
	POSTQ320	I learned from answering the exam questions.
	POSTQ33	My **understanding of the meaning** of CIS 675 course material has enhanced through the exam.
	POSTQ39	I am better able to **see** different course **components** of CIS 675 and organize them in a meaningful way through the exam.
	POSTQ313	My skill to **relate knowledge from several areas** to make my argument has been improved through the CIS675 exam.
	POST311R	My ability to **recognize patterns** of CIS 675 course material and their underlying meanings has remained the same as before the exam.
	POST37R	My ability to **solve problems** using what I learned in CIS 675 has NOT been improved through the exam.
	POSTQ318	I learned from reading the exam questions posted online.
	POST31R	My knowledge of major concepts, methods, and theories of CIS 675 has NOT been improved through the exam.
Satisfaction (Alpha=.81)	POSTQ312	The exam provided a comfortable timeframe.
	POST38R	I felt under a lot of pressure taking the exam this way.
	POSTQ32	I enjoyed the examination process.
	POSTQ321	I would recommend using the exam version that I did in this course in the future.
Fairness in Grading (Alpha=.73)	POST316R	The final grade that I received on this exam was NOT fair.
	POSTQ34	I felt the grading process was fair.

[a]: The course number "CIS675" is shown as an example and it is customized for each course.
[b]: All items were measured using the same scale from strongly disagree (1) to strongly agree (7) except for item POSTQ21R.
[c]: This item was customized for each course Three customizations are shown as examples.

CHAPTER 6

QUANTITATIVE DATA ANALYSIS – OVERALL RESULTS

Before testing hypotheses and answering research questions, Section 6.1 first analyzes the pre-existing differences in learning dispositions among the three exam modes using pre-exam factors found in the previous chapter. Section 6.2 reports students' responses in the post-exam survey on the individual question item level with univariate analysis. The one-way ANOVA test results comparing the post-exam responses on the construct level and grades are presented in Section 6.3.

6.1 Learning Predispositions

Before examining students' responses in the post-exam survey, pre-existing differences in learning styles were examined using pre-exam data to determine if there was any significant difference among the three exam modes before the exam. A one-way ANOVA test was conducted to analyze the differences among the three exam modes using the three constructs validated in the previous chapter: deep learning, surface learning, and collaborative learning. Results are shown in Table 6.1. Note the negative construct, surface learning orientation, was converted back so that the construct name matches the construct mean (i.e., the higher the number, the more surface orientation in learning). As the results show, the significance levels of all three constructs are above .05. This indicates that there is no significant difference among students in the three exam modes before the exam in terms of their learning predispositions.

Table 6.1 Learning Predispositions - ANOVA

Learning Predispositions	Exam Mode	N	Mean	S. D.	F	Sig.
Collaborative Learning	Traditional	146	5.16	1.18	.10	.90
	Participatory	119	5.18	1.16		
	Collaborative	155	5.12	1.17		
Deep Learning	Traditional	146	4.57	1.30	2.52	.08
	Participatory	119	4.90	1.16		
	Collaborative	155	4.71	1.13		
Surface Learning ^{T}	Traditional	146	.85	.42	.83	.44
	Participatory	119	.78	.40		
	Collaborative	155	.82	.45		

^{T}: Transformed scale

6.2 Univariate Analysis

This section presents the analysis of the post-exam survey data on the individual question item level using univariate data analysis. Only items that have been validated in the factor analysis in the previous sections are analyzed. Section 6.2.1-6.2.2 present the items related to exam study strategy constructs, including deep/surface exam study and social engagement. Section 6.2.3-6.2.5 presents the items related to the exam outcome constructs, including satisfaction, perceived learning, and perceived fairness in grading.

6.2.1 Deep/Surface Exam Study

Table 6.2 provides the univariate analysis results of the six items of the deep exam study factor, including the distributions from Strongly Disagree (1) to Strongly Agree (7), means, and standard deviations. In five out of six items, students in the online participatory or collaborative exams reported adopting a deeper approach to exam study compared with those in the traditional exam. For example, students in the participatory exam reported the highest adoption of deep exam study in terms of finding the topics interesting and willing to spend extra time to study (Traditional=4.32, Participatory=4.68, Collaborative=4.52), adopting professional roles in understanding materials (e.g., put oneself in the position of a system analyst and designer) (T=4.45, P=4.78, C=4.77), spending extra time to find out additional information (T=4.01, P=4.89, C=4.66), and relating course materials to other subject areas (T=4.91, P=5.28, C=5.11). Students in the collaborative exam reported the highest level of achieving satisfaction by researching topics to form their own conclusions in exam study (T=4.62, P=4.85, C=4.97). The only item which had the highest score in the traditional exam is regarding testing oneself on important topics until one understands them completely (T=4.77, P=4.22, C=4.45). The results suggest that students in the participatory and collaborative online exams adopted a deeper approach to exam study, including taking professional and academic perspectives in understanding materials, while students in the traditional exam tested themselves on important topics before the exams.

Table 6.2 Items of Deep Exam study

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
When I was studying for the CIS675 exam, I found most topics interesting and spent extra time trying to obtain more information about them.	T	5.8	7.9	20.1	18.0	20.9	17.3	10.1	4.32	1.66
	P	2.7	9.1	9.1	18.2	30.9	17.3	12.7	4.68	1.55
	C	4.4	7.3	16.1	18.2	22.6	21.2	10.2	4.52	1.61
I found that I had to do enough work on a topic so that I could form my own conclusions before I was satisfied in studying for the CIS675 exam.	T	1.4	7.2	8.6	33.1	19.4	20.9	9.4	4.62	1.42
	P	.9	4.5	10.9	20.0	29.1	24.5	10.0	4.85	1.35
	C	1.5	2.9	10.2	17.5	29.2	27.7	10.9	4.97	1.35
I tested myself on important CIS675 topics until I understood them completely while studying for exam.	T	5.0	7.9	13.7	10.8	20.1	25.9	16.5	4.77	1.75
	P	8.2	10.9	16.4	11.8	26.4	20.9	5.5	4.22	1.71
	C	8.0	7.3	13.1	11.7	29.2	24.8	5.8	4.45	1.67
When I was studying for the exam, I put myself in the position of *a system analyst and designer* to try to understand his/her role in organizations.	T	5.8	10.8	11.5	17.3	22.3	23.7	8.6	4.45	1.68
	P	1.8	6.4	14.5	17.3	20.0	29.1	10.9	4.78	1.52
	C	2.9	7.3	13.9	15.3	21.9	24.1	14.6	4.77	1.62
I spent a lot of time finding out more about interesting topics which have been discussed in CIS675 when I was studying for the exam.	T	7.9	15.1	13.7	23.0	20.9	10.1	9.4	4.01	1.71
	P	.9	5.5	12.7	20.0	20.9	25.5	14.5	4.89	1.48
	C	3.6	10.2	10.9	14.6	26.3	21.2	13.1	4.66	1.65
When I was studying for the exam, I found that I could relate CIS 675's material to other subject areas.	T	5.0	4.3	10.1	14.4	25.9	22.3	18.0	4.91	1.64
	P	1.8	5.5	5.5	9.1	26.4	31.8	20.0	5.28	1.47
	C	3.6	6.6	7.3	7.3	27.0	29.2	19.0	5.11	1.61

Note: The course number "CIS675" is shown as an example and it is customized for each course.
T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam
SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree

Table 6.3 shows the univariate analysis results of the two items of the surface exam study factor. Note the negative items were converted back so that the item statement matches the mean (i.e., the higher the number, the more adoption of surface exam study). In both items, students in the traditional exam adopted noticeably higher levels of surface approach in their processing of information compared with those in the online exams, including trying to remember answers to likely questions as the best way to pass the exam (T=4.35, P=2.96, C=3.15), and learning by rote (T=3.94, P=2.89, C=3.00). Results suggest that students in the traditional exam had the highest adoption of surface exam study strategy in their processing of information.

Table 6.3 Items of Surface Exam Study

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
I found the best way to pass the CIS675 exam I participated in is to try to remember answers to likely questions.	T	7.9	8.6	10.1	21.6	24.5	20.9	6.5	4.35	1.65
	P	29.1	17.3	14.5	19.1	10.0	7.3	2.7	2.96	1.75
	C	25.5	19.0	12.4	16.1	13.9	11.7	1.5	3.15	1.79
In studying for the CIS675 exam, I learned most of the things by rote, going over and over them until I knew them by heart even if I did not understand them.	T	12.9	13.7	11.5	14.4	25.2	20.9	1.4	3.94	1.75
	P	31.8	18.2	13.6	13.6	12.7	8.2	1.8	2.89	1.77
	C	28.5	22.6	13.1	7.3	17.5	6.6	4.4	3.00	1.86

Note: The course number "CIS675" is shown as an example and it is customized for each course.
T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam
SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree

6.2.2 Social Engagement

Table 6.4 provides the univariate analysis results of the four items of the social engagement construct. In all four items, students in the collaborative online exam had the noticeably highest level of social engagement, including getting to know other students through the exam process (T=3.45, P=2.98, C=4.49), perceiving the exam as a group process(T=2.73, P=2.46, C=4.17), forming a learning community through exam study (T=3.57, P=3.84, C=4.51), and enhancing understanding of course materials by interacting with other students (T=3.80, P=3.81, C=4.46). Results suggest that students in the collaborative exam had the highest adoption of social engagement in exam study. Interestingly, it is also noticeable that the level of social engagement in the participatory exam was lower than or as low as in the traditional exam on most items. This suggests that without the small group activities, the level of social engagement in the online environment is as low as or even lower than in the traditional settings.

Table 6.4 Items of Social Engagement

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
I was able to get to know some students better during the exam process.	T	23.0	15.8	11.5	17.3	14.4	11.5	6.5	3.45	1.93
	P	30.9	20.0	10.9	14.5	10.0	9.1	4.5	2.98	1.89
	C	12.5	4.4	10.3	16.2	22.1	19.9	14.7	4.49	1.88
The exam process was mainly a(n):		I.						G.		
	T	47.5	10.1	5.8	14.4	7.2	11.5	3.6	2.73	2.00
	P	48.2	17.3	6.4	8.2	11.8	4.5	3.6	2.46	1.84
	C	12.5	11.8	16.2	11.8	17.6	14.0	16.2	4.17	1.99
The CIS 675 exam allowed me to form a kind of learning community with other students.	T	18.7	18.0	11.5	21.6	10.1	10.1	10.1	3.57	1.94
	P	17.3	15.5	6.4	20.0	20.0	10.0	10.9	3.84	1.96
	C	10.3	8.8	8.1	8.8	30.1	24.3	9.6	4.51	1.81
My understanding of course material was enhanced by interacting with other students in the class through the exam process.	T	19.4	11.5	11.5	16.5	17.3	15.8	7.9	3.80	1.96
	P	18.2	13.6	10.9	16.4	16.4	16.4	8.2	3.81	1.96
	C	10.3	13.2	8.1	8.8	23.5	19.9	16.2	4.46	1.96

T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam
SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree
I: Individual Experience; **G**: Group Experience

6.2.3 Satisfaction

Table 6.5 presents the univariate analysis results of the four items of the satisfaction construct. In terms of timeframe, students in the traditional exam had the highest perception of the exam being conducted in a comfortable timeframe (T=5.28, P=4.47, C=4.81). This corresponds to students' responses in the interviews where the two-week timeframe of the online exam was reported as too long compared with the couple of hours typically spent in a traditional exam. Students in the online exams had slightly lower levels of pressure compared with the traditional exam (reversed score: T=4.27, P=4.32, C=4.41), and they reported more enjoyment with the exam process (T=4.09, P=4.37, C=4.43). There was no noticeable difference in terms of the preference of recommending the exam process to be used in the future (T=4.64, P=4.68, C=4.64). These results are congruent with the interview responses. Students commented that although they enjoyed the easy access to resources and less pressure with the online exams, the lengthy exam timeline was a main issue they had to overcome, which affected their recommendation of the process to be used in the future. See more discussions in Chapter 8.

Table 6.5 Items of Satisfaction

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
The exam provided a comfortable timeframe.	T	3.6	2.9	7.2	10.9	20.3	33.3	21.7	5.28	1.54
	P	9.2	9.2	14.7	11.0	19.3	22.0	14.7	4.47	1.88
	C	10.5	6.8	9.8	6.8	19.5	23.3	23.3	4.81	1.98
I felt under a lot of pressure taking the exam this way. R	T	6.5	13.0	21.7	10.9	15.2	18.8	13.8	4.27	1.85
	P	9.2	12.8	18.3	9.2	11.9	23.9	14.7	4.32	1.96
	C	10.5	9.8	14.3	14.3	15.8	15.0	20.3	4.41	1.98
I enjoyed the examination process.	T	10.9	12.3	10.1	24.6	18.8	11.6	11.6	4.09	1.81
	P	11.0	9.2	11.0	14.7	22.9	17.4	13.8	4.37	1.88
	C	17.3	5.3	9.0	8.3	22.6	21.1	16.5	4.43	2.06
I would recommend using the exam version that I did in this course in the future.	T	5.8	10.1	8.0	20.3	16.7	24.6	14.5	4.64	1.76
	P	11.0	7.3	5.5	13.8	22.9	20.2	19.3	4.68	1.92
	C	15.0	5.3	6.0	9.8	21.8	22.6	19.5	4.64	2.03

R: Reversed item

T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam

SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree

6.2.4 Perceived Learning

Table 6.6 provides the univariate analysis results of the 12 items of the perceived learning construct. In ten out of the 12 items, students in the collaborative exam reported the highest level of learning among the three. For the other two items, perceptions of students in the participatory exam were the highest. For example, students in the collaborative exam reported the highest level of improvement in their knowledge of course concepts, methods and theories (T=4.75, P=5.18, C=5.24), and their understanding of course materials (T=4.77, P=4.79, C=4.89). They also reported enhancement in skills such as using knowledge in new situations (T=4.71, P=4.91, C=4.97), solving problems (T=4.57, P=4.84, C=4.90), recognizing patterns (T=4.09, P=4.03, C=4.13), making generalizations and predictions (T=4.51, P=4.66, C=4.86), comparing and discriminating ideas (T=4.42, P=4.56, C=4.68), and making judgment and assessment of the quality of ideas and arguments (T=4.46, P=4.78, C=4.91). In addition, students in the collaborative exam reported the highest level of learning in reading exam questions posted online (T=4.38, P=4.64, C=4.77) and in answering exam questions (T=4.64, P=5.37, C=5.44). Students in the participatory exam reported the highest level of improvements in seeing different course components and organizing them in a meaningful way (T=4.18, P=4.55, C=4.50), and relating knowledge from different academic areas (T=4.57, P=4.84, C=4.76). In 11 out of the 12 items, students in the traditional exam reported the lowest level of perceived learning among the three exam modes.

Table 6.6 Items of Perceived Learning

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
My ability to **judge** the value of ideas and **assess** the quality of arguments has been improved through the CIS675 exam.	T	4.3	9.4	6.5	28.3	28.3	13.8	9.4	4.46	1.52
	P	1.8	5.5	5.5	22.0	35.8	23.9	5.5	4.78	1.29
	C	4.5	3.8	7.5	10.5	36.8	27.8	9.0	4.91	1.45
My ability to **use** methods, concepts, and theories I learned in CIS 675 in new situations has been improved through the exam.	T	4.3	5.8	10.1	15.9	31.2	23.2	9.4	4.71	1.53
	P	3.7	4.6	7.3	17.4	27.5	27.5	11.9	4.91	1.49
	C	3.8	5.3	5.3	12.8	37.6	19.5	15.8	4.97	1.50
My skill to **compare and discriminate** between ideas has been improved through the CIS675 exam.	T	5.1	6.5	10.1	29.0	28.3	10.9	10.1	4.42	1.51
	P	1.8	7.3	8.3	28.4	30.3	17.4	6.4	4.56	1.35
	C	5.3	2.3	11.3	16.5	37.6	18.8	8.3	4.68	1.45
My ability to use course material to **make generalizations or predictions** has been improved through the CIS 675 exam	T	4.3	7.2	8.7	24.6	30.4	17.4	7.2	4.51	1.48
	P	.9	5.5	9.2	30.3	27.5	18.3	8.3	4.66	1.31
	C	5.3	3.0	5.3	22.6	27.8	22.6	13.5	4.86	1.52
I learned from answering the exam questions.	T	5.1	9.4	6.5	17.4	29.7	21.0	10.9	4.64	1.62
	P	.9	4.6	.9	9.2	39.4	24.8	20.2	5.37	1.27
	C	3.8	2.3	3.0	5.3	30.8	32.3	22.6	5.44	1.42
My **understanding of the meaning** of CIS 675 course material has enhanced through the exam.	T	5.8	4.3	7.2	16.7	31.9	23.9	10.1	4.77	1.54
	P	4.6	5.5	9.2	15.6	25.7	31.2	8.3	4.79	1.54
	C	6.0	5.3	4.5	9.0	38.3	26.3	10.5	4.89	1.55
I am better able to **see** different course **components** of CIS 675 and organize them in a meaningful way through the exam.	T	8.0	8.0	11.6	30.4	19.6	17.4	5.1	4.18	1.58
	P	2.8	8.3	10.1	26.6	26.6	13.8	11.9	4.55	1.51
	C	4.5	4.5	13.5	22.6	30.8	16.5	7.5	4.50	1.46
My skill to **relate knowledge from several areas** to make my argument has been improved through the CIS675 exam.	T	3.6	6.5	8.0	26.8	31.2	13.8	10.1	4.57	1.46
	P	2.8	5.5	7.3	16.5	33.0	26.6	8.3	4.84	1.41
	C	6.8	2.3	9.8	14.3	34.6	21.1	11.3	4.76	1.56

Table 6.6 Items of Perceived Learning (Continued)

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
My ability to **recognize patterns** of CIS 675 course material and their underlying meanings has <u>remained the same</u> as before the exam. [R]	**T**	7.2	8.0	19.6	21.7	25.4	13.8	4.3	4.09	1.54
	P	3.7	12.8	21.1	23.9	22.0	11.0	5.5	4.03	1.49
	C	6.0	11.3	15.8	24.1	23.3	12.8	6.8	4.13	1.58
My ability to **solve problems** using what I learned in CIS 675 has <u>NOT</u> been improved through the exam. [R]	**T**	5.1	8.0	15.9	18.8	18.1	16.7	17.4	4.57	1.75
	P	3.7	7.3	8.3	14.7	27.5	24.8	13.8	4.84	1.59
	C	6.0	5.3	7.5	14.3	25.6	23.3	18.0	4.90	1.69
I learned from reading the exam questions posted online.	**T**	10.9	9.4	5.8	24.6	16.7	18.8	13.8	4.38	1.86
	P	2.8	11.9	5.5	19.3	29.4	21.1	10.1	4.64	1.56
	C	3.8	3.0	11.3	19.5	30.1	21.8	10.5	4.77	1.46
My knowledge of major concepts, methods, and theories of CIS 675 has <u>NOT</u> been improved through the exam. [R]	**T**	6.5	8.7	10.1	13.0	21.7	18.8	21.0	4.75	1.83
	P	3.7	5.5	10.1	9.2	19.3	25.7	26.6	5.18	1.70
	C	5.3	5.3	9.8	4.5	17.3	30.8	27.1	5.24	1.76

Note: The course number "CIS675" is shown as an example and it is customized for each course.
[R]: Reversed item
T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam
SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree

6.2.5 Perceived Fairness in Grading

Table 6.7 presents the univariate analysis results of the two items of the perceived fairness in grading construct. In terms of fairness of grades, students in the online exams had higher perceptions of fairness of the *exam grades* they received than those in the traditional exam (T=4.99, P=5.18, C=5.15). On the other hand, students in the traditional exam had the highest perceptions of the fairness of the *grading process* than the other two (T=5.23, P=5.15, C=4.91). This suggests that although students in the online exams questioned the fairness of the grading process (e.g., peer grading mechanism), they perceived the grades they received as more fair than students in the traditional exam who received grades from professors. See more discussions with students' interviews in Chapter 8.

Table 6.7 Items of Perceived Fairness in Grading

%		SD 1	2	3	N 4	5	6	SA 7	M	SD
The final grade that I received on this exam was NOT fair. [R]	**T**	2.2	7.2	10.1	23.2	8.7	23.9	24.6	4.99	1.70
	P	4.6	4.6	7.3	16.5	15.6	21.1	30.3	5.18	1.74
	C	3.0	7.5	6.8	15.0	15.0	27.1	25.6	5.15	1.69
I felt the grading process was fair.	**T**	1.4	5.8	5.1	18.8	15.2	31.9	21.7	5.23	1.52
	P	.9	5.5	9.2	12.8	22.9	31.2	17.4	5.15	1.47
	C	4.5	8.3	8.3	12.8	18.8	31.6	15.8	4.91	1.70

[R]: Reversed item
T: Traditional Exam; **P**: Participatory Exam; **C**: Collaborative Exam
SD: Strongly Disagree; **N**: Neutral; **SA**: Strongly Agree

6.3 Overall Comparison

After examining survey results on the individual item level, this section compares the differences of the post-exam survey responses among the three exam modes on the factor level. In addition, differences in grades in the three exam modes are discussed.

As discussed in the factor analysis and normality tests in the previous chapter, four factors are normally distributed (or have been transformed into normal): deep exam study, social engagement, perceived learning and satisfaction. Two factors are not normally distributed: surface exam study and perceived fairness in grading. In addition, the exam total grade has been transformed into a normal distribution, and the question quality and the grading quality grades are not normally distributed. To compare the differences among the three exam modes, a one-way ANOVA test was conducted on the constructs and grades that are normally distributed. The Kruskal Wallis Test, which is the non-parametric equivalent of a one-way ANOVA, was conducted on the two constructs that are not normally distributed. A Mann-Whitney test, which is the non-parametric equivalent of a T-test, was conducted on the two grades that are not normally distributed. The ANOVA results are shown in Table 6.8. The Kruskal Wallis test results and the Mann-Whitney results are shown in Table 6.9 and Table 6.10, where the higher the mean rank, the higher the score.

Table 6.8 Overall Results 1 - ANOVA Test

Exam Study Strategies and Outcomes	Exam Mode	N	Mean	S. D.	F	Sig.
Deep Exam Study	Traditional	139	4.51	1.17	2.20	.112
	Participatory	110	4.78	1.00		
	Collaborative	137	4.74	1.18		
Social Engagement	Traditional	139	3.38	1.62	22.45**	.000
	Participatory	110	3.27	1.45		
	Collaborative	136	4.41	1.43		
Perceived Learning ^{T}	Traditional	138	21.78	10.33	3.09*	.047
	Participatory	109	23.91	10.04		
	Collaborative	133	24.78	10.19		
Satisfaction	Traditional	138	4.57	1.31	.22	.803
	Participatory	109	4.46	1.57		
	Collaborative	133	4.57	1.64		
Exam Total Grade ^{T}	Traditional	164	2.50	.84	24.38**	.000
	Participatory	127	3.15	.82		
	Collaborative	157	2.86	.70		

^{T}: Transformed scale **: Significant at p<.01 level. *: Significant at P<.05 level

Table 6.9 Overall Result 2 – Kruskal Wallis Test

Exam Study Strategies and Outcomes	Exam Mode	N	Mean Rank	Chi-Square	Sig.
Surface Exam Study	Traditional	139	243.75	45.02**	.000
	Participatory	110	159.55		
	Collaborative	137	169.78		
Fairness in Grading	Traditional	138	192.28	.27	.876
	Participatory	109	193.02		
	Collaborative	133	186.59		

**: Significant at p<.01 level.

Table 6.10 Overall Result 3 – Mann-Whitney Test

Grade	Exam Mode	N	Mean Rank	Mann-Whitney U	Sig.
Question Quality Grade	Participatory	128	161.95	8388.50**	.004
	Collaborative	164	133.65		
Grading Quality Grade	Participatory	127	175.99	5581.00**	.000
	Collaborative	162	115.95		

In terms of the post-exam factors, the differences are significant among the three exam modes in social engagement (p<.01), perceived learning (p<.05), and surface exam study (p<.01). Figures 6.1-6.3 provide the diagrams of the three constructs with significant differences. In all three diagrams, the higher the score, the higher the value of the variable (i.e., social engagement, perceived learning, adoption of surface exam study strategy). Post-hoc analysis reveals that in terms of social engagement, students in the collaborative exam were significantly more engaged than students in the traditional (p<.01) and the participatory exam (p<.01). In terms of perceived learning, students in the collaborative exam reported significantly higher perceptions of learning than those in the traditional exam (p<.05). There is no post-hoc analysis for a Kruskal Wallis test. As shown in Figure 6.3, students in the participatory exam had the lowest level of surface exam study while those in the traditional exam had the highest. These results are congruent with the item-level univariate analysis in the previous section

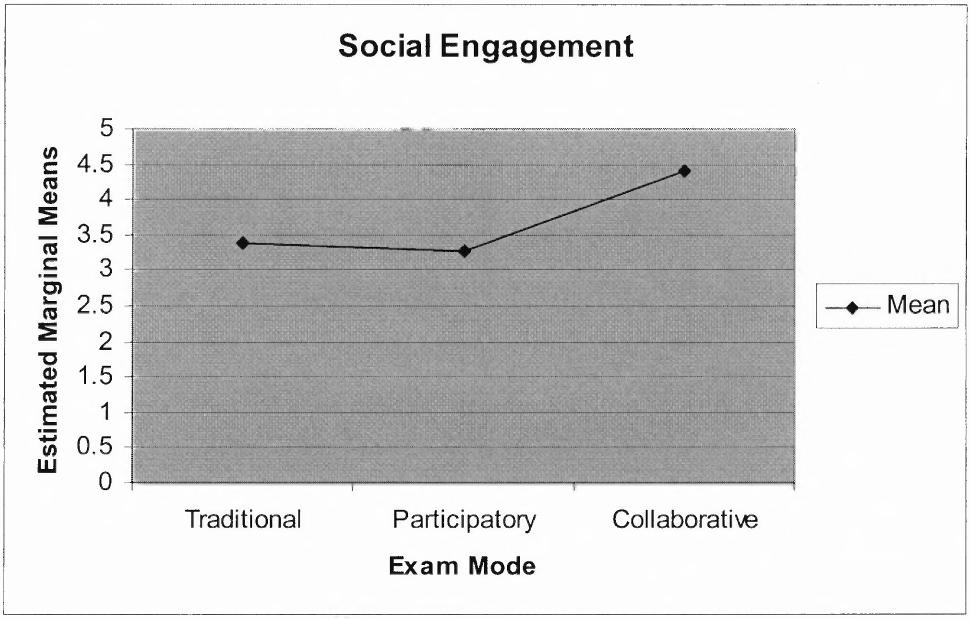


Figure 6.1 Overall significant result 1- social engagement.

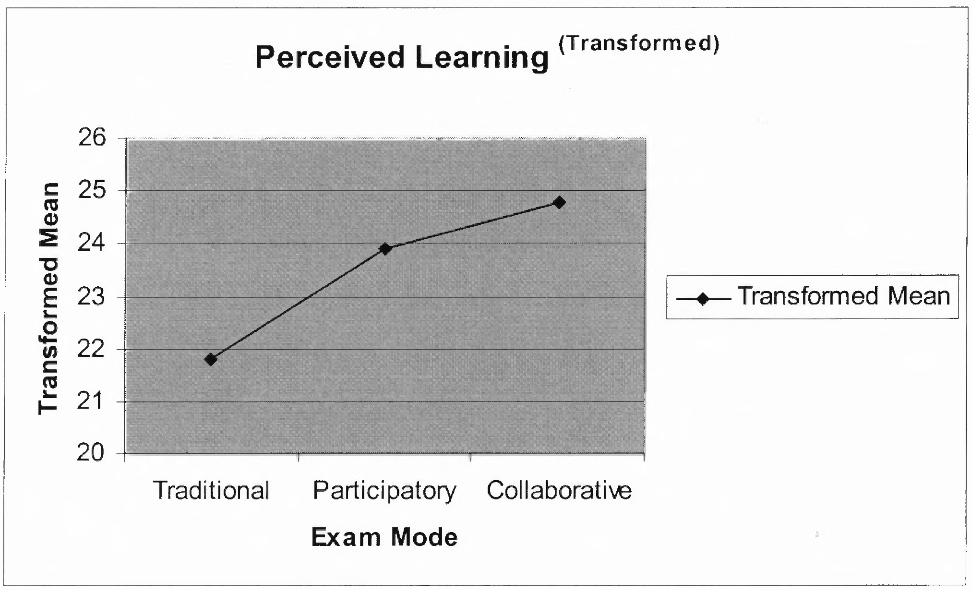


Figure 6.2 Overall significant result 2- perceived learning.

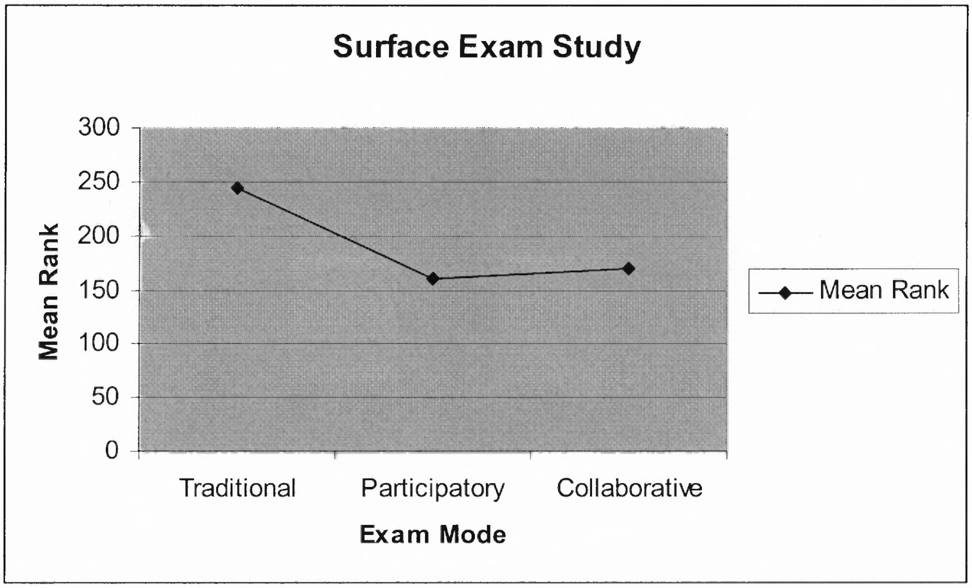


Figure 6.3 Overall significant result 3- surface exam study.

Regarding grades, the ANOVA analysis of exam total grades in Table 6.8 show that students in the participatory exam received the highest grades, while students in the traditional exam received the lowest. Post-hoc analysis revealed that the differences are significant. Similarly, the Mann-Whitney test results of the question quality and grading quality grades shown in Table 6.10 suggest that students in the participatory exam received higher grades for their question quality and grading quality than those in the collaborative exam, and the difference is significant. These findings are the opposite of what were hypothesized, as will be shown in the next chapter. See more discussions in Chapter 9.

To measure actual learning, additional data on the final exam/project were collected from the courses that participated in the experiment as the midterm exam. All the courses in the fall semester and one course (CIS 390) in the summer semester participated in the experiment as the midterm exam. Grades from the final exam were collected, or when the course did not offer the final exam but a final project, the final project grades were collected instead of the final exam grades. A one-way ANOVA test was conducted on the final exam/project grades and the results are shown in Table 6.11. Results reveal that students in the collaborative exam achieved the highest final exam/project grades than students in the other two exam modes. Using the final exam or project scores as the reflection of actual learning, this suggests that students in the collaborative online exam may have indeed learned the most from the midterm exam compared with the other two exams. However, the difference is not significant.

Table 6.11 Overall Results 4 – Final Exam/Project Grades

Grade	Exam Mode	N	Mean	S. D.	F	Sig.
Other (Final exam/project) Grade	Traditional	129	78.41	16.38	1.03	.359
	Participatory	25	75.64	13.23		
	Collaborative	74	81.42	24.68		

CHAPTER 7

QUANTITATIVE DATA ANALYSIS – HYPOTHESES TESTING

This chapter answers the first six sets of research questions and tests hypotheses one through six using the factors validated in Chapter 5. Research questions seven through nine are investigated qualitatively using log and interview data and are discussed in Chapter 8. Section 7.1 presents the main results of hypotheses testing. Section 7.2 provides the analysis of the additional survey data gathered from the two online exam modes. Note in this section the two negative constructs (surface learning orientation, surface exam study) were converted so that the construct names match the construct means (i.e., the higher the number, the more surface orientation in learning or use of surface exam study strategy).

7.1 Hypotheses Testing

This section presents the main results of hypotheses testing. Section 7.1.1 provides the results regarding the independent variable in comparing the three exam modes. Section 7.1.2 provides the results regarding the moderator variable of student learning predispositions on exam study strategies. Section 7.1.3 provides the results regarding the moderator variable of courses characteristics on the collaborative exam outcomes. Section 7.1.4 reports results regarding the dependent variables of exam outcomes and the relationships among them.

Note given deep exam study resulted in two factors (deep exam study, surface exam study) using factor analysis, hypotheses H1.2, H2.1, H3.1 and H3.3 were tested using the deep exam study and the surface exam study factors, respectively.

7.1.1 Comparison of Exam Modes

After presenting the overall results of the three exam modes in the previous chapter, this section presents the results in comparing the collaborative exam with the participatory exam (Section 7.1.1.1), and the collaborative exam with the traditional exam (Section 7.1.1.2).

7.1.1.1 Collaborative vs. Participatory Exam. This section tests the first hypotheses comparing the two modes of online exams, including social engagement in exam study (H1.1), deep/surface exam study (H1.2), perceived learning (H1.3), satisfaction (H1.4), grade distribution (H1.5), perceived fairness in grading (H1.6), question quality (H1.7), and grading quality(H1.8).

T-tests for two independent samples were conducted on post-exam constructs and grades that are normally distributed, and a Mann-Whitney test, the nonparametric equivalent of the T-test, was conducted on constructs and grades that are not normally distributed. Table 7.1 shows the result of the T-test, and Table 7.2 shows the result of the Mann-Whitney test. The results show that students in the collaborative exam reported significantly higher levels of social engagement than those in the participatory exam. However, the total exam grades students received in the collaborative exam were significantly lower than those in the participatory exam, and the question quality grade was lower too.

Table 7.1 Collaborative vs. Participatory Exam (T-test)

Exam Study Strategies and Outcomes	**T**	**df**	**P**
Social Engagement	6.15**	244	.000
Deep Exam Study	-0.29	245	.775
Perceived Learning ^{T}	0.67	240	.504
Satisfaction	0.55	240	.582
Exam Total Grade ^{T}	-3.22**	282	.001

^{T}: Transformed scale **: Significant at p<.01 level.

Table 7.2 Collaborative vs. Participatory Exam (Mann-Whitney Test)

Exam Study Strategies and Outcomes	**Mann-Whitney U**	**P**
Surface Exam Study	7057.50	.389
Post-exam Fairness in Grading	7016.00	.666
Question Quality Grade	8388.50**	.004

**: Significant at p<.01 level.

Therefore:

H1.1: Students taking the collaborative examination will have higher social engagement in the exam studying process than students taking the participatory exam.
Supported (mean=4.41, 3.27), significant (p<.01)

H1.2: Students taking the collaborative examination will have higher adoption of deep learning in the exam studying process than students taking the participatory exam.
Not in predicted direction using deep exam study (mean=4.74, 4.78), not significant
Not in predicted direction using surface exam study (mean rank=169.78, 159.55), not significant

<u>H1.3</u>: Students taking the collaborative examination will have higher perception of learning than students taking the participatory exam.
In predicted direction (mean=24.78, 23.91), but not significant

<u>H1.4</u>: Students taking the collaborative examination will have higher satisfaction of the exam than students taking the participatory exam.
In predicted direction (mean=4.57, 4.46), but not significant

<u>H1.5:</u> Students taking the collaborative examination will have higher grades than students taking the participatory exam.
Reverse effect found (mean=2.86, 3.15), significant (p<.01)

<u>H1.6</u>: Students taking the collaborative examination will have higher perception of fairness in grading than students taking the participatory exam.
Not in predicted direction (mean rank=186.59, 193.02), not significant

<u>H1.7</u>: Students' questions will have higher quality (judged by the question quality grade given by the instructor) than those from the participatory exam.
Reverse effect found (mean rank=133.65, 161.95), significant (p<.01)

To test the differences in student grading quality in the participatory and the collaborative exams, the correlation between students' grading and the instructor's grading was calculated for each of the two online exam modes, respectively. Given each student designed two to three questions and therefore graded two to three answers, the average score was calculated as student answer grading (i.e., student answer grading= average (student answer 1-3 grading)). The same method was used in calculating instructor's answer grading (i.e., instructor answer grading = average (instructor answer 1-3 grading)). As the results shown in Table 7.3, the correlations between student grading and instructor's grading are significant and are high in both the collaborative and the participatory exam. Furthermore, the correlation is higher in the collaborative exam than in the participatory exam. This suggests that students in the collaborative exam achieved higher grading quality than students in the participatory exam.

Table 7.3 Student Grading and Instructor Grading (Correlation)

		Student Answer Grading
Instructor Answer Grading	Participatory	.699**
	Collaborative	.876**

** Correlation is significant at the 0.01 level (2-tailed)

Therefore:

> **H1.8**: Student's grading will have higher correlation with the instructor's grading in the collaborative examination than in the participatory exam.
> **Supported**

7.1.1.2 Collaborative vs. Traditional Exam. This section tests the second sets of hypotheses comparing the collaborative online exam with the traditional in-class exam regarding deep/surface exam study (H2.1), social engagement (H2.2), perceived learning (H2.3), satisfaction (H2.4), grade distribution (H2.5), and perceived fairness in grading (H2.6).

A T-test and a Mann-Whitney test were conducted. Table 7.4 shows the result of the T-test, and Table 7.5 shows the result of the Mann-Whitney test. Results show that students in the collaborative exam reported significantly higher levels of social engagement, higher perceived learning, and lower levels of surface exam study than those in the traditional exam. Students in the collaborative exam also achieved significantly higher scores than those in the traditional exam.

Table 7.4 Collaborative vs. Traditional Exam (T-test)

Exam Study Strategies and Outcomes	T	df	P
Deep Exam Study	1.63	274	.10
Social Engagement	5.54**	273	.00
Perceived Learning ^{T}	2.41*	269	.02
Satisfaction	.02	269	.99
Exam Total Grade ^{T}	4.09**	319	.00

^{T}: Transformed scale **: Significant at p<.01 level. *: Significant at p<.05 level

Table 7.5 Collaborative vs. Traditional Exam (Mann-Whitney Test)

Exam Study Strategies and Outcomes	Mann-Whitney U	P
Surface Exam Study	5794.00**	.00
Fairness in Grading	8890.00	.65

**: Significant at p<.01 level.

Therefore:

H2.1: Students taking the collaborative examination will have higher adoption of deep learning in the exam studying process than students taking the traditional exam.
In predicted direction by deep exam study (mean=4.74, 4.51), but not significant
Supported by surface exam study (mean rank= 169.78, 243.75), significant (p<.01)

H2.2: Students taking the collaborative examination will have higher social engagement in the exam studying process than students taking the traditional exam.
Supported (mean=4.41, 3.38), significant (p<.01)

H2.3: Students taking the collaborative examination will have higher perception of learning than students taking the traditional exam.
Supported (mean=24.78, 21.78), significant (p<.05)

<u>H2.4</u>: Students taking the collaborative examination will have higher satisfaction of the exam than students taking the traditional exam.
Not in predicted direction (mean=4.57, 4.57), no significant differences

<u>H2.5</u>: Students taking the collaborative examination will achieve higher grades than students taking the traditional exam.
Supported (mean=2.86, 2.50), significant (p<.01)

<u>H2.6</u>: Students taking the collaborative examination will have lower perception of fairness in grading than students taking the traditional exam.
In predicted direction (mean rank= 186.59, 192.28), but not significant

7.1.2 Exam Study Strategies

This section tests the third sets of hypotheses exploring the relationship between student's learning predispositions and their exam studying strategies (H3.1, H3.2). Additional factorial ANOVA tests are presented in addition to the correlation analysis tests in analyzing hypothesis H3.3 and H3.4.

7.1.2.1 Deep/Surface Exam Study. To examine the relationship between students' deep/surface learning predispositions and their deep/surface exam study strategies, the correlation analysis using Pearson's R was conducted on deep exam study strategy. The Spearman's rho test was conduct on surface exam study, which is not normally distributed. As show in Table 7.6, the learning predispositions are significantly correlated with the exam study strategies. The correlation between deep learning and deep exam study is higher than that of surface learning and surface exam study.

Table 7.6 Deep/Surface Study and Learning Predispositions (Correlation)

	Surface Exam Study [a]	Deep Exam Study
Pre-exam Surface Learning	.232**	-0.424**
Pre-exam Deep Learning	-0.161**	.549**

** Correlation is significant at the 0.01 level (2-tailed)
[a] Spearman's rho is used in stead of Pearson's R in this column

Therefore:

H3.1: Students' adoption of deep learning in the exam studying process will be positively related to their predispositions in deep learning.
Supported by surface learning and surface exam study (Spearman's rho =.232, p<.01)
Supported by deep learning and deep exam study (r =.55, p<.01)

To explore the differences of deep/surface exam study between the traditional and the collaborative exam mode, the correlation between deep exam study and the deep learning predispositions was calculated for the two exam modes, respectively. As the results show in Table 7.7, the correlation between deep exam study and deep learning predisposition is higher in the traditional exam than the collaborative exam. The same was done for the correlation between the surface exam study and the surface learning predisposition, and the results in Table 7.8 are congruent with the conclusions above.

Table 7.7 Deep Study and Deep Orientation in Mode 1 & 3 (Correlation)

		Deep Exam Study
Pre-exam Deep Learning	Traditional	.583**
	Collaborative	.488**

** Correlation is significant at the 0.01 level (2-tailed)

Table 7.8 Surface Study and Surface Orientation in Mode 1 & 3 (Correlation)

		Surface Exam Study [a]
Pre-exam Surface Learning	Traditional	.317**
	Collaborative	.113

** Correlation is significant at the 0.01 level (2-tailed)
[a] Spearman's rho is used in stead of Pearson's R

Therefore:

H3.3: The correlation between students' adoption of deep learning in the exam studying process and their predispositions in deep learning will be higher in the collaborative exam than the traditional exam.
Not supported

<u>Additional Analysis of H3.3</u> To further explore the effect of exam mode on the relationship between the learning predisposition and deep/surface exam study strategy, additional factorial ANOVA tests were conducted to test the interaction effect of exam mode and learning predispositions on exam study strategy. The test answers the following question:

> **<u>Q3.3.a</u>**: Is the difference in the adoption of deep exam study between students who are least deep-oriented and those who are most deep-oriented larger in the collaborative exam than in the traditional exam?

To answer this question, deep learning predisposition as measured by pre-exam survey was first categorized into low, medium, and high. The two extreme categories were used and the medium category was excluded in further analysis. A 2*2 Factorial ANOVA test was then conducted to compare the two exam modes and the two deep learning predispositions. Table 7.9 and 7.10 show the means and the significance test results in Factorial ANOVA. As show in the tables, there is no significant interaction effect of exam mode and learning predispositions on deep exam study strategy. Figure 7.1 shows that the least deep-oriented students had a slightly higher increase in the use of deep exam study strategy than those who were most deep-oriented, and the difference between the least and most deep-oriented students was slightly smaller in the collaborative exam than the traditional exam. However, the differences are not statistically significant.

Table 7.9 Deep Exam Study (Exam Mode*Pre-exam Deep Learning)

Exam Mode	Pre-exam Learning Disposition: Least Deep-oriented	Pre-exam Learning Disposition: Most Deep-oriented	Total
Traditional	3.83 (.96)	5.29 (1.17)	4.47 (1.28)
Collaborative	4.11 (1.16)	5.32 (1.09)	4.75 (1.27)
Total	3.95 (1.05)	5.31 (1.12)	4.60 (1.28)

Table 7.10 Factorial ANOVA of Deep Exam Study

Dependent variable	Source		F	Sig.	Partial Eta^2
Deep Exam Study	Main Effect	Exam Mode (1,3)	.800	.373	.005
	Main Effect	Pre-exam Deep Learning	59.677	.000	.277
	Interaction Effect	Exam Mode * Pre-Deep	.536	.465	.003

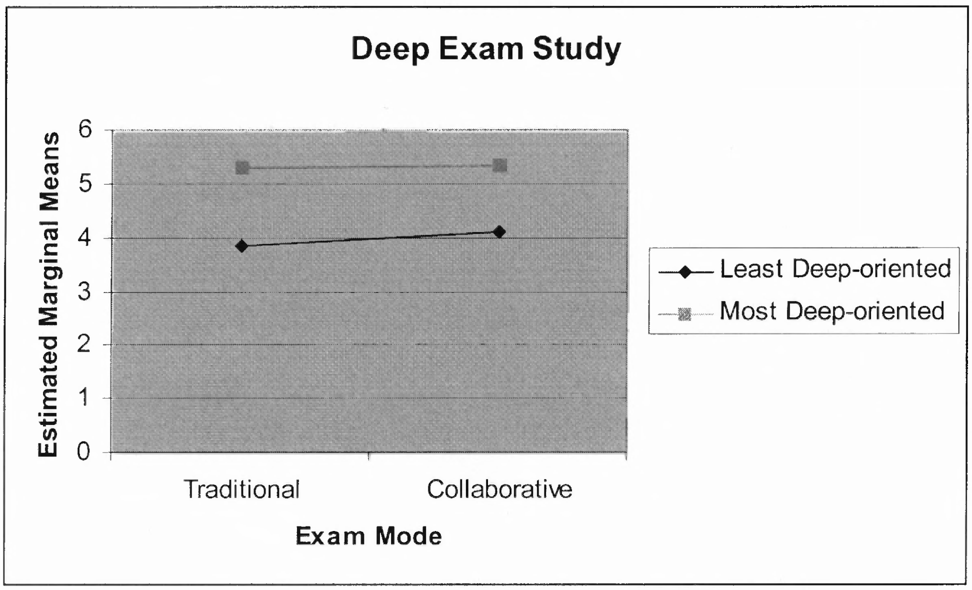


Figure 7.1 Effects of learning predisposition and exam mode on deep exam study.

For investigation purposes, the effects of exam mode and the surface learning predisposition on surface exam study strategy were examined, similar to the procedures above. Surface learning predisposition as measured by pre-exam survey was first categorized into low, medium, and high. The two extreme categories were used and the medium category was excluded in further analysis. Since the surface exam study construct is not normally distributed, the factorial ANOVA test cannot be conducted to test the interaction effect. For analytical purpose, Table 7.11 shows the means in the 2*2 table, and the means are drawn in Figure 7.2. As shown, the most surface-oriented students had a larger drop in the use of surface exam study strategy than the least surface-oriented students (i.e., the slope of the square line is more steep than the diamond line). In addition, the difference in surface exam study between the two learning orientations was smaller in the collaborative exam than the traditional exam.

Table 7.11 Surface Exam Study (Exam Mode* Pre-exam Surface Learning)

	Pre-exam Learning Disposition		
Exam Mode	**Least Surface-oriented**	**Most Surface-oriented**	**Total**
Traditional	3.49 (1.83)	4.59 (0.93)	4.08 (1.51)
Collaborative	2.78 (1.57)	3.26 (1.30)	3.01 (1.46)
Total	3.11 (1.72)	3.98 (1.29)	3.55 (1.58)

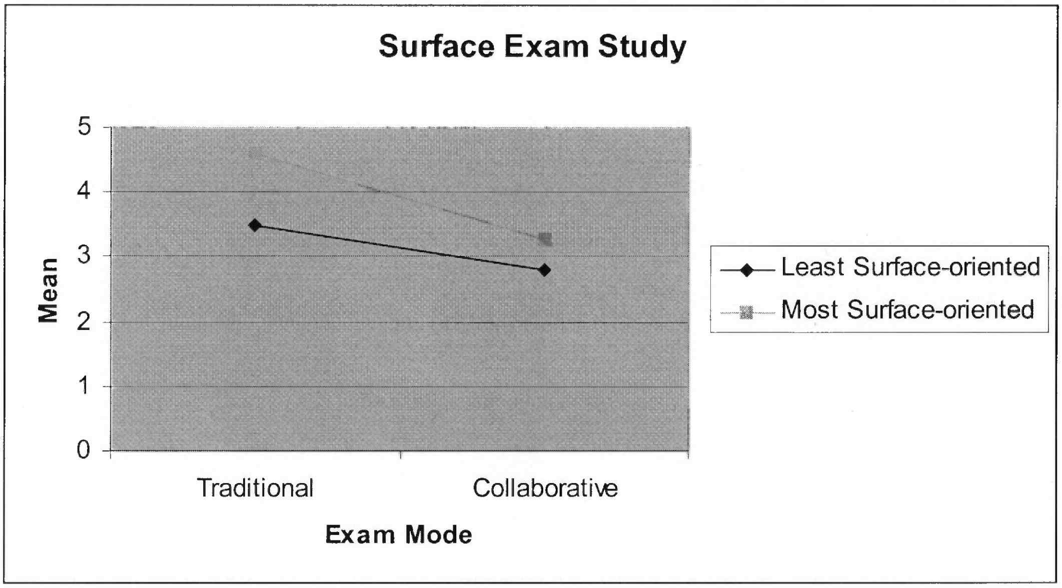


Figure 7.2 Effects of learning predisposition and exam mode on surface exam study.

In summary, the use of deep/surface exam study strategies was significantly correlated to the deep/surface learning predispositions, but the correlations were lower in the collaborative exam than in the traditional exam. The collaborative exam tended to be more effective in decreasing surface exam study and increasing deep exam study among students who were originally the most surface-oriented.

7.1.2.2 Social Engagement. To examine the relationship between students' pre-exam collaborative learning and their social engagement in exam studying, the correlation analysis using Pearson's R was conducted. Result in Table 7.12 shows that the two variables are significantly correlated at r =.43 level.

Table 7.12 Social Engagement and Collaborative Learning (Correlation)

	Social Engagement
Pre-exam Collaborative Learning	.425**

** Correlation is significant at the 0.01 level (2-tailed).

Therefore:

<u>H3.2</u>: Students' level of social engagement in the exam studying process will be positively related to their predispositions in collaborative learning.
Supported (r =.43, p<.01)

To explore the differences of social engagement between the traditional and the collaborative exam mode, the correlation between social engagement and collaborative learning was calculated for the two exam modes, respectively. As the results show in Table 7.13, the correlation is higher in the traditional exam than the collaborative exam.

Table 7.13 Social Engagement and Collaborative Learning in 1 & 3 (Correlation)

		Social Engagement
Pre-exam Collaborative Learning	Traditional	.507**
	Collaborative	.442**

** Correlation is significant at the 0.01 level (2-tailed)

Therefore:

<u>H3.4</u>: The correlation between students' level of social engagement in the exam studying process and their predispositions in collaborative learning will be higher in the collaborative exam than the traditional exam.
Not Supported.

<u>Additional Analysis of H3.4</u> To further explore the effect of exam mode on the relationship between the collaborative learning predisposition and social engagement, an additional factorial ANOVA test was conducted to test the interaction effect of exam mode and collaborative learning predisposition on social engagement in exam study. The test answers the following question:

<u>Q3.4.a</u>: Is the difference in the adoption of social engagement in exam studying between students who are least collaborative-oriented and most collaborative-oriented larger in the collaborative exam than in the traditional exam?

To answer this question, collaborative learning predisposition as measured by pre-exam survey was first categorized into low, medium, and high. The two extreme categories were used, and a 2*2 Factorial ANOVA test was conducted. Table 7.14 and 7.15 show the means and the significance test results in Factorial ANOVA tests. As show in Figure 7.3, the least collaborative-oriented students had a larger increase in the level of social engagement than those who were most collaborative-oriented (i.e., the slope of the diamond line is more steep than the square line). In addition, the difference in social engagement between the two types of students was smaller in the collaborative exam than the traditional exam. However, the interaction effect is not significant.

Table 7.14 Social Engagement (Exam Mode* Pre-exam Collaborative Learning)

	Pre-exam Learning Dispositions		
Exam Mode	**Least Collaborative-oriented**	**Most Collaborative-oriented**	**Total**
Traditional	2.44 (1.12)	4.46 (1.62)	3.42 (1.71)
Collaborative	3.53 (1.48)	4.77 (1.30)	4.16 (1.52)
Total	2.99 (1.42)	4.62 (1.46)	3.80 (1.65)

Table 7.15 Factorial ANOVA of Social Engagement

Dependent variable	**Source**		**F**	**Sig.**	**Partial Eta^2**
Social Engagement	Main Effect	Exam Mode (1,3)	9.608	.002	.061
	Main Effect	Pre-exam Collaborative Learning	52.470	.000	.260
	Interaction Effect	Exam Mode * Pre-Collab	2.960	.087	.019

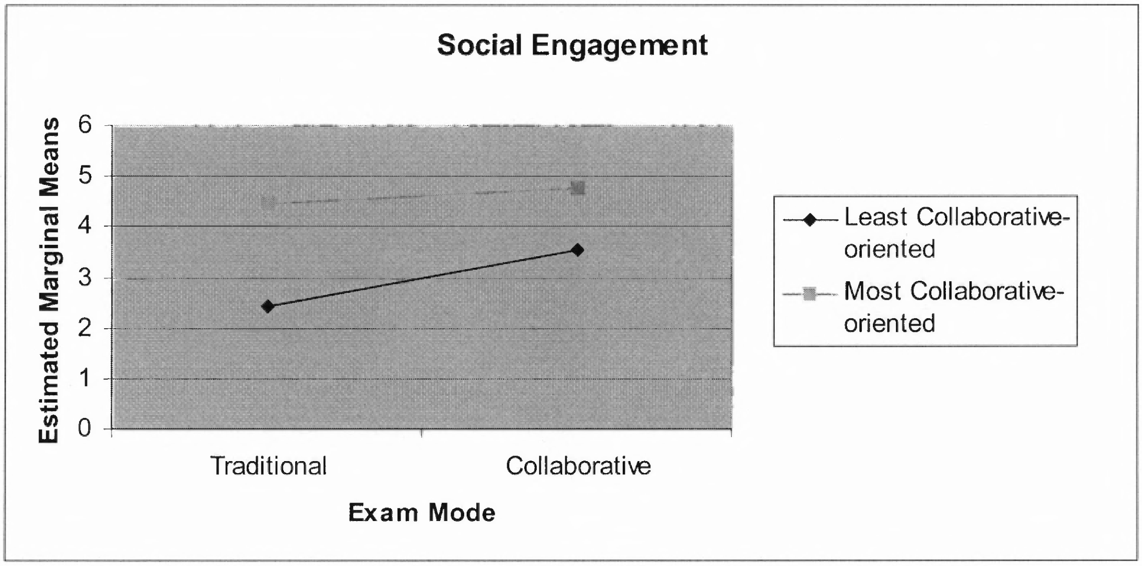


Figure 7.3 Effects of learning predisposition and exam mode on social engagement.

In summary, the level of social engagement in exam studying was significantly correlated to the collaborative learning predisposition, and the correlation was lower in the collaborative exam than in the traditional exam. The collaborative exam tended to be more effective in increasing social engagement in exam studying among students who were originally the least collaborative-oriented.

7.1.3 Course Characteristics

This section tests the fourth set of hypotheses regarding the effect of course characteristics on students' perceptions of the collaborative exam. Therefore, only the data for students who participated in the collaborative exam process are used. Section 7.1.3.1 compares the DL with the blended classes (H4.1). Section 7.1.3.2 compares the undergraduate with the graduate classes (H4.2). Section 7.1.3.3 compares the technical vs. the discussion-oriented courses (H4.3). Section 7.1.3.4 compares the classes with varying class sizes (4.4).

7.1.3.1 Blended vs. DL. To test the differences in students' perceptions of the collaborative exam in blended and DL courses, the T-test was conducted on perceived learning and satisfaction, and the Mann-Whitney test was conducted on perceived fairness in grading. Although no hypothesis was made regarding exam study strategies comparing the two types of courses, tests were also conducted. Results are shown in Table 7.16 and 7.17. As show in the tables, there were some tendency for students in DL courses to have higher perceptions of learning in the collaborative exam than students in the blended courses, and to be less satisfied and have less perception of the fairness in grading, but none of the above difference is significant.

Table 7.16 Collaborative Exam - Blended vs. DL Courses (T-test)

Exam Study Strategies and Outcomes	Course	N	Mean	S. D.	T	df	P
Social Engagement	Blended	54	4.44	1.34	.239	134	.811
	DL	82	4.38	1.50			
Deep Exam Study	Blended	55	4.55	1.24	-1.632	135	.105
	DL	82	4.88	1.12			
Perceived Learning ^{T}	Blended	53	23.59	8.84	-1.104	131	.272
	DL	80	25.58	10.98			
Satisfaction	Blended	53	4.79	1.61	1.229	131	.221
	DL	80	4.43	1.66			

^{T}: Transformed scale

Table 7.17 Collaborative Exam - Blended vs. DL Courses (Mann-Whitney Test)

Exam Study Strategies and Outcomes	Course	N	Mean Rank	Mann-Whitney U	P
Surface Exam Study	Blended	55	79.97	1651.500**	.008
	DL	82	61.64		
Fairness in Grading	Blended	53	70.34	1943.0	.412
	DL	80	64.79		

**: Significant at p<.01 level.

Therefore:

H4.1: DL students will have higher perceptions of the collaborative exam than blended students.
H4.1.1: perceived learning:
In predicted direction (mean=25.58, 23.59), not significant

H4.1.2: satisfaction:
Not in predicted direction (mean=4.43, 4.79), not significant

H4.1.3: perceived fairness in grading:
Not in predicted direction (mean rank=64.79, 70.34), not supported

The tests on exam strategies show that students in DL courses tended to adopt higher levels of deep exam study strategy than those in blended courses (mean=4.88, 4.55), but the differences are not significant. Students in DL courses adopted much less surface exam study strategy than those in blended courses (mean rank=61.64, 79.97), and the difference is significant at .01 level. There was no significant difference in the level of social engagement of students in DL courses or in blended courses (mean=4.38, 4.44) for the collaborative exam. These results indicated that the collaborative exam may be used equally well in blended as well as DL courses.

7.1.3.2 Undergraduate vs. Graduate. To test the differences in students' perceptions of the collaborative exam in undergraduate and graduate courses, a T-test and a Mann-Whitney test were conducted. The same tests were also conducted on exam study strategies. Results are shown in Table 7.18 and 7.19. As show in the tables, students in undergraduate courses tended to have lower perceptions of the collaborative exam than those in graduate courses in all aspects investigated, except for the perceptions of fairness in grading. However, only differences in perceived learning are significant at .05 level.

Table 7.18 Collaborative Exam - Undergraduate vs. Graduate Courses (T-test)

Exam Study Strategies and Outcomes	Course	N	Mean	S. D.	T	df	P
Social Engagement	Undergraduate	63	4.16	1.53	-1.903	134	.059
	Graduate	73	4.62	1.32			
Deep Exam Study	Undergraduate	64	4.48	1.20	-2.547*	135	.012
	Graduate	73	4.98	1.11			
Perceived Learning ^{T}	Undergraduate	61	22.55	10.30	-2.368*	131	.019
	Graduate	72	26.68	9.77			
Satisfaction	Undergraduate	61	4.57	1.84	-.049	113.6	.961
	Graduate	72	4.58	1.46			

^{T}: Transformed scale *: Significant at p<.05 level

Table 7.19 Collaborative Exam - Undergrad vs. Graduate (Mann-Whitney)

Exam Study Strategies and Outcomes	Course	N	Mean Rank	Mann-Whitney U	P
Surface Exam Study	Undergraduate	64	75.73	1905.500	.062
	Graduate	73	63.10		
Fairness in Grading	Undergraduate	61	69.96	2015.500	.412
	Graduate	72	64.49		

Therefore:

H4.2: Undergraduate students will have lower perceptions of the collaborative exam than graduate students.
H4.2.1: perceived learning:
Supported (mean=22.55, 26.68), significant (p<.05)

H4.2.2: satisfaction:
In predicted direction (mean=4.57, 4.58), not significant

H4.2.3: perceived fairness in grading:
Not in predicted direction (mean rank=69.96, 64.49), not significant

The tests on exam strategies show that students in graduate courses adopted higher levels of deep exam study strategy than those in undergraduate courses (mean=4.98, 4.48), and the difference is significant at .05 level. Students in graduate courses also tended to adopt less surface exam study strategies (mean rank= 63.10, 75.73) and to have higher levels of social engagement (mean=4.62, 4.16) than those in undergraduate courses, but both of these are not significant. Therefore, we can conclude that while collaborative exams are especially appropriate for graduate courses that emphasizing deep learning, they can also be used on the undergraduate level as well.

7.1.3.3 Technical vs. Discussion-oriented. To test the differences in students' perceptions of the collaborative exam in technical and discussion-oriented courses, a T-test and a Mann-Whitney test were conducted. The same tests were also conducted on exam study strategies. Results are shown in Table 7.20 and 7.21. As show in the tables, students in technical courses were more satisfied (significant at .05 level) than those in discussion-oriented courses. There is no significant difference in perceptions of learning and perceived fairness in grading in students in discussion-oriented vs. technical courses.

Table 7.20 Collaborative Exam - Technical vs. Discussion Courses (T-test)

Exam Study Strategies and Outcomes	Course	N	Mean	S. D.	T	df	P
Social Engagement	Technical	47	4.43	1.34	.134	134	.894
	Discussion	89	4.40	1.49			
Deep Exam Study	Technical	48	4.48	1.24	-1.960	135	.052
	Discussion	89	4.89	1.12			
Perceived Learning T	Technical	46	23.55	8.94	-1.018	131	.311
	Discussion	87	25.44	10.79			
Satisfaction	Technical	46	4.99	1.55	2.184*	131	.031
	Discussion	87	4.35	1.65			

T: Transformed scale *: Significant at p<.05 level

Table 7.21 Collaborative Exam - Technical vs. Discussion (Mann-Whitney Test)

Exam Study Strategies and Outcomes	Course	N	Mean Rank	Mann-Whitney U	P
Surface Exam Study	Technical	48	81.89	1517.500**	.005
	Discussion	89	62.05		
Fairness in Grading	Technical	46	70.85	1824.000	.399
	Discussion	87	64.97		

**: Significant at p<.01 level

Therefore,

H4.3: Students taking the technical courses will have lower perceptions of the collaborative exam than students taking the discussion-oriented courses.
H4.3.1: perceived learning:
In predicted direction (mean=23.55, 25.44), not significant

H4.3.2: satisfaction:
Reverse effect found (mean=4.99, 4.35), significant (p<.05)

H4.3.3: perceived fairness in grading:
Not in predicted direction (mean rank=70.85, 64.97), not significant

Tests on exam strategies show that the only significant difference is that students in discussion-oriented courses adopted less surface exam study strategies (mean rank= 62.05, 81.89) than students in technical courses (significant at $p<.01$ level). There is no significant difference in the adoption of social engagement or deep exam study strategies. Therefore we can conclude that the collaborative exam can be used in both technical and discussion-oriented courses.

7.1.3.4 Class size. To test the differences in students' perceptions of the collaborative exam in classes of different sizes, the classes of small sizes (1-19 students per class) and classes of large sizes (30 and above per class) were compared. A T-test and a Mann-Whitney test were conducted. The same tests were also conducted on exam study strategies. Results are shown in tables 7.22 and 7.23. There was no significant difference in perceived learning, perceived fairness in grading, or satisfaction

Table 7.22 Collaborative Exam - Class Size (T-test)

Exam Study Strategies and Outcomes	Class Size	N	Mean	S. D.	T	df	P
Social Engagement	Small	53	4.67	1.21	1.134	84	.260
	Large	33	4.36	1.28			
Deep Exam Study	Small	53	5.07	1.03	2.977**	85	.004
	Large	34	4.31	1.33			
Perceived Learning ^{T}	Small	53	26.78	9.10	1.455	84	.149
	Large	33	23.82	9.34			
Satisfaction	Small	53	4.56	1.55	-1.790	84	.077
	Large	33	5.16	1.43			

^{T}: Transformed scale **: Significant at $p<.01$ level

Table 7.23 Collaborative Exam - Class Size (Mann-Whitney Test)

Exam Study Strategies and Outcomes	Class Size	N	Mean Rank	Mann-Whitney U	P
Surface Exam Study	Small	53	37.56	559.500**	.003
	Large	34	54.04		
Fairness in Grading	Small	53	40.80	731.500	.200
	Large	33	47.83		

**: Significant at p<.01 level

Therefore:

> H4.4: Students in small classes will have higher perceptions of the collaborative exam than students in large classes.
> H4.4.1: perceived learning:
> **In predicted direction (mean=26.78, 23.82), not significant**
>
> H4.4.2: satisfaction:
> **Not in predicted direction (mean=4.56, 5.16), not significant**
>
> H4.4.3: perceived fairness in grading:
> **Not in predicted direction (mean=40.80, 47.83), not significant**

Tests on exam strategies show that students in small classes adopted higher levels of deep exam study strategy than students in large classes (mean=5.07, 4.31), and the difference is significant at .01 level. They also adopted much less surface exam study strategies (mean rank= 37.56, 54.04) and the difference is also significant at the .01 level. There is no significant difference in the adoption of social engagement (mean=4.67, 4.36). Therefore, overall, collaborative exams seem appropriate for both small and large classes, though the deep learning strategies are most likely to be adopted in small classes.

7.1.4 Exam Outcomes

This section tests the fifth and the sixth hypotheses on exam outcomes, including exam grades (H5), satisfaction, perceived learning, and perceived fairness in grading (H6).

Additional factorial ANOVA tests are presented in addition to the correlation analysis tests in analyzing hypothesis H6.2-6.5.

7.1.4.1 Grades. This section tests the relationship between grades and students' perceptions of the exam, including satisfaction and perceived fairness in grading. The correlation analysis using Pearson's R was conducted between exam total grade and satisfaction, both of which are normally distributed. Given perceived fairness in grading is not normally distributed, the spearman's rho test was used for the correlation between exam total grade and perceived fairness in grading. As the results show in Table 7.24, both constructs are significantly correlated with exam grades. Yet the correlation is relatively low between perceived fairness and exam grade, and is low between satisfaction and exam grade.

Table 7.24 Exam Grade and Exam Outcomes (Correlation)

		Exam Total Grade [T]
Exam Outcomes	Satisfaction	.194**
	Perceived Fairness in Grading [a]	.356**

[T]: Transformed scale ** Correlation is significant at the 0.01 level (2-tailed)
[a] Spearman's rho is used in stead of Pearson's R in this row

Therefore:

H5.1: Students' satisfaction will be positively related to students' grades.
Supported (r =.192, p<.01)

H5.2: Students' perceived fairness in grading will be positively related to students' grades.
Supported (Spearman's rho=.356, p<.01)

7.1.4.2 Satisfaction. The correlation analysis using Pearson's R was conducted on perceived learning and satisfaction. As shown in Table 7.25, the two constructs are significantly correlated, and the correlation is high at r =.617 level.

Table 7.25 Satisfaction and Perceived Learning (Correlation)

	Satisfaction
Perceived Learning ^{T}	.617**

** Correlation is significant at the 0.01 level (2-tailed).
^{T}: Transformed scale

Therefore:

<u>H6.1</u>: Students' satisfaction with the exam will be positively related to students' perceived learning.
Supported (r =.62, p<.01)

Next, to explore the differences in the correlation between satisfaction and deep exam study in the traditional and the collaborative exam modes, the correlation between satisfaction and deep exam study was calculated for the two exam modes, respectively. As the results show in Table 7.26, the correlation is higher in the traditional exam than the collaborative exam. The correlation is low in the collaborative exam mode.

Table 7.26 Satisfaction and Deep Exam Study in Exam Mode 1 & 3 (Correlation)

		Satisfaction
Deep Exam Study	Traditional	.442**
	Collaborative	.254**

** Correlation is significant at the 0.01 level (2-tailed)

Therefore:

<u>H6.2</u>: The correlation between students' adoption of deep learning in the exam studying process and students' satisfaction with the exam will be higher in the collaborative exam than the traditional exam.
Not Supported.

<u>Additional Analysis of H6.2</u> To further explore the effect of exam mode on the relationship between deep exam study and satisfaction, an additional factorial ANOVA test was conducted to test the interaction effect of exam mode and deep exam study on satisfaction. The test answers the following question:

<u>Q6.2.a</u>: Is the difference in satisfaction between students who adopted the highest level of deep exam study and those who adopted the lowest level of deep exam study larger in the collaborative exam than in the traditional exam?

To answer this question, deep learning as measured in the post-exam survey was first categorized into low, medium, and high. The two extreme categories were used, and a 2*2 Factorial ANOVA test was conducted. 7.27 and 7.28 show the means and the Factorial ANOVA significance test results. As shown, the interaction effect is not significant, even though Figure 7.4 shows the distance between the two lines is slightly smaller in the collaborative exam than the traditional exam.

Table 7.27 Satisfaction (Exam Mode * Deep Exam Study)

	Exam Study Strategy		
Exam Mode	**Least Deep Learning**	**Most Deep Learning**	**Total**
Traditional	3.98 (1.22)	5.24 (1.09)	4.44 (1.32)
Collaborative	4.06 (1.83)	5.09 (1.37)	4.65 (1.65)
Total	4.01 (1.47)	5.15 (1.26)	4.54 (1.48)

Table 7.28 Factorial ANOVA of Satisfaction

Dependent variable	**Source**		**F**	**Sig.**	**Partial Eta^2**
Satisfaction	Main Effect	Exam Mode (1,3)	.04	.848	.000
	Main Effect	Deep Exam Study	29.63	.000	.144
	Interaction Effect	Exam Mode * Deep Exam Study	.33	.567	.002

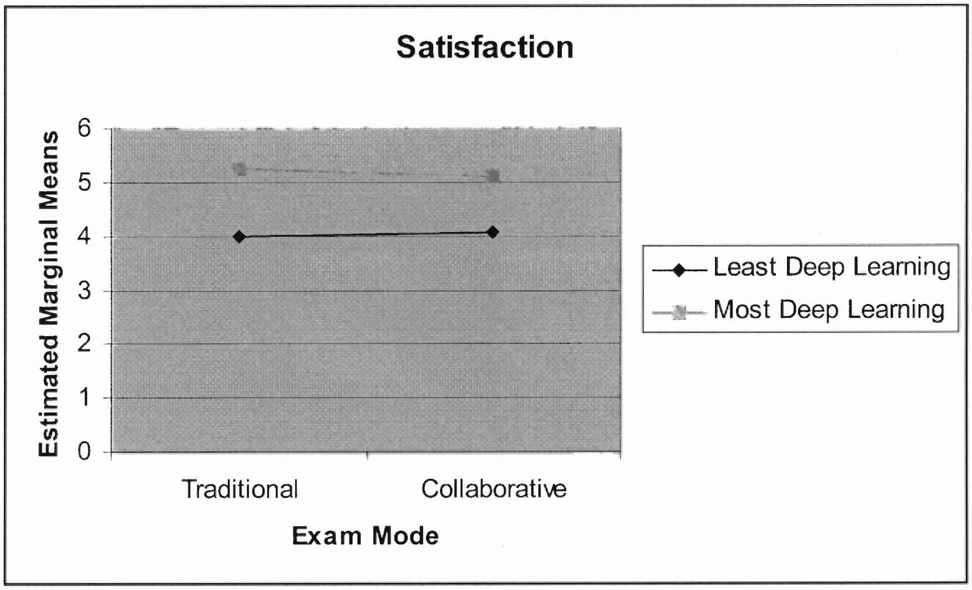


Figure 7.4 Effects of deep exam study and exam mode on satisfaction.

Therefore, the difference in satisfaction between students who adopted the highest level of deep exam study and those who adopted the lowest was slightly smaller in the collaborative exam than in the traditional exam, but the result is not significant.

Similarly, to explore the differences in the correlation between satisfaction and social engagement in the traditional and the collaborative exam mode, the correlation between satisfaction and social engagement was calculated for the two exam modes, respectively. As the results show in Table 7.29, the correlation is higher in the collaborative exam than the traditional exam. The correlation in the traditional exam mode is low.

Table 7.29 Satisfaction and Social Engagement in Mode 1 & 3 (Correlation)

		Satisfaction
Social Engagement	Traditional	.237**
	Collaborative	.496**

** Correlation is significant at the 0.01 level (2-tailed)

Therefore:

> **H6.3**: The correlation between students' level of social engagement in the exam studying process and students' satisfaction with the exam will be higher in the collaborative exam than the traditional exam.
> **Supported.**

Additional Analysis of H6.3 To further explore the effect of exam mode on the relationship between social engagement and satisfaction, an additional factorial ANOVA test was conducted to test the interaction effect of exam mode and social engagement on satisfaction. The test answers the following question:

> **Q6.3.a**: Is the difference in satisfaction between students who were most socially engaged in exam study and those who were least engaged larger in the collaborative exam than in the traditional exam?

To answer this question, social engagement as measured in the post-exam survey was first categorized into low, medium, and high. The two extreme categories were used and a 2*2 Factorial ANOVA test was conducted. Table 7.30 and 7.31 show the means and the Factorial ANOVA significance test results. Figure 7.5 shows that the line representing the least socially engaged students has a drop in satisfaction in the collaborative exam, where the line representing the most socially engaged students stays almost the same in the two exam modes. This indicates that there is a negative synergistic interaction effect where students who were least socially engaged in exam studying were less satisfied with the collaborative exam. The interaction effect is significant at .01 level.

Table 7.30 Satisfaction (Exam Mode*Social Engagement)

Exam Mode	Exam Study Strategy: Least Socially Engaged	Exam Study Strategy: Most Socially Engaged	Total
Traditional	4.43 (1.37)	5.27 (1.09)	4.72 (1.34)
Collaborative	3.06 (1.79)	5.19 (1.25)	4.67 (1.67)
Total	4.08 (1.60)	5.21 (1.19)	4.70 (1.50)

Table 7.31 Factorial ANOVA of Satisfaction

Dependent variable	Source		F	Sig.	Partial Eta^2
Satisfaction	Main Effect	Exam Mode (1,3)	10.667	.001	.057
	Main Effect	Social Engagement	44.721	.000	.203
	Interaction Effect	Exam Mode * Social Engagement	8.526	.004	.046

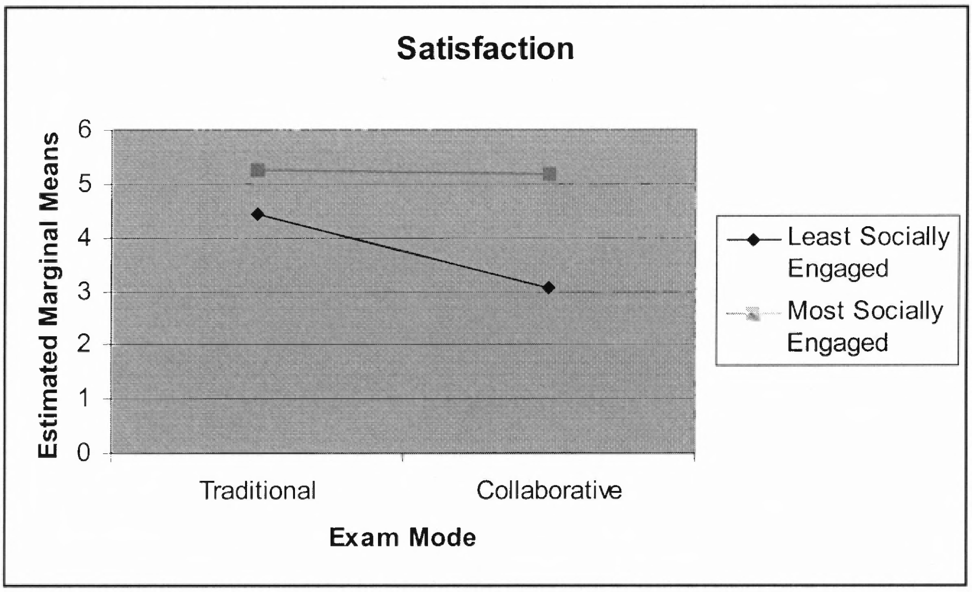


Figure 7.5 Effects of social engagement and exam mode on satisfaction.

Therefore, the difference in satisfaction between students who were most socially engaged in exam study and those who were least engaged was larger in the collaborative exam than the traditional exam, and the result is significant at .01 level.

In summary, the correlation between satisfaction and the adoption of deep exam study was higher in the traditional exam than the collaborative exam. In contrast, the correlation between satisfaction and social engagement was higher in the collaborative exam than the traditional exam. In addition, students who did not engage with other students in exam study were significantly less satisfied with the collaborative exam than those who engaged with others.

7.1.4.3 Perceived Learning. Similarly, to explore the differences in the correlation between perceived learning and deep exam study in the traditional and the collaborative exam mode, the correlation between perceived learning and deep exam study was calculated for the two exam modes, respectively. As the results show in Table 7.32, both correlations are relatively strong, and the correlation is higher in the traditional exam than the collaborative exam.

Table 7.32 Perceived Learning and Deep Exam Study in Mode 1 & 3 (Correlation)

		Perceived Learning ^{T}
Deep Exam Study	Traditional	.593**
	Collaborative	.487**

^{T}: Transformed scale

** Correlation is significant at the 0.01 level (2-tailed)

Therefore:

> <u>H6.4</u>: The correlation between students' adoption of deep learning in the exam studying process and students' perceived learning will be higher in the collaborative exam than the traditional exam.
> **Not Supported.**

Additional Analysis of H6.4 To further explore the effect of exam mode on the relationship between deep exam study and perceived learning, an additional factorial ANOVA test was conducted to test the interaction effect of exam mode and deep exam study on perceived learning. The test answers the following question:

> **Q6.4.a**: Is the difference in perceived learning between students who adopted the least deep learning and the most deep learning higher in the collaborative exam than in the traditional exam?

To answer this question, deep learning as measured in the post-exam survey was first categorized into low, medium, and high. The two extreme categories were used, and a 2*2 Factorial ANOVA test was conducted. Table 7.33 and 7.34 show the means and the Factorial ANOVA significance test results. As shown, the interaction effect is not significant, even though the distance between the two lines is slightly smaller in the collaborative exam than the traditional exam (as shown in Figure 7.6).

Table 7.33 Perceived Learning ^{T} (Exam Mode * Deep Exam Study)

	Exam Study Strategy		
Exam Mode	**Least Deep Learning**	**Most Deep Learning**	**Total**
Traditional	16.23 (8.44)	29.22 (8.91)	21.02 (10.63)
Collaborative	19.02 (8.95)	30.30 (8.29)	25.53 (10.20)
Total	17.28 (8.70)	29.85 (8.52)	23.15 (10.65)

^{T}: Transformed scale

Table 7.34 Factorial ANOVA of Perceived Learning ^{T}

Dependent variable	Source		F	Sig.	Partial Eta^2
Perceived Learning ^{T}	Main Effect	Exam Mode (1,3)	2.179	.142	.012
	Main Effect	Deep Exam Study	85.315	.000	.326
	Interaction Effect	Exam Mode * Deep Exam Study	.422	.517	.002

^{T}: Transformed scale

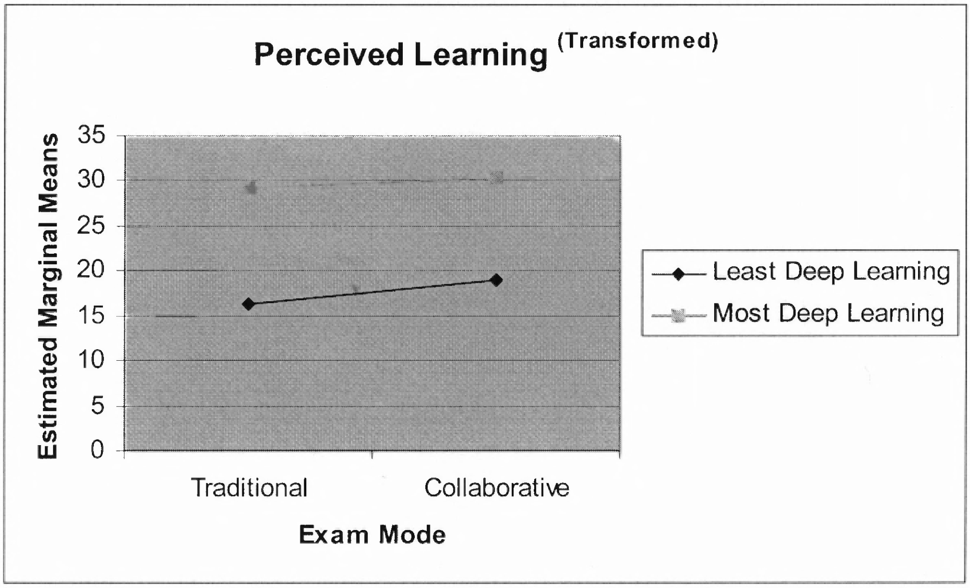


Figure 7.6 Effects of deep exam study and exam mode on perceived learning.

Lastly, to explore the differences in the correlation between perceived learning and social engagement in the traditional and the collaborative exam mode, the correlation between perceived learning and social engagement was calculated for the two exam modes, respectively. As the results show in Table 7.35, the correlation is higher in the collaborative exam than the traditional exam. The correlation is low in the traditional mode.

Table 7.35 Perceived Learning and Social Engagement in Mode 1 & 3 (Correlation)

		Perceived Learning [T]
Social Engagement	Traditional	.369**
	Collaborative	.574**

** Correlation is significant at the 0.01 level (2-tailed)
[T]: Transformed scale

Therefore:

<u>H6.5</u>: The correlation between students' level of social engagement in the exam studying process and students' perceived learning will be higher in the collaborative exam than the traditional exam.
Supported.

<u>**Additional Analysis of H6.5**</u> To further explore the effect of exam mode on the relationship between social engagement and perceived learning, an additional factorial ANOVA test was conducted to test the interaction effect of exam mode and social engagement on perceived learning. The test answers the following question:

<u>Q6.5.a</u>: Is the difference in perceived learning between students who are most socially engaged in exam study and those who are least engaged higher in the collaborative exam than in the traditional exam?

To answer this question, social engagement as measured in the post-exam survey was first categorized into low, medium, and high. The two extreme categories were used and a 2*2 Factorial ANOVA test was conducted. Table 7.36 and 7.37 show the means and the Factorial ANOVA significance test results. Figure 7.7 shows that the least socially engaged have a drop in perceived learning in the collaborative exam, while the most socially engaged stay almost the same in the two exams. This indicates that there is a tendency towards negative synergistic interaction effect where students who were least socially engaged in exam studying had significantly less perceptions of learning in the collaborative exam, but the interaction effect is not significant.

Table 7.36 Perceived Learning ^{T} (Exam Mode * Social Engagement)

Exam Mode	Exam Study Strategy		Total
	Least Socially Engaged	Most Socially Engaged	
Traditional	19.37 (11.58)	28.60 (8.66)	22.61 (11.48)
Collaborative	15.87 (14.92)	29.33 (8.20)	26.04 (11.70)
Total	18.48 (12.51)	29.08 (8.32)	24.25 (11.68)

^{T}: Transformed scale

Table 7.37 Factorial ANOVA of Perceived Learning ^{T}

Dependent variable	Source		F	Sig.	Partial Eta^2
Perceived Learning ^{T}	Main Effect	Exam Mode (1,3)	.642	.424	.004
	Main Effect	Social Engagement	43.023	.000	.196
	Interaction Effect	Exam Mode * Social Engagement	1.506	.221	.008

^{T}: Transformed scale

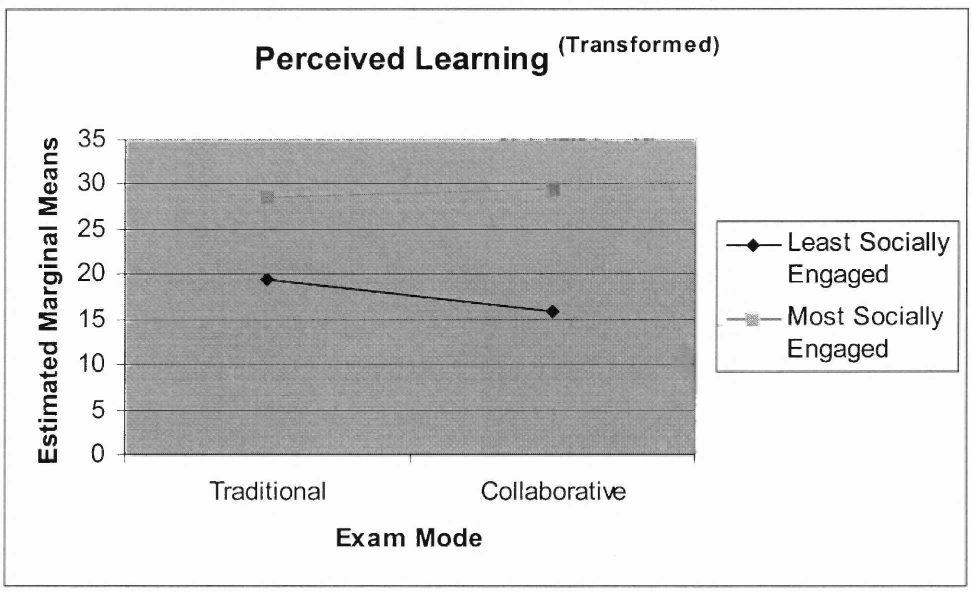


Figure 7.7 Effects of social engagement and exam mode on perceived learning.

In summary, the correlation between perceived learning and the adoption of deep exam study was higher in the traditional exam than the collaborative exam. In contrast, the correlation between perceived learning and social engagement in exam study was higher in the collaborative exam than in the traditional exam. Students who did not engage with other students in exam study tended to have less perceptions of learning in the collaborative exam than those who engaged with others, but the effect is not significant.

7.2 Additional Analysis on Questions in the Online Exams

This section discusses the additional 22 questions in the questionnaires for the two online exam modes. Section 7.2.1 provides students' perceptions of the anonymity of the online exam processes. Section 7.2.2 compares students' responses to the additional questions regarding perceived learning and perceived fairness in grading in the two online exam modes. Section 7.2.3 analyzes the group process in the collaborative exam and its correlations with the exam study strategy and outcomes.

7.2.1 Anonymity

Given the participatory and collaborative exams were conducted anonymously using IDs, students in these two exams were asked about their perceptions of the anonymity mechanism. As the results show in Table 7.38-7.40, the majority of students did not guess the person who authored or graded the questions, did not share their IDs with others, and liked the anonymity in the online exams. Chi-square tests were conducted and there is no significant difference in these responses between the participatory and the collaborative exam modes.

Table 7.38 Anonymity 1- Identity Guess

		Did you try to guess the person who authored or graded the question?		Total
		Yes	No	
Exam Mode	Participatory	24 21.8%	86 78.2%	110 100%
	Collaborative	23 16.9%	113 83.1%	136 100%
Total		47 19.1%	199 80.9%	246 100%

Table 7.39 Anonymity 2–ID Sharing

		Have you shared your question IDs with anyone else?		Total
		Yes	No	
Exam Mode	Participatory	8 7.3%	102 92.7%	110 100%
	Collaborative	13 9.6%	123 90.4%	136 100%
Total		21 8.5%	225 91.5%	246 100%

Table 7.40 Anonymity 3- Anonymity Preference

		I like the anonymity in the exam		Total
		Yes	No	
Exam Mode	Participatory	98 89.1%	12 10.9%	110 100%
	Collaborative	120 88.2%	16 11.8%	136 100%
Total		218 88.6%	28 11.4%	246 100%

Students in the collaborative exam were also asked whether their online group communication was anonymous, and whether they liked it. Given most of the exam groups were project teams where students already knew each other, anonymity in the group online communication was optional. 75% of students reported that they did not communicate with team members anonymously. Two thirds of the remaining students reported that they liked the group online communication as anonymous.

7.2.2 Addition Comparison of Participatory and Collaborative Exams

Eight additional questions were included in the participatory and collaborative exam questionnaires to investigate the perceived learning and the perceived fairness in grading. Principle component factor analysis with PROMAX rotation was conducted on the eight items. Two factors were found and one item (POSTQ327) was eliminated for low loading onto the factors. The final factor analysis with the seven remaining items extracted the same two factors, which are labeled as perceived learning additional items; and perceived fairness in grading additional items. The levels of internal consistency of the factors were measures using Cronbach's alpha, and both are above .7. Table 7.41 shows the items in the perceived learning additional items scale, and Table 7.42 shows the items in the perceived fairness in grading additional items scale. A T-test was conducted to compare the differences between the participatory and the collaborative exams, and the results are also shown in the tables. As the results show, there is no significant difference in perceptions of learning and perceived fairness in grading between students who took the collaborative exam vs. the participatory exam. These results are congruent with the analysis using the main perceived learning and perceived fairness in grading constructs in Chapter 6.

Table 7.41 Perceived Learning Additional Items: P vs. C Exam (T-test)

Item No.	Item [a]	Exam Mode	N	Mean	S. D.	T	df	P
POSTQ322R	I did NOT learn from making up questions [R]	P	109	5.39	1.50	-.196	240	.85
		C	133	5.44	1.75			
POSTQ323	I learned from reading other's answers	P	109	5.04	1.49	-.387	240	.70
		C	133	5.11	1.55			
POSTQ324R	I did NOT learn from grading students' answers [R]	P	109	5.15	1.53	-.340	240	.73
		C	133	5.22	1.69			
POSTQ325	I learned from reading grading posted online	P	109	4.75	1.42	.421	240	.67
		C	133	4.67	1.62			

Cronbach's alpha=.81

P: participatory Exam; **C**: Collaborative Exam
[R] Reversed Item
[a]: All items were measured using the same scale from Strongly Disagree (1) to Strongly Agree (7)

Table 7.42 Perceived Fairness in Grading Additional Items P vs. C Exam (T-test)

Item No.	Item [a, b]	Exam Mode	N	Mean	S. D.	T	df	P
POSTQ326R	I do NOT believe CIS 675 students were able to design questions of good quality for the exam [R]	P	109	5.13	1.55	.78	240	.44
		C	133	4.96	1.73			
POSTQ328R	I do NOT believe CIS675 students were capable of grading the responses to the questions they designed [R]	P	109	4.71	1.56	.39	240	.70
		C	133	4.62	1.73			
POSTQ329R	It would have been an improvement if only professor graded [R]	P	109	3.90	1.69	.21	240	.83
		C	133	3.85	1.87			

Cronbach's alpha=.74

P: participatory Exam; **C**: Collaborative Exam
[R] Reversed Item
[a]: The course number "CIS675" is shown as an example and it is customized for each course.
[b]: All items were measured using the same scale from Strongly Disagree (1) to Strongly Agree (7)

7.2.3 Analysis of Group Process in the Collaborative Exam

Ten questions were included in the collaborative exam questionnaire to investigate the group process. Principle component factor analysis with PROMAX rotation was conducted on the ten items. Three items (POSTQ210R, POSTQ211, POSTQ216) were eliminated for low loading onto the factors. The final factor analysis with the seven remaining items was conducted, and they all loaded onto one factor, which is labeled as group process. The levels of internal consistency of the factor was measured using Cronbach's alpha, it is at .91. Table 7.43 shows the items in the group process scale with means and the standard deviation.

Table 7.43 Group Process

Item No.	Item		Mean	SD
POSTQ27	Solution Satisfaction	How satisfied or dissatisfied were you with the quality of your group's questions and grades? Very Dissatisfied 1 –7 Very Satisfied	5.46	1.62
POSTQ28	Involvement 1	There was a high degree of participation on the part of members. Strongly Disagree1-7 Strongly Agree	5.16	1.83
POSTQ212	Involvement 2	Members worked together as a group. Strongly Disagree 1-7 Strongly Agree	5.04	1.89
POSTQ29	Discourse Quality	The issues discussed during the group's discussions were: Trivial 1-7 Substantial	4.69	1.66
POSTQ213	Trust	To what extent did you trust the members in your group? No trust at all 1-7 Great deal of trust	5.16	1.75
POSTQ214	Process Gain	The group process uncovered valid alternatives that I had not considered. Strongly Disagree 1-7 Strongly Agree	4.38	1.85
POSTQ215	Cohesiveness	To what extent were the people in your group helpful? Not helpful at all 1-7 Very helpful	5.08	1.72

Cronbach's alpha=.91

To analyze the effect of group process on exam study strategies and outcomes, correlation analysis was conducted. Since the individual item scores are not normally distributed, correlation analysis was conducted using spearman's rho. Table 7.44 shows that in terms of exam study strategies, significant positive correlations are found between all aspects of the group process and social engagement, as well as deep exam study strategy. The correlations range from .34 to .54, and they are significant at .01 level. This suggests that the better the group process, the better use of exam strategies (deep exam study, social engagement). In terms of exam outcomes, significant positive correlations are found between all aspects of group process and perceptions of learning and satisfaction. In addition, perceived fairness in grading is positively correlated to three aspects (items) of the group process (solution satisfaction, involvement 1, and cohesiveness). These correlations suggest that the better the group process, the better perceptions of the exam outcomes.

Table 7.44 Group Process and Exam Study and Outcomes (Correlation)

Item	Social Engagement	Deep Exam Study	Surface Exam Study	Perceived Learning	Satisfaction	Perceived Fairness in Grading
Solution Satisfaction	.357**	.339**	.014	.500**	.431**	.422**
Involvement 1	.449**	.381**	.063	.370**	.282**	.263**
Involvement 2	.525**	.460**	.046	.530**	.392**	.139
Discourse Quality	.423**	.427**	.040	.383**	.436**	.095
Trust	.380**	.341**	.057	.404**	.326**	.143
Process Gain	.509**	.462**	.105	.565**	.329**	.095
Cohesiveness	.536**	.446**	.180*	.517**	.419**	.264**

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

CHAPTER 8

LOG AND QUALITATIVE DATA ANALYSIS

This chapter investigates research questions 7 through 9 using the log data and qualitative interview data. Research questions 7 and 8 are first investigated quantitatively using student log data in Section 8.1, and qualitatively using student interview data in Section 8.2.1. Given the small number of faculty (only one) who recorded the instructor log, and the breath and depth of information collected from the instructor interviews, research question 9 is explored qualitatively using interviews in Section 8.2.2.

8.1 Student Log Data Analysis

This section answers research questions 7 and 8 regarding students' use of mobile devices and the time and effort they spent in exam study. The following three types of information were collected in student logs:

1. Exam study length – how long was the study activity
2. Exam study scope – what materials were covered in the study
3. Exam study medium – what medium and devices were used during the study

Exam study length was recorded as the number of hours spent daily by students in exam study. Exam study scope dealt with whether the study covered required exam materials (i.e., exam materials), such as textbooks, lecture notes, required readings, etc., or materials not explicitly required for the exam (i.e., additional materials), such as additional readings, etc. In addition to the time and scope of exam study, students were also asked what medium they used in exam study: traditional paper-and-pen (e.g., textbook, notes, papers), the Internet, or mobile devices (e.g., mobile phone, PDAs, notebook). In addition to the above three aspects, students were asked to indicate whether the exam study activity was an individual and/or group process, and were asked to fill out the above three aspects for the individual and group exam study activities, respectively.

A total of 1848 valid log entries were recorded in the online log database, which is referred to as the *log entry database* in this chapter. Before further analysis, the log entry database was checked for data consistency and was treated in the following steps:

1. Transformed different time format into the hour format (e.g. changed "15 minutes" into .25 hour). The time field was then transformed from a *string* variable to a *number* variable to allow further calculation and analysis.

2. Treated time outliers. The time spent on daily exam study as students reported initially ranged from .05 to 300 hours. Given students were instructed to fill out the log daily, the maximum time spent in exam study per day should not be more than 24 hours. In some cases, given the multiple log records filled out by the same student, the number was treated as minutes rather than hours (e.g., a student persistently used the "xxx minutes" format in the time field in all records but one where s/he put in "30". Then the number 30 was treated as minutes and was transformed to .5). After the above transformation process, there were still 18 records where the students reported the study time as more than 15 hours in the daily log. These 18 time fields were regarded as outliers and were eliminated in further analysis.

3. Consolidated records by students. To calculate the total time students spent in exam study, the 1848 log records were aggregated in SPSS by student. This results in 416 student records, indicating that a total of 416 students filled out the log. The 416 records were saved into a new file which is referred to as the *student log database* in this chapter.

8.1.1 Descriptive Statistics

This section analyzes the *student log database*. Individual exam study activities are first analyzed and presented, followed by group exam study activities. For each of the two types of exam activities, the following three aspects are presented: exam study length, scope, and medium.

As indicated previously, a total of 416 student records were extracted into the *student log database.* Table 8.1 summaries the basic information of the student log database. On average each student recorded 4.44 log entries, with a minimum of one entry and a maximum of 23 entries. Among the 416 students, 411 reported at least one of the exam activities as individual exam study, with a maximum of 22 times and an average of 4.27 times. 139 students recorded at least one of the exam studies as group activity, with a maximum of 12 times and an average of 2.09 times.

Table 8.1 Consolidated Student Log Database

Student Record	N	Mean	S. D.	Min	Max
Log Entry	416	4.44	3.97	1	23
Individual Study	411	4.27	3.77	1	22
Group Study	139	2.09	1.87	1	12

Table 8.2-8.4 provide more details of the individual exam study activities from the 411 students who reported individual exam activities. In terms of the time spent in the exam, 397 students reported at least once the number of hours they spent for the exam. On average, students reported spending 9.78 hours in the exam with a wide range from .05 hours to 86 hours total. Comparing the three exam modes, students in the collaborative exam reported spending the most amount of hours in the exam (mean=11.23 hours) while students in the traditional exam reported the least (mean=7.83 hours). Given the fact that students in the traditional exam were asked to record the exam log till the date of the exam, they did not report the time they spent in taking the traditional exam (one to three hours). When the in-class exam hours are added to the

exam hours, there is no significant difference among the three exam modes in terms of the number of hours spent in the exam.

Table 8.2 Individual Exam Study –Length

Individual Study - Length	Exam	N	Mean	S. D.	Min	Max
Exam Hours	T	125	7.83*	9.01	0.05	55.5
	P	124	10.02	11.56	0.25	86
	C	148	11.23	11.56	0.25	68.9
	Total	397	9.78	10.89	0.05	86

*: Exam hours for the traditional exam do not include the one to three hour in-class exam time.

In terms of the content of exam study, students were asked whether they studied materials explicitly required for the exam (i.e., exam materials), or materials not explicitly required for the exam (i.e., additional materials). Table 8.3 provides the results from students who recorded individual exam activities. Given students reported on average individual study activity 4.27 times (Table 8.1), the result suggests the majority of the time students studied materials explicitly required for the exam (mean=3.39 times), and sometimes they studied additional materials (mean= 2.12 times). Students in the two online exams reported studying exam materials more frequently (P=3.62, C=3.81) than students in the traditional exam (T=2.6 times). The ANOVA test and post-hoc analysis reveal the differences between both online exams and the traditional exam are significant. Students in the two online exam modes also reported slightly higher frequency in using additional materials, but the difference is not significant. The result suggests the online exams resulted in students studying the required exam materials and the additional exam materials more frequently than the traditional exam, and the difference in the required exam materials is significant.

Table 8.3 Individual Exam Study –Materials

Individual Study - Materials	Exam	N	Mean	S. D.	Min	Max	F	Sig.
Exam Materials	T	110	2.60	1.94	1	11	7.21**	.001
	P	119	3.62	2.79	1	14		
	C	144	3.81	2.96	1	17		
	Total	373	3.39	2.69	1	17		
Additional Materials	T	44	1.86	1.30	1	7	1.16	.317
	P	72	2.31	1.61	1	9		
	C	84	2.08	1.59	1	8		
	Total	200	2.12	1.54	1	9		

**: Significant at p<.01 level

In terms of the medium students used in exam study, Table 8.4 shows the results from students who recorded individual exam activities. Given students reported on average individual study activities 4.27 times (Table 8.1), most frequently they used the Internet in their exam study (mean=3.23), and sometimes they used paper-and-pen (mean=2.58) and mobile devices (mean=2.08). The ANOVA test reveals that there is no significant difference in the use of paper-and-pen and mobile devices in their exam study among students in three exam modes. However, students in the two online exams used the Internet more frequently (P=3.37, C=4.09) in exam study than students in the traditional exam (T=1.90). This suggests the online exams resulted in students significantly using the Internet more in exam study than the traditional exam. Although the result may seem trivial given the fact that the online exams required students to use the Internet to participate in the exam process, a closer look revealed that the online exams encouraged students to use the Internet in searching for additional materials and in understanding course materials and concepts. See more discussions in the interview section.

Table 8.4 Individual Exam Study –Medium

Individual Study - Medium	Exam	N	Mean	S. D.	Min	Max	F	Sig.
Paper-and-Pen	**T**	99	2.21	1.66	1	8	2.68	.08
	P	89	2.74	2.10	1	10		
	C	94	2.81	2.12	1	10		
	Total	282	2.58	1.97	1	10		
Internet	**T**	98	1.90	1.39	1	8	15.8**	.000
	P	108	3.37	2.71	1	18		
	C	133	4.09	3.84	1	21		
	Total	339	3.23	3.07	1	21		
Mobile Devices	**T**	14	1.71	1.20	1	5	.174	.841
	P	17	2.12	1.83	1	6		
	C	21	2.29	4.00	1	18		
	Total	52	2.08	2.79	1	18		

**: Significant at p<.01 level

Next the group exam study activities are examined. Table 8.5 -8.7 provide the information the 139 students who reported group exam study activities. In terms of time spent in group exam study, 128 students recorded the time and on average, students reported spending 3.15 hours in the exam with a range from .05 hours to 26.5 hours. The ANOVA test reveals that there is no significant difference in the amount of time students spent in group exam study. This is unexpected considering that students in the collaborative exam were supposed to collaborate with other students and therefore to spend more time in group study. Several factors may have contributed to this result. One primary factor is that only a small percentage of students from the three exam modes kept a log of the group exam study activities (24, 18, 86 who recorded group activity logs compared with 235, 152, and 199 who participated in the three exam modes). The low participation rate in group exam logs reduced the accuracy of the group log. Another important factor reducing the accuracy of the log may be the fact that it was a self-reported log. For example, using the answers collected from the group activity purpose field, it is obvious that students in the participatory exam regarded a wide range of activities as exam group activities, including taking classes and working on projects together (which are not part of the exam).

Table 8.5 Group Exam Study - Length

Group Study - Length	Exam	N	Mean	S. D.	Min	Max
Exam Hours	T	24	2.39*	2.96	0.05	14
	P	18	4.30	6.28	0.25	26.5
	C	86	3.13	3.63	0.08	20.08
	Total	128	3.15	4.00	0.05	26.5

*: Exam hours for the traditional exam do not include the one to three hour in-class exam time.

In terms of the content of the group exam study, Table 8.6 shows that on average students used materials explicitly required for the exam 1.53 times and used additional materials 1.23 times. Given students reported on average group study activity 2.09 times (Table 8.1), the result is similar to the individual exam study pattern, where the majority of the time student groups studied required materials and sometimes they studied additional materials. The ANOVA test reveals that there is no significant difference among the three exam modes in this regard.

Table 8.6 Group Exam Study - Materials

Group Study - Materials	Exam	N	Mean	S. D.	Min	Max	F	Sig.
Exam Materials	T	18	1.22	0.55	1	3	1.650	.198
	P	15	1.47	0.74	1	3		
	C	57	1.65	0.99	1	5		
	Total	90	1.53	0.89	1	5		
Additional Materials	T	8	1.25	0.71	1	3	.156	.856
	P	6	1.33	0.82	1	3		
	C	12	1.17	0.39	1	2		
	Total	26	1.23	0.59	1	3		

Table 8.7 shows the results of the medium students used in group exam study. Given students reported on average group study activity 2.09 times (Table 8.1), almost all the time they used the Internet in group exam study (mean=2.04), frequently they used mobile devices (mean=1.53), and sometimes they used paper-and-pen (mean=1.36). Compared with the medium used in individual exam study, Internet and mobile devices are used more frequently in group study. The ANOVA test reveals that the only significant difference in the use of the medium is that students in the traditional exam significantly used mobile devices more than those in the participatory exam. This may be due to the fact that students in traditional exams are more likely to use mobile communication methods (e.g. mobile phones) to coordinate synchronous face-to-face group study meetings where students in the online exams are more likely to use the Internet (e.g., web conferences or emails) to communicate asynchronously.

Table 8.7 Group Exam Study - Medium

Group Study - Medium	Exam	N	Mean	S. D.	Min	Max	F	Sig.
Paper-and-Pen	**T**	16	1.19	0.54	1	3	.466	.631
	P	11	1.45	0.69	1	3		
	C	18	1.44	1.15	1	5		
	Total	45	1.36	0.86	1	5		
Internet	**T**	13	1.54	1.20	1	5	1.616	.205
	P	10	1.40	0.70	1	3		
	C	62	2.24	1.94	1	10		
	Total	85	2.04	1.76	1	10		
Mobile Devices	**T**	7	2.14	1.46	1	5	3.506*	.042
	P	6	1.00	0.00	1	1		
	C	23	1.48	0.59	1	3		
	Total	36	1.53	0.84	1	5		

*: Significant at p<.05 level

8.1.2 Investigating Research Questions

After providing the general information of student logs, this section investigates research questions 7 and 8 using log data. Research question 7 is concerned about the use of mobile computing devices and its impact on students' learning in exam study. Each sub question will be discussed in turn in this section.

Q7.1: What mobile computing devices do students use in exam studying?

To answer this question, log entry database was examined in terms of the mobile devices students specified and the results are shown in Table 8.8. In individual exam study, 102 log records contain valid information on the mobile devices used in the exam study. The most frequently used mobile device was notebook (70 times), followed by mobile phone (22 times). The two were also used together sometimes (8 times). Tablet PC was mentioned, and used together with mobile phones. In group exam study, 53 log entries contained valid mobile device information. The most frequently used mobile device was mobile phone (40 times), and notebook took second place (10 times). The two were also used together sometimes, and tablet PC was also mentioned. Results suggest that notebooks and mobile phones are the most frequently used mobile devices in exam study, where notebooks are most often used in individual exam study and mobile phones are most often used in group exam study.

Table 8.8 Mobile Devices Used in Exam Study (Frequency)

Mobile Devices	Individual Exam Study	Group Exam Study
Notebook	70	10
Mobile Phone	22	40
Mobile Phone & Notebook	8	2
Tablet PC	1	1
Mobile Phone & Tablet PC	1	0
Total	102	53

Q7.2: For what purpose do students use mobile computing devices in exam studying?

To answer this question, log entry database was examined regarding the purpose of using the mobile devices. A total of 96 individual log entries and 45 group log entries contain valid responses. The different purposes are grouped into five categories based on their common characteristics:

1. Communication/Discussion: communicate with other students, coordinate meetings, discuss with other students
2. Access to the Internet: connect to the Internet, browse the Internet

3. Documentation: write down exam questions/answers/grades/disputes
4. Exam study: study for the exam, such as review lectures, organize concepts of the course, etc.
5. Following exam procedures: such as read exam procedures, submitting answers online, etc.

Table 8.9 provides the frequency of using mobile devices for each of the five purposes in individual and group exam study. As shown, communication and discussion with other students is the primary reason for using mobile devices in both individual and group exam studies. Access to the Internet is the second most frequent reason for using mobile devices. Writing down exam answers, questions, etc., is the third highest reason for using mobile devices, followed by exam study and following exam procedures.

Table 8.9 Purpose of Using Mobile Devices (Frequency)

Purpose of Using Mobile Devices	Individual Exam Study	Group Exam Study
Communication/Discussion with Other Students	34	35
Access Internet	23	3
Documentation	20	2
Exam Study	13	2
Following Exam Procedures	6	3
Total	96	45

Q7.3: How frequently do students use mobile computing devices in exam studying?

As discussed in the previous section on the analysis of student log database, on average each student recorded individual exam activity 4.27 times (Table 8.1), and they reported using mobile devices 2.08 times (Table 8.4). On average, each student recorded group exam activity 2.09 times (Table 8.1), and they reported using mobile devices 1.53 times (Table 8.7). This suggests that mobile devices are used frequently in exam studies, and the frequency is higher in group exam study (1.53/2.09= 73.2%) than in individual exam study (2.08/4.27= 48.7%). In comparison, students reported using the Internet and paper-and-pen more often than mobile devices in individual exam study (Table 8.4). In group exam study, the use of Internet also surpassed mobile devices, but the use of mobile devices surpassed paper-and-pen (Table 8.7).

Q7.4: Does the use of mobile computing devices affect students' adoption of deep learning in exam studying?

As shown in Table 8.9, students mentioned using mobile devices in direct exam study 13 out of the 96 times when mobile devices were used in individual exam study and 2 out of 45 times in group exam study. Specifically, students reported using mobile devices (including notebooks, mobile phones, Tablet PCs) to read lecture slides, search for materials, use online dictionary, make notes, organize key concepts, testing programming ideas, and coding. The easy access to the Internet, the ability to write down exam answers/questions whenever the student wants to, and the ability to easily get connected to other students to discuss exam issues together all appear to contribute to students' learning. Although the limited data does not warrant a conclusive answer to this question, the result suggests that the convenience of anywhere, anytime computing enhanced students' access to information on the Internet and connectivity with other students, which have contributed to students' learning in exam study.

<u>Q7.5</u>: Does the use of mobile computing devices affect students' social engagement in exam studying?

As shown in Table 8.9, the most frequently reported reason for students to use mobile devices in exam study was to communicate and discuss with other students. This is especially the case in group exam study, where 35 out of 45 times students reported using mobile devices to communicate with other students in group exam study. More specifically, students reported using both synchronous (e.g., mobile phones) and asynchronous (e.g., email, posting in web conferences) communications to connect with other students in order to coordinate meetings, keep in touch with team members, reminding each other of deadlines, emailing and posting on Webboard or WebCT to discuss exam question and answers, etc. Although the limited data does not warrant a conclusive answer to this question, the result suggests that the convenience provided by mobile devices allows students to be easily connected anytime, anywhere, which contributed to the level of social engagement in exam study.

<u>Q8</u>. Do students spend more or less time and effort in preparing for, studying, and taking the online exam than the traditional exam?

The time aspect of exam study was examined using the student log data with the number of hours students reported in exam study. The effort aspect of exam study was examined using the scope of students' exam study considering whether students studied required exam materials or additional materials. As shown in Table 8.2, in individual exam study, students in the collaborative exam tended to spend more time in exam study than the traditional exam (T=10.83, C=11.23) after adjusting the three hours students spent in the traditional exam, but the difference is not significant. Students in the two online exam modes spent more effort (Table 8.3) in exam study in terms of more frequently studying required exam materials (T=2.60, P=3.62, C=3.81) and additional materials (T=1.83, P=2.31, C=2.08). The difference in required exam materials is significant. Overall, the results suggest that students spent more time and effort in studying for the online exams than the traditional exam.

8.2 Student and Instructor Interview Data Analysis

A total of 29 interviews were conducted with students selected from the online participatory and collaborative exams. Selected interviewees included students who had high and low perceptions of the exam based on the post exam survey data. Students who received high scores and low scores in the exam were also selected. Actual selection of interviewees was inevitably influenced by students' availability.

27 student interviews were conducted either face to face or over the phone, and were tape recorded. One student interview was conducted via Instant Messaging due to the student's preference, and another student interview was answered via email when the student was overseas on business. All the professors conducting the online exams were interviewed after the exams, and the interviews were tape recorded. Pre-exam interviews with professors were conducted and notes were taken. Table 8.10 provides the course information of the student interviewees with pseudo names used in this section.

Table 8.10 Students Selected for Interviews

2004	Participatory Exam	Collaborative Exam	Total
Spring	CIS675-BLD: Student A, Student B	MIS635-DL: Student M, Student N	4
Summer	MIS635-DL& BLD: Student C, Student D	CIS675-DL: Student O, Student P	
		CIS390-DL: Student Q, Student R	6
Fall	CIS433-BLD: Student E, Student F, Student G	CIS365- BLD: Student S, Student T Student U	
	MIS635-BLD: Student H, Student I	MIS635-DL: Student V, Student W Student X	
	CIS684- BLD: Student J, Student K, Student L	CIS675-DL: Student Y, Student Z Student AA	
		IE685–BLD: Student AB, Student AC	19
Total	12	17	**29**

All interviews were transcribed and imported into NVivo® for analysis. Section 8.2.1 explores research questions 7 and 8 using student interview data. Section 8.2.2 investigates research question 9 using faculty interview data.

8.2.1 Student Interviews

Regarding their use of mobile devices as related to research question 7, students' responses in the interviews match with the data collected from the log. Most frequently students mentioned using notebook computers and the wireless Internet. In addition, mobile phones are used to communicate and coordinate meetings, or keep track of group progress including reminding people of deadlines. While students had the option to use the web conferences or emails to communicate with other students asynchronously, or to use the phone to talk with people synchronously, the choice tend to be made based on the nature of the task as characterized by one student:

> "I used mobile phones for scheduling and communicating with my teammates to discuss something right away. To sort out confusions, for example, when is this due? Is it tomorrow or today? If I post online, I don't know when I will get a response."
>
> \- Student W

One student's response pointed out another interesting mode of communication which is in-between the asynchronous and synchronous modes: text-messaging using mobile phones. The example suggests text messaging on mobile phones is used to get a relatively fast (almost synchronous) response without posing imminent disturbance in the recipient's environment (one of the characteristics of asynchronous communication).

> "I used text messaging. I used them not to ask a question, but things like when can we meet? Or where are you? Are you in class? I used it because I am not sure where they are at six pm. You do not want to call someone who might be in class. So it is better to use text messaging. But there are times when I call them. But that will be like during the day when they are not in class."
>
> \- Student AB

Related to research question 8, most of the students being interviewed reported spending *more* time and effort in the online exams compared with a traditional exam. Some reported spending time in each phase of the online exam process to ensure the quality of their work (e.g. questions, answers, grading justifications). For example:

> "Probably I put in a little more time. You are more involved and you have to spend time in each phase. "
>
> \- Student L

> "I'd say overall I spent more time, a lot of time on each component. For the questions, I wanted them to be challenging and I wanted them to make complete sense. So I spent quite a bit of time designing them and assuring that occurred. For grading, I wanted to be sure that I was as fair as I could possibly be with

providing constructive criticism. My answers I spent the least amount of time on, but I did spend quite a bit of time on them too."

- Student R

"Because they are essay questions that require 700 to 1700 words (in answers), you have to write a lot of text… We were given two questions. I spent about one day thinking about how to answer it, and looking through my textbook and my notes. And I think I wrote my answer a few times until I was sure, probably until Sunday... So it took me a couple of days to study the information and settle the references and to re-organize the answer."

- Student J

In addition to spending time to ensure the quality of the exam questions, answers, and the grading, students in the collaborative exam also talked about the extra time they spent to coordinate group activities and follow postings online. These were mentioned especially by the group leaders in the collaborative exam. For example:

"If I was not the group leader, I probably say I would spend less time studying. But being the group leader push(ed) me to read some of the questions they put on, and look over some of the chapters they mentioned. Taking a sample of what they did is an illustration of what everybody else would be doing. So I took that as a guideline as what type of questions I should expect. "

- Student U

"Probably more. There was a lot of overhead such as coordinating with the group, understanding what the project was, email my questions and waiting to get them back, and reading other posts. There were so many overheads. Maybe hour for hour it could be about equivalent. But I sometimes just spent an hour and a half just looking over the posts to get them straight."

- Student Q

On the other hand there is a wider range of the time spent in studying for the online exam compared with the traditional, as suggested in the log data (Table 8.2). The following student comment helped to explain the difference:

> "I think I spent a lot more time doing this. But I think, because the quality of question varies substantially, I can see people spend a lot less time than the traditional exam if they do not want to provide the quality. So they can spend a lot less time. But if you want to provide the quality, then you will spend a lot more time. "
>
> - Student H

A small number of students being interviewed reported spending *less* time in exam study. They reported that the time needed to memorize was less as they had access to the resources in the online exams and therefore they felt they spent less time in studying for the online exams.

> "In terms of time straight up I probably put in less. However I'm the type of person that has to memorize everything so studying for a traditional exam takes much longer. I personally find this format to be more enjoyable and I feel I get more out of the course and subject matter this way."
>
> - Student C

> "Compared with a traditional exam, I probably studied less. But the reason was because I knew that taking the online exam I would have access to textbooks and other resources. So to study for it, I went through the book to understand where things are laid out and where I can find things more so than or actually learning or studying like memorizing."
>
> - Student S

Some students commented on the *different* approaches they used in preparing for the online exam. For example:

> "In traditional exam I study one or two days before the exam. This exam I studied everyday for a few hours for over a week."
>
> - Student V

> "Cumulatively, the time spent probably might be the same. But you used it over a period of time. Although the deadline was also there for the online exam, the effort was distributed."
>
> - Student X

> "It is less preparation to answer questions knowing there are only going to be two questions. In this course with the amount of materials to read, it is a relief. But what I had done instead was that I went across all the articles and got a sense of

things, but I did not commit to anything memory per se. What I did was mark off in my notes, highlight the notes that would be useful if I had to answer something. But there was less trying to remember things."

- Student Y

Two distinctive features of the way students studied in online exams emerged from the interviews. First, students mentioned the online exams allowed them to be more focused on certain topics and allowed them to conduct in-depth research. For example:

"Traditionally when you take an exam you study for the topics that are on the exam. This (exam) is basically any type of topics that you want to choose to place your questions on, and it made it littler easier to study. Because you are not studying so much; you are studying what you would like to design your questions on. It is more focused."

- Student F

"Cumulative hours probably came out close to the same and I still prefer the online open book to a (closed book) exam ...because I do well with a point of reference... I need something to trigger my memory. What you end up doing is you spent less time real "studying" study, because you are not looking at everything. But you spent a ton of hours, at least for me, to answer the questions because I went through every angle and possibilities. I think it is good to have two questions because it gave you a chance to really elaborate and really bring it together as a whole."

- Student O

Secondly, students mentioned the exam motivated them to use and synthesize materials both within and outside the course material list, and tried to look at the overall course structure and content rather than memorizing specific facts.

"What I did is that I went to the library. The professor said he does not want answers to be copied directly from the textbook. He wants us to do research as well.... I went to the library and read three reference books. Initially you went to read what your question is about, but you end up reading something else that interests you when you read it, and you learn the knowledge."

- Student AB

"I really tried to strike or get an overall feeling of what we have learned. I tried to cover everything. By doing that, I ended up covering a lot of COBLE code. Every of my questions involve code, which I feel was the focus of my class. In

> the definition question, I had two parts. One was defining or identifying different sections of COBLE code. The other part of the question asked about how files and record are related to each other. I asked another question that asked students to relate records and Abstract Data Types (ADT), but professor has thrown that out. But I thought those are the things that required to be understood."
>
> - Student S

The student interviews reveal that the online exams encouraged students to cover more breadth and depth of knowledge in exam study compared with a traditional exam. The exam allowed them to see a bigger picture related to the course subject rather than memorizing specific detailed facts. The online exams also allowed them to conduct in-depth research and discover a deeper level of knowledge. These findings are consistent with the survey results, where students in the participatory and collaborative exams reported more improvement in skills such as relating knowledge from several areas to make arguments, seeing different course components and organizing them in a meaningful way, recognizing patterns and the underlying meanings of course content, and using methods, concepts, and theories in solving new problems.

8.2.2 Instructor Interviews

A total of eight post-exam interviews were conducted with the five faculty members who conducted the online exams (except for the author whose classes also participated in the online exams). One faculty member participated in the study in all three semesters and was interviewed three times. Another faculty member participated in the study in two semesters and was interviewed twice. Faculty is referred to as professor A through E in this section. Interview data were used to investigate research question 9:

> **Q9**. Do instructors spend more or less time in preparing for and conducting the online exam than the traditional exam?

To examine this question, faculty members were asked in the interview whether they felt it was a lot of work in:

1. Adapting the online exam to the existing course structure
2. Grading the online exam compared with the traditional exam
3. Resolving student disputes of grades

Related to the first aspect, three out of five professors reported that it was not a lot of work to adapt the online exam to their courses. This is expected given the fact that the experimenter (the author) provided the initial documents regarding the exam procedure, guidelines, etc., and worked with each of the professors to customize the documents to fit their individual courses. One professor mentioned some changes made to fit the online exam schedule into the course syllabus.

> "We canceled the final project mostly because the exam takes two weeks and the summer semester is short."
>
> - Professor C

Another professor reported that it was more work because the course did not have an exam before.

> "Since I usually do not have the exam, it is more work for me. I would say I would have got to do it as a midterm or before the end of the semester. I had 26 students in the class. The timeframe was probably good from the student's perspective. "
>
> - Professor A

Regarding the second aspect on grading, it is clear from the interviews with faculty and the log data (only one) that grading was the activity where they spent the most time in conducting the online exams. For example, professor E reported spending a total of 16.5 hours in conducting the online exam in a class of about 35 students with the following time allocation:

- Administration & Communication (posting announcements, answer questions): 2.5 hours
- Question (review, revision, assigning): 4.5 hours
- Grading (question/answer/grading justification quality): 9.5 hours

The same professor reported spending 10.5 hours in conducting the online exam in another class with about 17 students with the following time allocation:

- Administration & Communication: 1 hour
- Question: 1.5 hours
- Grading: 8 hours

As seen from the two examples above, the professor spent the largest amount of hours in grading. From the interviews, three out of the five faculty members reported it was more work for them to grade the online exam compared with a traditional exam. Two key sources of the increased workload are identified. One comes from the fact that the professors had to grade not only the answers, as in a traditional exam, but also the questions and the grading justifications to provide the question grade and the grading grade. Another reason is that given students designed their own questions, professors had to deal with different questions which increased their workload. For example:

> "It is definitely more work... It takes more time not because of this type of exam, but because the questions are different. This means you have to pay attention to each question."
>
> \- Professor C

> "Yes, for me it is a lot more work. In a normal exam, I take the same exam and I mixed it up to make three or four exams. So I have an answer key. It is very simple for me to go through and see what is going on. In this grading, I have 180 different questions and I have to look at each one. And I have to figure out whether the grading the student replied makes sense... We did three questions

(per student), and that is about the maximum. Look at the number of hours we spent doing it. A larger test would just kill you."

- Professor D

One of the two faculty members who participated in the exam more than once initially reported the exam was more work:

"Yes. (The online exam is) very much more (work). They are all different questions. And I am grading questions, the answers, and the grading justifications."

- Professor A

The following semester the same professor reported:

"I was much quicker this time. But I did have to change (my grading template) because although I graded using the same team template I had before, we adjusted the grading criteria."

- Professor A

Another professor who conducted the exam more than once reported that the exam was less work. But after considering grading of questions and grading justifications, it was probably more work:

"It is less to grade because you have students to grade it before, and you can certainly base your grade on theirs. So I did not find it took any more time. In fact I skimmed some of the answers fast so that is not a problem. But when you talk about the total time when you have to grade the questions, then grade the grading, then also grade the answer, probably total is more time."

- Professor B

The following semester the same professor reported that the total amount of time probably evened out was less than the traditional exam:

"I do not think it is much different (than the amount of time I spent in grading the traditional exam). Grading the (online) exam is easier because you have someone else's comments there, so that's definitely easier. But then the fact that you have to grade the questions and grade the grading, it probably evens out. I do not think it is any more work for the midterm. Overall, maybe it is even a little bit less."

- Professor B

students did not follow the guidelines and posted the messages in the wrong places. For example:

> "I did post a root item for students to reply to with their questions etc. But some said my message was not clear. So some groups ended up posting their questions by replying to the root, others just posted their questions. There were probably 12 questions that were posted as new messages but not replies. So they all came up at the end of the page."
>
> - Professor A

> "We had the same problems this semester as we got last semester where students put their answers in a different place, as a different reply."
>
> - Professor B

In one case the professor had about 60 students in one WebCT course conference. The lack of a sorting mechanism made it hard for the professor to organize the conference:

> "The only difficulty I dealt with is with the number: we are looking at 180 questions. It is hard to keep track where you are and look at the flow. It would be so much easier to be able to put them into a certain order. You are sort of bouncing around. It is sort of hard to keep your mind on what you have just done, where you are at, and you have no clue in terms of where you are in the process."
>
> - Professor D

In summary, most faculty members who conducted the online exam the first time reported spending more time in grading the online exams compared with the traditional exam. On the other hand, faculty members who conducted the online exams multiple times reported becoming quicker and spending less time in grading the online exams than the traditional exam. In terms of the ALN environment, the systems have been shown as effective in conducting online exams in regular-sized classes with a moderate amount of postings. When the class size increases, the systems show lack of functionalities to support large amount of postings.

CHAPTER 9

CONCLUSIONS AND FUTURE RESEARCH

9.1 Summary

Table 9.1 and 9.2 summarize the hypotheses and the results. Table 9.1 shows the hypotheses and results grouped by the intervening and the dependent variables, including deep exam study, surface exam study, social engagement, perceived learning, satisfaction, perceived fairness in grading, and grades (including exam grade, question quality grade, grading quality grade). Given deep exam study resulted in two factors (deep exam study, surface exam study) using factor analysis, hypotheses H1.2, H2.1, H3.1 and H3.3 were tested using the deep exam study and the surface exam study factors, respectively.

Table 9.1 Summary of Hypotheses I – Comparison of the Three Exam Modes

Measure	Hypothesis	Results	Significant
Deep Exam Study	H1.2: C>P	C=4.74 P=4.78 T=4.51	Not Supported
	H2.1: C>T		Not Supported
	H3.1: Pre-exam Deep Learning	.55**	Supported
	H3.3: Correlation: C>T	C=.49** T=.58**	Not Supported
Surface Exam Study	*H1.2: P>C*	C=169.78 P=159.55 T=243.75	Not Supported
	H2.1: T>C		Supported
	H3.1: Pre-exam Surface Learning	.23**	Supported
	H3.3: Correlation: C>T	C=.11 T=.32**	Not Supported
Social Engagement	H1.1: C>P	C=4.41 P=3.27 T=3.38	Supported
	H2.2: C>T		Supported
	H3.2: Pre-exam Collaborative Learning	.43**	Supported
	H3.4: Correlation: C>T	C=.44** T=.51**	Not Supported
Perceived Learning	H1.3: C>P	C=24.78 P=23.91 T=21.78	Not Supported
	H2.3: C>T		Supported
	H6.4: Correlation with Deep Exam Study: C>T	C=.49** T=.59**	Not Supported
	H6.5: Correlation with Social Engagement: C>T	C=.57** T=.37**	Supported
Satisfaction	H1.4: C>P	C=4.57 P=4.46 T=4.57	Not Supported
	H2.4: C>T		Not Supported
	H6.1: Perceived Learning	.62**	Supported
	H6.2: Correlation with Deep Exam Study: C>T	C=.25** T=.44**	Not Supported
	H6.3: Correlation with Social Engagement: C>T	C=.50** T=.24**	Supported
Exam Grades	H1.5: C>P	C=2.86 P=3.15 T=2.50	Reverse Effect Found
	H2.5: C>T		Supported
Perceived Fairness in Grading	H1.6: C>P	C=186.59 P=193.02 T=192.28	Not Supported
	H2.6: C<T		Not Supported
Question Quality Grades	H1.7: C>P	C=133.65 P=161.95	Reverse Effect Found
Grading Quality	H1.8: Correlation with Instructor's Grading: C>P	C=.88** P=.70**	Supported

C: Collaborative Exam; **P**: Participatory Exam; **T**: Traditional Exam

As shown in Table 9.1, there are many significant and stimulating findings from this study. First students' adoptions of deep or surface exam study strategies were significantly associated with their deep or surface learning predispositions (H3.1). More importantly, the participatory and collaborative online exams resulted in students adopting a significantly lower level of the surface exam study strategy (H2.1). Students in the online exams reported significantly less use of memorization and rote learning without understanding the materials in exam study compared with those in the traditional exam. Instead, students reported in the interviews that the online exams allowed them to prepare for the exams differently than for the traditional exam, motivating them to conduct in-depth research and cover a broader breadth of knowledge and synthesizing materials in exam study.

Second, the level of social engagement in exam study was not only significantly associated with students' pre-exam disposition in collaborative learning (H3.2), but more importantly, was significantly higher in the collaborative exam than both the participatory exam and the traditional exam (H1.1, H2.2). Students in the collaborative exam reported perceiving the exam process as a group experience, forming a sense of learning community, and interaction with other students enhanced their understanding of course materials. Strikingly students' level of social engagement in the participatory online exam was significantly lower than that in the collaborative exam, and was at the same level of the traditional exam. This suggests that without incorporating the small group activities into the online process, the level of engagement in learning from others and the sense of a learning community are as low as in the traditional settings.

The moderating effect of social engagement on exam outcomes was also significant. There were higher levels of correlation between social engagement and students' satisfaction (H6.3) and perceived learning (H6.5) in the collaborative exam than the traditional exam. This suggests students who socially engaged themselves in exam study were more satisfied and perceived they learned more in the collaborative exam than the traditional exam. Additional analysis using the least and the most socially engaged students in the collaborative and traditional exams shows that there was a negative synergistic effect of social engagement, where the least socially engaged students had the lowest level of satisfaction (significant, Figure 7.5) and perceived learning (Figure 7.7) in the collaborative exam.

Significant differences were also found in exam outcomes in terms of perceived learning, but not in satisfaction or perceived fairness in grading. Students in the collaborative exam reported significantly higher levels of perceived learning than those in the traditional exam (H2.3). Students reported improvement in their understanding of course materials, and enhancement in skills such as using knowledge in new situations, solving problems, recognizing patterns, making generalizations and predictions, and comparing ideas and making judgments. These correspond well to the interview responses, where students reported the collaborative exam motivated them to synthesize materials, conduct in-depth research, and search for additional materials. Although the level of satisfaction is significantly correlated with perceived learning (H6.1), there was no significant difference in satisfaction and perceived fairness in grading among the three exam modes. In addition, the correlation between satisfaction and deep exam study is lower than the correlation between perceived learning and deep exam study, and so is the correlation between satisfaction and social engagement as compared with perceived learning and social engagement.

Some interesting results were unveiled when grades were analyzed. While it was hypothesized that students in the collaborative exam would receive the highest exam grades and question quality grades, students in the participatory exam actually received the highest (H1.5) exam grades among the three and higher question grades than the collaborative exam (H1.7). Students in the collaborative exam did receive higher exam grades than those in the traditional exam, and the result is significant (H2.5). This suggests that students did the best job in the participatory exam as judged by the exam grades given by the instructors in the exams, followed by those in the collaborative exam. One possible confounding issue to this result is that, although the courses were carefully matched in terms of levels and characteristics when allocated in the three exam modes, grades provided by different professors in different courses are not directly comparable. On the other hand, provided that the exam grades are comparable, then students in the participatory exam might be most academically well-prepared before the exam process. Regarding student grading quality, there is a higher correlation between students' grading and the instructor's grading in the collaborative exam compared with the participatory exam (H1.8), indicating that working in small groups produced higher quality grades than working alone online.

As discussed above, the collaborative online exam was effective in promoting social engagement, reducing surface learning in exam study, and increasing perceived learning. Was there any difference in the perceptions of the collaborative online exam among different courses? Table 9.2 provides the summary of hypotheses four, which is grouped by course characteristics, including distance-learning vs. blended, undergraduate vs. graduate, technical vs. discussion-oriented, and classes of small vs. large sizes.

Table 9.2 Summary of Hypotheses II – Collaborative Exam in Different Courses

Hypothesis		Results	Significant
DL>Blended	H4.1.1: Learning	DL= 25.58 BLD= 23.59	Not Supported
	H4.1.2: Satisfaction	DL=4.43 BLD =4.79	Not Supported
	H4.1.3: Fairness	DL= 64.79 BLD =70.34	Not Supported
Under<Graduate	H4.2.1: Learning	U=22.55 G=26.68	Supported
	H4.2.2: Satisfaction	U=4.57 G=4.58	Not Supported
	H4.2.3: Fairness	U=69.96 G=64.49	Not Supported
Technical<Discussion	H4.3.1: Learning	TCH=23.55 DISS=25.44	Not Supported
	H4.3.2: Satisfaction	TCH=4.99 DISS=4.35	Reverse Effect Found
	H4.3.3: Fairness	TCH=70.85 DISS=64.97	Not Supported
Small Class>Large Class	H4.4.1: Learning	SM=26.78 LG=23.82	Not Supported
	H4.4.2: Satisfaction	SM=4.56 LG=5.16	Not Supported
	H4.4.3: Fairness	SM=40.80 LG=47.83	Not Supported

As shown in Table 9.2, there is no significant difference in perceived learning, satisfaction, and perceived fairness in grading in DL vs. blended courses. The result shows that the collaborative online exam worked equally well in courses offered completely online and courses with the online and in-class components. The same conclusion can be made regarding classes of small size (<20 students per class) vs. larger size (>=30 students per class), where no significant difference was found in perceived learning, satisfaction, and perceived fairness in grading. Additional analysis on exam strategies revealed that there was significantly lower level of surface exam study in DL courses compared with blended courses in the collaborative exam. Similarly, students in small classes had significantly lower level of surface exam study and higher level of deep exam study than students in large classes. Taken together, these results suggest that the collaborative online exam worked equally well in DL/Blended courses and classes of small/large class size, and more deep exam learning and less surface learning was best achieved through small class size and DL courses.

A significant difference was found between students in undergraduate and graduate courses in the collaborative exam, where students in graduate courses reported significantly higher perceptions of learning compared with the undergraduates (H4.2.1). Additional analysis showed that students in graduate courses also adopted significantly higher levels of deep exam study than their undergraduate counterparts. The results suggest that there was no significant difference in satisfaction, perceived fairness in grading, and social engagement between undergraduate and graduate, and the collaborative exam was most effective in encouraging deep exam study and achieving higher perceptions of learning in graduate courses.

Interestingly, the comparison between students in technical vs. discussion-oriented courses show that, opposite to what was hypothesized, students in technical courses were more satisfied than students in discussion-oriented courses (H4.3.2). Most of the technical courses in this study had programming components, and interviews with students and faculty revealed that the online exam allowed students to actually write, compile, debug and execute the programming code, which are not possible in traditional exams without access to computers. On the other hand, additional analysis on exam study strategies revealed that students in discussion-oriented courses adopted significantly lower levels of surface exam strategy. This is congruent with the interview results where students reported searching for additional resources and synthesizing materials rather than memorizing facts in preparing for discussion-oriented courses. The results suggest that there was no significant difference in perceived learning, perceived fairness in grading, and social engagement between technical and discussion-oriented courses, and the collaborative exam was most effective in increasing the satisfaction in technical courses while lowering the level of surface exam study in discussion-oriented courses.

9.2 Conclusions and Discussions

Literature in ALN suggests that *interaction* and *higher-order learning* are keys to effective online learning (Lorenzo and Moore, 2002). In particular, it has been shown that active involvement in online learning activities (Alavi, 1994; Alavi and Dufner, 2004), interaction with classmates as well as instructors and course content (Vogel, Wagner et al., 1999; Swan, 2004), and small group discussions are specially effective in enhancing online learning (Hiltz, Coppola et al., 2000; Swan and Shea, 2004).

Regarding *interaction*, this study has shown that incorporating small group activities in the online learning process can significantly increase interactions among students and enhance their sense of an online learning community. Significant differences were found between the collaborative exam and the other two exam modes in the level of social engagement, which is defined as the extent of students' active involvement in learning from other students in exam studying, and the sense of forming a learning community. Results show that the collaborative online exam which incorporated small group activities resulted in significantly higher levels of interaction among students which enhanced in their understanding of course materials. Students got to know other students better through the exam process, perceived the exam as a group process, and formed a sense of a learning community. In contrast, the participatory online exam which did not incorporate small group activities resulted in the same low level of social engagement as in the traditional exam. This finding indicates that without incorporating small group activities into the online learning process, the level of interaction and the sense of a learning community are as low as in traditional settings.

Regarding *higher-order learning*, this study demonstrates that active involvement in the exam process can significantly reduce the use of surface learning in exam study. Significant differences were found between both online exam modes and the traditional exam in the level of deep/surface exam study, which is defined as the extent of the student's search for knowledge and understanding in exam study process rather than memorizing facts to pass the exam. Results show that by being actively involved in designing exam questions and grading exam answers, students in both participatory and collaborative online exams adopted significantly less surface learning strategies in exam study. Instead, they adopted professional perspectives to understand course material, and adopted academic perspectives to relate course content to other subject areas. Student log data confirmed that students in the online exams used exam materials more frequently to learn during the exam than students in the traditional exam.

In addition to deep/surface exam study strategy, *higher order learning* was examined through perceived learning as one of the exam outcomes. This study reveals that significant differences exist between the collaborative exam and the traditional exam regarding perceived learning developed based on Bloom's taxonomy, measuring learning from lower levels such as understanding the materials to higher levels such as solving problems, recognizing patterns, comparing and discriminating ideas, and making generalizations or prediction. This study shows that through active involvement and small-group activity, students in the collaborative exam reported significantly higher perceptions of learning than students in the traditional exam, including enhancement in skills in using knowledge in new situations, solving problems, recognizing patterns, making generalizations and prediction, and making judgment and assessing quality of arguments.

To summarize the findings discussed above, collaborative examinations significantly increased the adoption of higher order learning and interaction. Collaborative examinations also significantly increased perceived learning. These findings are congruent with the interview comments from the students. When asked what they had learned from the online exams, students commented that the exam process broadens their knowledge through reading others' questions and answers. They also gained real-world perspectives by applying the concepts into problem solving, idea presentation, and communication.

> "I did learn how to apply many of the concepts in the book when designing questions and answering questions. I also gained great perspectives and saw how applicable this material is and could be in the real work. I guess the true value of the materials showed during the midterm process… The exam made the concepts more obvious in how they apply in the real world."
>
> \- Student R

> "I think the fact that you have to sit down and present a good writing (in the online exam) applies to the working environment very well. A lot of times, especially what I do, you cannot just deliver the results. You have to let people understand what you are trying to do. So I think this helps me in terms of preparing and presenting in project work kind of situation. And this helps me to look at issues from different angles."
>
> \- Student H

> "Before I really did not know a lot about all the stuff available in IT such as Wi-Fi and the capabilities of cell phones now and what's start to become available. I think I became a lot more aware of the technologies that are out there because of the questions I had to answer and reading through other questions and responses in the exam conference."
>
> \- Student D

In particular, students from the collaborative exam commented that the group experience was a valuable learning opportunity, and reported improvement in teamwork and project management skills:

> "I would say the group experience. Whenever you work with groups, it is valuable. "
>
> \- Student Z

> "I learned a lot about project management and trying to hold the team together because I am the group leader not only for the midterm but also for the semester-long project. I always find myself coordinating meetings making sure everybody is doing something to contribute. That in itself was a valuable thing to take away."
>
> \- Student S

> "I learned management skills in terms of managing my team, such as motivating them to turn in their questions quickly and coming up with a mechanism to select for questions. "
>
> \- Student W

From the faculty's perspective, most of the faculty who did the online exams for the first time reported spending more time in conducting the online exam than the traditional exam. Yet faculty who conducted the exam multiple times reported spending less time in the online exam than the first time, and reported that the time was reduced to about the same level as the traditional exam. More importantly, the instructor's role shifted from the designer of the exam as in the traditional exam, to the facilitator of students' learning and consistency keeper in grading. Instead of designing exam questions, the instructor spent time in reviewing the questions students designed and in providing revisions and feedback. Instead of grading the exam answers from scratch, the instructor spent time in reviewing students' grading and in providing consistency in grading. The following comment from a professor provides a good description of the instructor's activities in grading:

> "To be fair, I have to grade every answer. I skimmed the question and looked at the grading justification. If it is very detailed and very good and relatively high grade, I did not spend time on it. If it is detailed and low grade, then I spent time grading the answer myself. If it is minimum details, high or low, then I have to grade it (the answer) myself because there was not enough detail."
>
> \- Professor A

Both students and faculty raised a number of issues and suggestions to improve the online exam. Two unique features of the online exams are the student question design and peer answer grading. As discussed in Chapter 7, overall, students in the online exams thought their peers were able to design questions of good quality for the exam. One of the issues raised in the student interviews was the question equality. Some students commented that although the instructor's review of the questions enhanced the quality of questions, some questions were easier or much more focused than others to answer. For example,

> "The professor did review the questions and he did modify a few of them. But one of mine was very broad and I thought that was not fair… Based on what I saw, some questions were much more focused, and it was easier to answer."
>
> \- Student J

To help address this issue, instructors in this study were asked to evaluate the difficulty level (e.g., low, medium, high) and topic area of each question when they reviewed and revised the questions (see the instructor's manual in Appendix E). When assigning the questions to students, instructors were asked to balance questions in terms of topic areas and difficulty levels. Professors reported that balancing was easier to achieve in large classes or with large number of questions than in small classes or with small number of questions. For example:

> "But one disadvantage of it (the exam) is that obviously some questions are harder than others so it's not exactly even for students. I mean you try to balance out the questions. But when you have a small number of groups, like the three groups that did it (the exam), I am not sure I was able to do a good balancing job as I did in (another class) because you had a larger class and you had more choice. So that's one concern with a small class."
>
> \- Professor B

> "In my undergraduate class, each student gets three questions. So it was easier to adjust difficulty levels for three questions. But in the graduate course there were only two questions and it made it harder to balance the questions."
>
> \- Professor E

Regarding grading, overall, students in the online exams believed that their peers were capable of grading the answers, although the level of confidence was less compared with their responses in peers' abilities in designing questions (Chapter 7). Fairness in peer grading was not brought up as a main issue in interviews as compared with previous studies (Shen, Cheng et al., 2000; Shen, Hiltz et al., 2001). This may be due to a few changes made in this study. First, the grading guideline was designed to reduce the number of individual dimensions (categories) for grading. The new grading guideline has only four categories to allow students to make global judgments. In addition, the elimination of the second level grader not only shortened the length of the exam, but also allowed each answer to be evaluated directly by the instructor as opposed to only the answers that were disagreed about by the first and the second level student graders. Yet this increased the workload on the instructors. To solve this issue, some students and faculty suggested that the student who designed the question should provide the answer, and give the answer as part of the answer justification. This may indeed enhance student learning experience as well. For example,

> "Someone who generated the questions should also reply to their own questions with an answer that would get full credit."
>
> - Student Y

> "One of the changes I want to see is that when they post questions, they have to give you the answer beforehand. That will help us to know what they want and help both parties to learn."
>
> - Professor D

In addition to question design and grading, a few other issues and suggestions were raised in the interviews. In terms of the exam schedule, the online exams were streamlined into two-weeks in this study with three days for each main phase of student activities, including question design, answering, and grading. While most of the students felt the schedule was fine, some students reported the schedule was a little tight and preferred different due dates. For example,

> "The whole process took two weeks. The only thing I can say is that sometimes I felt a little under the gun as far as time. I understand it is a summer course and it is shorter duration. I think three days is efficient, I just do not know whether weekend days are the best days (the answering phase was on Saturday, Sunday and Monday). Personally I am married with three children."
>
> - Student D

This is raised especially by those students who did the online exam for the first time, who commented that reading the instructions and understanding the way the online exam works took time.

> "I understand that this is something new. It will take a while to catch on. Since this is the first time we are doing it, we have to go through the instructions page by page. Of course it will be easier if we have to do the procedures again."
>
> \- Student AB

As part of the solution, students and faculty suggested making the exam instructions more succinct, posting the exam instructions well in advance of the exam, and making sure students understand the exam procedures before the exam starts. For example,

> "The instructions you wrote was very clear and well written. But not only myself but a lot of other students had trouble understanding it. I think partly due to the fact that this is a new concept that was presented and we had some confusion. But I know when I read through the directions, what I was looking for was very succinct one sentence in the middle of four or five. I had to read it a couple of times before I really understood. I think the professor should run through the instructions maybe at class before the exam started."
>
> \- Student S

A more serious issue concerning schedule is that students in the participatory exam and the collaborative exam were given the same amount of time in each phase of the exam. Students in both exams were given three days for each main phase of the exam, including designing questions, answering, and grading. Considering the question design and grading phases were group activities for the collaborative exam, there may not have been enough time for the groups to thoroughly discuss and enhance the quality of the questions or grading. As Leidner and Fuller (Leidner and Fuller, 1997) pointed out in their study of collaborative learning using GSS, when collaborating under time constraints, students did not process and assimilate information into their own cognitive framework. They rushed to contributing their ideas and did not have time to seriously think about others' ideas. Although the collaboration was conducted in asynchronously environment in this study as opposed to the synchronous settings in Leidner and Fuller's study, the timeframe may have limited the extent of idea assimilation and eventually the quality of group questions and grading.

While the online exams in the spring and summer studies were conducted as the final exam in all but one class, faculty preferred to have the online exam as a midterm exam rather than a final exam from the workload and learning point of view. For example,

> "But I would suggest us to use it (the exam) as a midterm, not final, so we will not have all the work in the end."
>
> \- Professor A

> "In addition, if this was my midterm, I could do some work in response to their feedback, such as problems shown through the exam. When it is a final exam there is no motivation for me or them to do much. Probably I would like to have it as my midterm rather than the final."
>
> \- Professor C

To summarize, this study shows that collaborative online examination significantly increased the adoption of higher order learning and interaction in exam study. Collaborative examinations also significantly increased perceived learning. These findings are congruent with the interview comments from the students regarding their learning in real-world perspectives and teamwork skills. Suggestions were raised to improve question equality, grading fairness, exam schedule, instruction clarity, and to reduce instructor's workload.

9.3 Contributions

This dissertation offered both theoretical and practical contributions to the research in ALN, collaborative learning, and online collaboration. This study extended theories in collaborative learning into online assessment, and proposed the collaborative online examinations model. Through rigorous field experiments, this study has shown that collaborative examinations significantly enhanced interaction and promoted higher order learning. In particular, small group activities in the online learning process significantly increased interactions among students which enhanced their sense of an online learning community. Active involvement in the online exam process significantly reduced the use of surface learning in exam study. Overall, students reported significantly higher perceptions of learning in the collaborative exam than the other exam modes. A number of significant relationships were discovered, which are congruent with other studies in ALN. For example, studies have found that students' overall perception of social presence was a predictor of their perceived learning in online courses (Picciano, 2002; Richardson and Swan, 2003). Results in this study indicate social engagement is significantly correlated with perceived learning, and satisfaction. This confirms that the level of social engagement is an important mediator of students' learning experience in the online environment.

Practically, the results from this dissertation study provide guidance not only to ALN instructors and course designers, but also assessment officials in higher education. For example, researchers pointed out the importance of developing assessment methods to promote deep learning (Suskie, 2000) in the AAHE's (American Association for Higher Education) conference on assessment. Both the participatory and the collaborative examinations significantly reduced the adoption of surface exam study strategy, and promoted deeper approach in exam study. Detailed online exam procedure guidelines were designed for both students and instructors, and the feedback from student and faculty interviews provided practical guidance in implementing the exams in courses. This study shows that the role of the faculty in online exams changed from the designer of the exam to facilitator of students' learning. The study has also shown the effectiveness of the collaborative exam in a variety of courses, including courses at graduate and undergraduate levels offered in face-to-face or distance learning modes.

Relevant to both theoretical and practical contributions, several scales were developed and showed high reliability, and are now available for other researchers to use. For example, the perceived learning scale was developed based on Bloom's taxonomy with 12 items measuring perceived learning from lower levels such as understanding the materials to higher levels such as solving problems, recognizing patterns, comparing and discriminating ideas, and making generalizations or prediction. Reliability analysis shows the scale is highly reliable with Chronbach's alpha at .93. The deep exam study scale was developed with six items which measure the adoption of deep learning strategy in exam study such as adoption of professional perspectives, academic perspectives, etc. The deep exam study scale was tested as highly reliable with Chronbach's alpha at .80. Other scales were successfully adapted and have also shown high reliability, such as social engagement (4 items, alpha=.80), collaborative learning (7 items, alpha=.87), deep learning (5 items, alpha=.83), and surface learning (5 items, alpha=.78).

The exploration of the exam study medium, including the Internet and the mobile devices, contributes to the understanding of the use of computer technology and systems in exam study, and the design of new learning systems using mobile devices. While this study suggests the current use of the Internet surpasses the use of mobile devices such as mobile phones and laptops in exam study, the rapidly increasing capabilities of mobile devices and their adoption in education makes it possible to move from E-Learning to M-Learning (Mobile-learning) in the near future.

9.4 Limitations

One main limitation of the study that may affect the validity of the experiment is the fact that students in different courses were used in different exam conditions. Yet as a field experiment, it would not be feasible to have total randomization in subject assignment. As discussed before, the random assignment of students in each class to one of the three exam modes may result in reactivity effect, which refers to subjects' awareness of participation in a study. In addition, the workload on the instructor to conduct three different exams in one course would be unrealistic. In this study, students were assigned to different exam modes based on the course they were taking. In most cases, each course section was assigned to only one exam mode. To avoid the selection effect that may confound the experiment results, which is the pre-existing differences between conditions, two measures were taken. First, courses were evaluated in terms of their characteristics, such as undergraduate vs. graduate, technical vs. discussion-oriented. Courses with similar characteristics were then assigned to one of the three exam modes to help counterbalance the pre-existing differences in students before the exam. Second, a pre-exam survey was distributed before the study to measure the pre-existing differences among students. The pre-exam survey results were examined to determine pre-existing differences. As discussed in Chapter 5, there were no significant pre-existing differences among the three-exam modes.

Despite the good result that there was no significant pre-existing difference among students, a few issues related to the fact that different courses participated in the study may confound the findings. Most of these issues are related to the variations and deviations from the standard exam procedures, which are discussed in Section 4.6.2. First, all the experiments in the spring and summer studies were conducted during the final exam period except for one course, while the rest were conducted during the midterm. Although no obvious difference was observed, instructors reported preferring the online exam as the midterm as opposed to the final due to the busy schedule in the final period. Second, even though instructors were told to make a separate time slot for the online exams without having other assignments due in between, a few courses had other assignments or group projects going on in parallel with the exam, while other courses did not. Students reported in the interviews that the projects or assignments going on in parallel with the exam increased their confusion of the exam process and reduced their satisfaction with the exam. Third, although the online exams followed the general 2.5 week exam schedule, the actual length varied from 15 days to 24 days depending on the individual course and the instructor's schedules. Fourth, the number and actual type of questions elicited from students slightly varied in different courses to accommodate different course requirements. Based on the interviews and the observations, issues one and two seem to be more important than issues three and four. While there were variations in the traditional exam procedures as they were conducted in different courses, as discussed in Section 4.6.2, the traditional exams followed the same basic procedures and they are adequate in serving the purpose of the baseline condition in this experiment.

Related to the quality of group work in the collaborative exam, the consistency may have been affected by the differences in groups, including group history, size, and homogeneity. Regarding group history, to leverage cohesion of existing groups (Mennecke, Hoffer et al., 1995), groups from previous assignments or ongoing group projects (termed assignment groups) were used in the collaborative exam. In some cases the assignment groups had too few or too many people to satisfy the three to four people per group requirement for the collaborative exam. So some assignment groups were divided to form the collaborative exam groups, some were combined, and some were newly formed. This results in groups with different levels of group history and cohesion with group members. In addition, while most of the groups in the collaborative exam were of size three or four, some groups had members who dropped the course and ended with two or in a very few cases just one person. While the number of questions and the amount of grading was not affected by group size (e.g., two questions per student regardless of group size), the level of engagement, learning, and other outcomes may be different depending on the number of people in a group. Lastly, group homogeneity may vary between different courses and levels. While students in undergraduate blended courses may be fairly homogenous, students in graduate DL courses may have a greater variety in terms of background and experiences. Studies have shown that subjects from a narrowly defined population may lead to the formation of groups whose members have the same basic knowledge domains and hence with a group size logically smaller than the physical size (Dennis, Jay F. Nunamaker et al., 1990/91). These issues warrant investigation in future studies.

Besides the possible confounding issues discussed above, the generalizability of the findings on faculty in the online exam is limited. This is due to the small number of faculty who participated in the study. While close to 600 students participated in the experiment, only six faculty members conducted the online exams and were studied in this research. The limited number of faculty prevented extensive data collection and comprehensive analysis of the impact of the online exams on faculty. While all six faculty were interviewed (except for the author), only one recorded the instructor's exam log. The long-term effect of the online exam on faculty's teaching and workload also remains to be further investigated.

Lastly, while the survey instruments were tested as reliable, the reliability of the exam log is a limitation of this study. Despite the convenience of the student exam log provided online, and the instructions to fill out the exam log daily, whether students accurately recorded their exam study activities using the log was unknown. Thus the validity of the results on student exam study details remains to be tested further.

9.5 Future Research

This research can be extended in several areas in the future. One main area of interest is to further study the group process in the collaborative exam. While this study has shown the effectiveness of small group activities in increasing interaction and higher order learning, future studies are needed to more carefully examine what makes group work a good collaborative learning experience. While the analysis in this study was mainly on the individual student level, future research is needed to analyze the phenomena on the group level. It has been observed in this study that some groups in the collaborative exam truly worked together. Members of the group were highly involved in the group process, and they worked together to enhance the quality of the group questions and grading. In contrast, the level of involvement in some groups was very low and members of the group simply participated in the exam individually. Possible areas for further exploration in group process include: group composition (Pimmel, 2003), group communication (Hathorn and Ingram, 2002), and leadership styles (Paulsen, October 2004). For example, the group conferences on Webboard or WebCT can be examined and coded to further examine involvement levels of group members, communication patterns, and leadership styles in different groups.

Another area that requires further exploration, which is already underway, is the effect of the online exam on the faculty. Literature suggests that the role of faculty shifts from the lecturer to facilitator in online education (Berge, 1995; Dzuiban, Shea et al., 2004). This study with a limited number of faculty pointed out that the role of the faculty in online exams changed from the designer of the exam to facilitator of students' learning. Further studies need to be conducted to comprehensively evaluate the role of faculty in facilitating students' learning as well as the workload in conducting the online exams on faculty (Shen and Patten, April 2005).

The study of the use of mobile devices in students' exam studying and their effect on learning is another area that is worth further investigation. While this study reveals that the Internet was the most frequently used medium in exam study, students often used mobile devices in their exam studies. The mobile devices most often used are notebooks and mobile phones. With the rapidly evolving capabilities of mobile computing devices, and with the increasing use of mobile devices in education (Okada, Tarumi et al., April 2001; Rößling, Bär et al., June 2003), it is expected that the use of mobile devices will increase in the near future. It remains to be investigated what effects the mobile devices will have on collaborative learning and online exams as well as students' learning.

One area that requires further exploration but may be more difficult to implement is the long-term effect of the collaborative examination on student's collaboration and learning. While the effect of the collaborative exam in this study was measured only once at the end of the semester, there were some indications that the exam may have long-term effects on students as well as faculty. For example, students who participated in the online exam more than once (due to enrollment in several courses or in multiple semesters) reported more familiarity with the online exam process, spending more time in studying and less time in learning the exam process, and enjoying the exam process more. Similarly, faculty who participated in the exam process more than once also reported a reduction in time spent in conducting the exam. It would be interesting and useful to examine the impact of the collaborative exam using longitudinal studies with the exam implemented in multiple courses in multiple semesters. A longitudinal study would also be useful to more systematically examine student's actual learning, which despite a tendency to favor collaborative learning, did not match with the perceived learning result in this study.

APPENDIX A

CONSENT FORM

Appendix A contains the consent form used in this study. Student subjects were asked to sign the form before the experiment began.

NEW JERSEY INSTITUTE OF TECHNOLOGY
323 MARTIN LUTHER KING BLVD.
NEWARK, NJ 07102

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

TITLE OF STUDY: Collaborative Examination in Asynchronous Learning Networks

RESEARCH STUDY:

I,____, have been asked to participate in a research study under the direction of Jia Shen. Other professional persons who work with her as study staff may assist to act for her.

PURPOSE:

The purpose of this study is to find out, in a scientific manner, students' opinions of the examination formats in this class.

DURATION:

My participation in this study will last from a few days before the exam to a few days after the exam (until I submit the questionnaires, or until I decide not to submit them).

PROCEDURES:

I have been told that, during the course of this study, the following will occur:

1. I will be asked to voluntarily complete one questionnaire before the exam.
2. I will be asked to voluntarily keep a time log of my exam activities during the exam.
3. I will be asked to voluntarily complete one questionnaire after the exam.
4. I may also be asked to voluntarily participate in an interview after the exam.

PARTICIPANTS:

I will be one of about 140 participants in several classes and sections to participate in this study.

EXCLUSIONS:

I will inform the researcher if any of the following apply to me:

- I do not wish to complete the questionnaires for any reason.

RISK/DISCOMFORTS:

I have been told that the study described above may involve the following risks and/or discomforts:

- None known or anticipated

There also may be risks and discomforts that are not yet known.

CONFIDENTIALITY:

Every effort will be made to maintain the confidentiality of my study records. Officials of NJIT will be allowed to inspect sections of my research records related to this study. If the findings from the study are published, I will not be identified by name. My identity will remain confidential unless disclosure is required by law.

PAYMENT FOR PARTICIPATION:

I have been told that I will receive no monetary compensation for my participation in this study. However, extra credits towards my course grade will be awarded for my complete participation in the study.

CONSENT AND RELEASE:

I fully recognize that there are risks that I might be exposed to by volunteering in this study which are inherent in participating in any study. I understand that I am not covered by NJIT's insurance policy for any injury or loss I might sustain in the course of participating in the study.

RIGHT TO REFUSE OR WITHDRAW:

I understand that my participation is voluntary and I may refuse to participate, or may discontinue my participation at any time with no adverse consequence. I also understand that the investigator has the right to withdraw me from the study at any time.

INDIVIDUAL TO CONTACT:

If I have any questions about my treatment or research procedures, I may discuss them with the principal investigator. If I have any addition questions about my rights as a research subject, I may contact:

Richard Greene, M.D., Ph.D., Chair, IRB (973) 596-3281

SIGNATURE OF PARTICIPANT

I have read this entire form, or it has been read to me, and I understand it completely. All of my questions regarding this form or this study have been answered to my complete satisfaction. I agree to participate in this research study.

Student Name:______________________________

Signature:__________________________________

Date:______________________________________

SIGNATURE OF INVESTIGATOR OR RESPONSIBLE INDIVIDUAL

To the best of my knowledge, the participant,

____, has understood the entire content of the above consent form, and comprehends the study. The participants and those of his/her parent/legal guardian have been accurately answered to his/her/their complete satisfaction.

Investigator's Name: Jia Shen

Signature:__________________________________

Date:______________________________________

APPENDIX B

RESEARCH INSTRUMENTS – STUDENTS

Appendix B contains research instruments used to collect data from student subjects in this study. The documents include pre-exam student questionnaire (Appendix B.1), student exam log (Appendix B.2), post-exam student questionnaire (Appendix B.3), and post-exam student interview guide (Appendix B.4).

B.1 Pre-exam Student Questionnaire

Appendix B.1 contains the pre-exam student questionnaire used in the study. The actual pre-exam survey was conducted online, and the paper version below shows all the questions in their actual order as displayed in the online survey. The presentation of the online survey was slightly different from the paper version. The consent form as shown in appendix A was displayed after the student logins the survey system. The online survey used Active Server Pages technology to dynamically customize questions for different courses (e.g. showing different course names). In the paper version below, assume the student selected CIS675 on the login page.

Thank you for your participation in the Student Examination Study at NJIT. As the first step of your participation, you will fill out this pre-exam questionnaire, which will take approximately 15 minutes. This questionnaire has **two** parts and a consent form. Please make sure you print your name in the consent form and answer **ALL** the questions in part 1 and 2.

Identifying Information:

The identification information is elicited for the purpose of matching your pre-exam survey results with the post-exam survey results. All information collected in this study will be confidential and will be used for research purpose only. Your responses will NOT affect your course grade in any way.

First Name: ______________________
Last Name: ______________________

Your Email: ______________________ (Optional)
(You will be contacted at this email address within 24 hours if your data are not correctly recorded in the database.)

I am in ________ (please select one course)
Course list: *(Omitted)*

(Note: If you are taking **more than one course** from the list above, please complete the survey for each course individually.)

Consent Form
(As shown in appendix A, omitted)

Part 1. Your Learning Approach

The following 18 questions ask about the approach you have used so far in studying the course you selected above. Please choose the one most appropriate response to each question on a 7-point scale from Strongly Disagree to Strongly Agree. Fill the answer that best fits your immediate reaction. **Do not spend a long time on each item: your first reaction is probably the best one.** Do not worry about projecting a good image. Your answers are strictly confidential and will NOT affect your course grade in anyway.

	Strongly Disagree						Strongly Agree
1. My aim is to pass CIS675 while doing as little work as possible.	1	2	3	4	5	6	7
2. I find that at times studying CIS675 gives me a feeling of deep personal satisfaction.	1	2	3	4	5	6	7
3. I keep my work to the minimum because I do not find CIS675 very interesting.	1	2	3	4	5	6	7
4. I feel that virtually any topic in CIS675 can be highly interesting once I get into it.	1	2	3	4	5	6	7
5. I find I can get by in most exams in CIS675 by memorizing key sections rather than trying to understand them.	1	2	3	4	5	6	7
6. I find that studying CIS675 can at times be as exciting as a good novel or movie.	1	2	3	4	5	6	7
7. I find it is NOT helpful to study CIS675 topics in depth because it wastes time, when all you need is a passing acquaintance with topics.	1	2	3	4	5	6	7
8. I work hard at CIS675 because I find the material interesting.	1	2	3	4	5	6	7
9. I see no point in learning material that is not likely to be in the CIS675 examination.	1	2	3	4	5	6	7
10. I come to most CIS675 classes with questions in mind that I want answered.	1	2	3	4	5	6	7
11. I enjoyed working with other students on group assignments/projects in CIS675.	1	2	3	4	5	6	7
12. I would rather have done the group assignments/projects in CIS675 individually.	1	2	3	4	5	6	7
13. I have learned a lot from other students in our group assignments/projects in CIS675.	1	2	3	4	5	6	7
14. The help I got from other students in this course was useless or misleading.	1	2	3	4	5	6	7
15. Students in my CIS675 class tend to be very cooperative in sharing knowledge and learning together.	1	2	3	4	5	6	7
16. Contact with other students in CIS675 has not played an important part in my learning.	1	2	3	4	5	6	7
17. Interacting with other students has played an important role in my learning in CIS675.	1	2	3	4	5	6	7
18. Students in my CIS675 class tend to be very competitive against each other.	1	2	3	4	5	6	7

Part 2. Background Information

1. Your gender:

___ Male ___ Female

2. Is English your native or first language?

__ No __ Yes

3. I am a

___ Undergraduate student. Years in undergraduate program ___________

___ Graduate student. Years in graduate program _______________

4. I have ____ years of work experience related to this course.

____ None (The is my first exposure to this area.)

____ Less than 1 year

____ 1- 3 years

____ 4 - 9 years

____ 10-15 years

____ More than 15 years

5. How many courses have you taken that use WebBoard, WebCT or a similar system before this course?

___ None (This is my first course with an online component)

___ One other course

___ 2-4 other courses

___ 5 or more other courses

6. I expect my exam grade to be (select one range):

1	2	3	4	5	6	7	8	9
100 | 96	95 | 91	90 | 86	85 | 81	80 | 76	75 | 71	70 | 66	65 | 61	Below 60

7. Previous Online Collaborative Exam Experience

A series of studies have been conducted at NJIT in recent years on an online examination process called Collaborative Examination. Different from a traditional exam where instructors make up questions and grade students' answers, the online Collaborative Exam allows student participation in various steps of the examination process, such as making up exam questions and grading others' answers, using Webboard or WebCT. Please select one of the following:

____ **I never heard about the "Collaborative Exam" before this course.**

____ **I have heard about the "Collaborative Exam" from other students but I have never experienced it myself.**

____ **I did the complete "Collaborative Exam" before, in which I contributed questions and graded others' answers. Please list the course numbers and semesters when you took the exam at NJIT:**

Course Number ________Semester and Year: ________ Section: __DL __FTF
Course Number ________ Semester and Year: ________ Section: __DL __FTF

____ **I did a partial "Collaborative Exam" before, in which I contributed questions. Please list course numbers and semesters when you took the exam at NJIT:**

Course Number ________ Semester and Year: ________ Section: __DL __FTF
Course Number ________ Semester and Year: ________ Section: __DL __FTF

Thank you for completing the pre-exam questionnaire!

Thank you for your participation in this study! You have just finished the first step in the Student Examination Study. In the next few days you will receive information regarding the following steps in participating in this study, including keeping a record of your exam-related activities during the process, and completing a post-exam survey right after the exam. Information will be posted on your course Webboard/WeCT. Your complete and thorough participation in these following steps is important to us. Thank you very much in advance! If you have any comments or questions regarding the survey, please email Jia Shen (jxs1866@njit.edu).

B.2 Student Exam Log

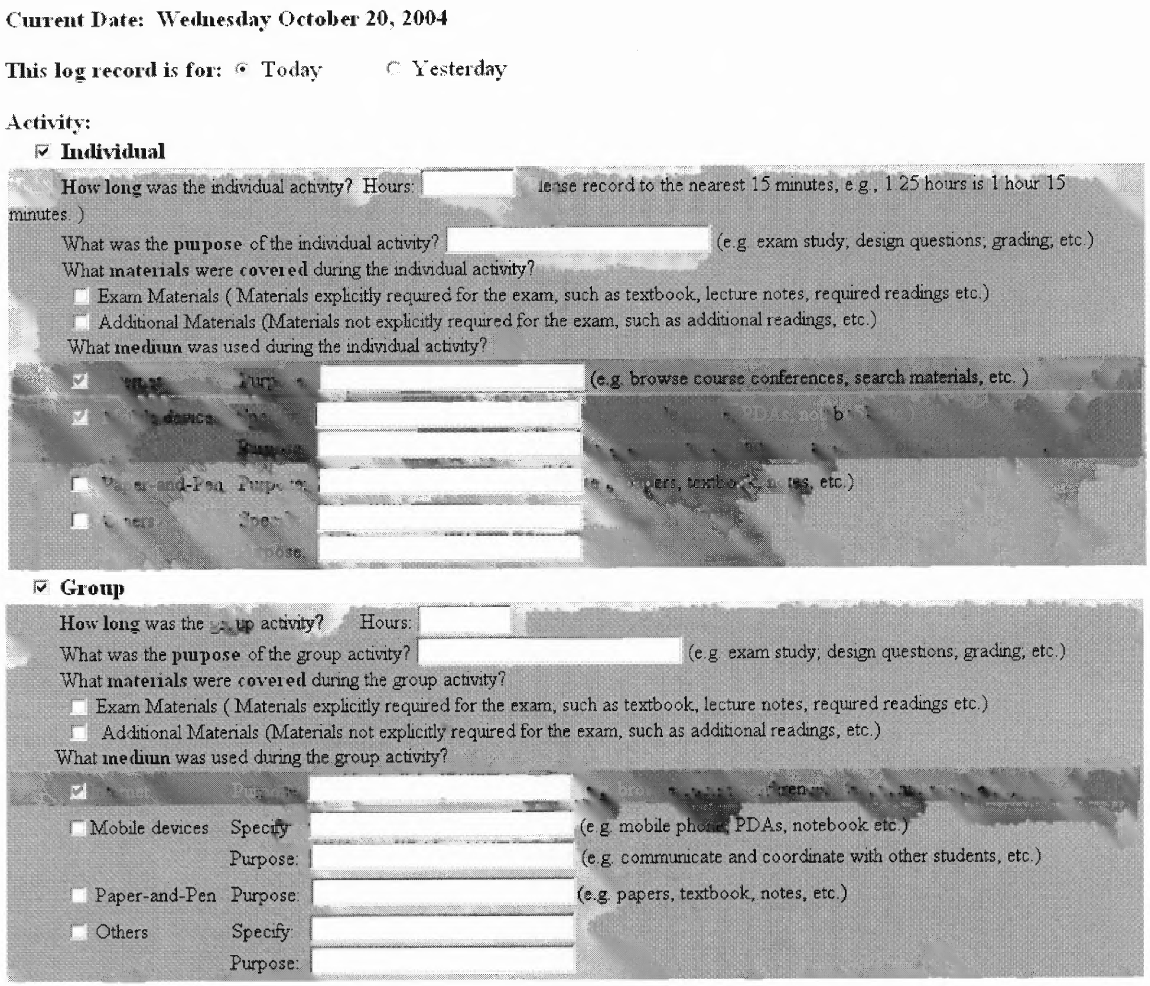
Appendix B.2 contains the student exam log used in this study. The actual log was provided online. This section shows the instructions for the log and the log form as it is displayed online.

Please use the following form to record your exam studying activities in (course name). You may fill out the log many times daily, however you can only record your activities for **today and yesterday** to ensure the accuracy of the log.

In the log form below, please first indicate whether the study activity was an individual or group activity (you may select both types), then for each of the them, indicate the duration (in hours), purpose, study materials (e.g., textbook, lecture notes, required readings; or additional outside materials), and what medium was used in the activity (e.g., PDAs, mobile phones, wireless network, traditional paper and pen, or any other devices).

Do not worry about projecting a good image. Please be completely honest. Your information is strictly confidential. The data collected will be used for research purposes ONLY and will not affect your course grade in any way.

Note if you are taking the online exam, you need to keep a record of your preparation before the exam and your exam activities during the online exam process. If you are taking the traditional in-class exam, you only need to keep a record of your exam preparation till the exam date.



Current Date: Wednesday October 20, 2004

This log record is for: Today Yesterday

Activity:

Individual

How long was the individual activity? Hours: [illegible] lease record to the nearest 15 minutes, e.g., 1.25 hours is 1 hour 15 minutes.)

What was the **purpose** of the individual activity? (e.g. exam study, design questions, grading, etc.)

What **materials** were **covered** during the individual activity?

Exam Materials (Materials explicitly required for the exam, such as textbook, lecture notes, required readings etc.)

Additional Materials (Materials not explicitly required for the exam, such as additional readings, etc.)

What **medium** was used during the individual activity?

[illegible] (e.g. browse course conferences, search materials, etc.)

[illegible] PDAs, not [illegible]

Paper-and-Pen [illegible] pers, textbo[illegible] tes, etc.)

[illegible]

Group

How long was the [illegible] activity? Hours:

What was the **purpose** of the group activity? (e.g. exam study, design questions, grading, etc.)

What **materials** were **covered** during the group activity?

Exam Materials (Materials explicitly required for the exam, such as textbook, lecture notes, required readings etc.)

Additional Materials (Materials not explicitly required for the exam, such as additional readings, etc.)

What **medium** was used during the group activity?

[illegible]

Mobile devices Specify (e.g. mobile phone, PDAs, notebook etc.)

Purpose: (e.g. communicate and coordinate with other students, etc.)

Paper-and-Pen Purpose: (e.g. papers, textbook, notes, etc.)

Others Specify

Purpose:

Figure B.1 Student log form online.

The “Internet” and “Mobile device” options in the “individual” activity section and the “Internet” option in the “group” section of the log form are selected and highlighted as an example.

B.3 Post-exam Student Questionnaire

Appendix B.3 contains the post exam questionnaire used in this study. The actual post-exam survey was conducted online, and the paper version shows all the questions in the actual order as they appeared in the online survey. The presentation of the online survey was slightly different from the paper version. The online survey used Active Server Pages technology to dynamically generate three post surveys based on the exam mode the students selected on the login page: traditional; online individual, i.e., participatory online exam; online small group, i.e., collaborative online exam). In the paper version below, C2 denotes questions shown only to the participatory exam students. C3 denotes questions shown only to the collaborative exam students. If there is no notation, then the question is common to all three modes. In addition, some of questions were customized for different courses, which are denoted using quotation (“”) in the paper version below.

The following questions ask about the strategies you used in studying for the final exam. Please choose the one most appropriate response to each question on a 7-point scale from Strongly Disagree to Strongly Agree. Fill the answer that best fits your immediate reaction. **Do not spend a long time on each item**: your first reaction is probably the best one. **Do not worry about projecting a good image**. Your answers are strictly confidential and will NOT affect your course grade in anyway.

Please note "**exam studying**" in the questions below includes studying before the exam and studying when you answered questions in class/ includes studying before the exam and studying during question design, answering, and grading phases.

	Strongly Disagree						Strongly Agree
1. In studying for the CIS675 exam, I learned most of the things by rote, going over and over them until I knew them by heart even if I did not understand them.	1	2	3	4	5	6	7
2. When I was studying for the exam, I put myself in the position of "an Information Systems Evaluation researcher" to try to understand his/her role in organizations.	1	2	3	4	5	6	7
3. I found the best way to pass the CIS675 final exam I participated in is to try to remember answers to likely questions.	1	2	3	4	5	6	7
4. When I was studying for the exam, I found that I could relate CIS 675's material to other subject areas.	1	2	3	4	5	6	7
5. I spent a lot of time finding out more about interesting topics which have been discussed in CIS675 when I was studying for the exam.	1	2	3	4	5	6	7
6. I restricted my study to what was specifically set in the CIS675 final exam scope, as I thought it was unnecessary to do anything extra.	1	2	3	4	5	6	7
7. I found that I had to do enough work on a topic so that I could form my own conclusions before I was satisfied in studying for the CIS675 exam.	1	2	3	4	5	6	7
8. I tested myself on important CIS675 topics until I understood them completely while studying for exam.	1	2	3	4	5	6	7
9. When I was studying for the CIS675 exam, I found most topics interesting and spent extra time trying to obtain more information about them.	1	2	3	4	5	6	7

Use of Mobile Devices (select as many answers as appropriate)
1. Have you used any of the following mobile computing devices in studying for the final exam (e.g., taking notes, coordinating study meetings, communication with others, etc.)?
______ Mobile phones (voice)
______ Text messaging using mobile phones
______ PDAs
______ Wireless Notebook computers
______ Other (Please specify _______text________)

2. If you selected any of the above, what did you use the mobile devices for?
______ Taking notes for myself
______ Scheduling and coordinating meetings with classmates
______ Discussing topics for the exam
______ Other (Please specify ______text_________)

Part 2 of 4 – Exam Studying Strategies (continued)
In these questions, the "**exam process**" includes "includes your exam preparation before the exam and answering questions in class."/ "includes your exam preparation before the exam and question design, answering, and grading phases online."

1. The exam process was mainly a(n):
Individual experience 1----2----3----4----5----6----7 Group experience

	Strongly Disagree						Strongly Agree
2. My understanding of course material was enhanced by interacting with other students in the class through the exam process.	1	2	3	4	5	6	7
3. I felt my relationship with other students in the class was mainly competitive during the exam process.	1	2	3	4	5	6	7
4. Anything I learned through the exam, I learned on my own.	1	2	3	4	5	6	7
5. The CIS 675 final exam allowed me to form a kind of learning community with other students.	1	2	3	4	5	6	7
6. I was able to get to know some students better during the final exam process.	1	2	3	4	5	6	7

(C3) Please answer the following 10 questions based on your experience with your exam group only:
7. How satisfied or dissatisfied were you with the quality of your group's questions and grades?
Very Dissatisfied 1----2----3----4----5----6----7 Very Satisfied
8. There was a high degree of participation on the part of members.

Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

9. The issues discussed during the group's discussions were:

Trivial 1----2----3----4----5----6----7 Substantial

10. Did all group members participate equally in the discussions?

Equal participation 1----2----3----4----5----6----7 Unequal participation

11. One person influenced the group's work more than the rest of the group.

Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

12. Members worked together as a group.

Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

13. To what extent did you trust the members in your group?

No trust at all 1----2----3----4----5----6----7 Great deal of trust

14. The group process uncovered valid alternatives that I had not considered.

Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

15. To what extent were the people in your group helpful?

Not helpful at all 1----2----3----4----5----6----7 Very helpful

16. I experienced pressure, either to conform to a particular viewpoint or not to contradict others.

Not at all 1----2----3----4----5----6----7 Very much

Anonymity

(C2, C3) 1. Did you try to guess the person who authored or graded the questions?

___ Yes ___ No

(C2, C3) 2. Have you shared your question IDs with anyone else?

___ Yes ___ No

(C2, C3) 3. I like the anonymity in the exam.

___ Yes ___ No

(C3) 4. I like the anonymity in my small-group discussion.

__ Yes __ No ___Not applicable (Our group discussion was not anonymous)

Part 3 of 4 – Exam Outcomes

In answering the questions below, please think about the exam you have just experienced in your course.

	Strongly Disagree						Strongly Agree
1. My knowledge of major concepts, methods, and theories of CIS 675 has <u>NOT</u> been improved through the exam.	1	2	3	4	5	6	7
2. I enjoyed the examination process.							
3. My **understanding of the meaning** of CIS 675 course material has enhanced through the exam.	1	2	3	4	5	6	7
4. I felt the grading process was fair.							
5. My ability to **use** methods, concepts, and theories I learned in CIS 675 in new situations has	1	2	3	4	5	6	7

been improved through the exam.	
6. The instructor organized the exam very well.	
7. My ability to **solve problems** using what I learned in CIS 675 has NOT been improved through the exam.	1 2 3 4 5 6 7
8. I felt under a lot of pressure taking the exam this way.	
9. I am better able to **see** different course **components** of CIS 675 and organize them in a meaningful way through the exam.	1 2 3 4 5 6 7
10. The grading criteria given by the professor were explicit enough.	
11. My ability to **recognize patterns** of CIS 675 course material and their underlying meanings has remained the same as before the exam.	1 2 3 4 5 6 7
12. The exam provided a comfortable timeframe.	
13. My skill to **relate knowledge from several areas** to make my argument has been improved through the CIS675 exam.	1 2 3 4 5 6 7
14. The exam did NOT allow me to demonstrate what I learned in class.	
15. My ability to use course material to **make generalizations or predictions** has been improved through the CIS 675 exam.	1 2 3 4 5 6 7
16. The final grade that I received on this exam was NOT fair.	
17. My skill to **compare and discriminate** between ideas has been improved through the CIS675 exam.	1 2 3 4 5 6 7
18. I learned from reading the exam questions posted online.	
19. My ability to **judge** the value of ideas and **assess** the quality of arguments has been improved through the CIS675 exam.	1 2 3 4 5 6 7
20. I learned from answering the exam questions.	
21. I would recommend using the exam version that I did in this course in the future.	

(C2, C3) 22. I did NOT learn from making up questions.
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 23. I learned from reading other's answers.
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 24. I did NOT learn from grading students' answers.

Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 25. I learned from reading grading posted online.
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 26. I do NOT believe CIS 675 students were able to design questions of good quality for the exam
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 27. The question design criteria given by the CIS 675 professor were explicit enough.
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 28. I do NOT believe CIS675 students were capable of grading the responses to the questions they designed
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

(C2, C3) 29. It would have been an improvement if only professor graded.
Strongly Disagree 1----2----3----4----5----6----7 Strongly Agree

Part 4 of 4 - Overall Comments and Suggestions

1. If given a choice, which version of the exam would you prefer to take?

______ the traditional exam (where the instructor designs questions and grades answers, and students do the exam in class.)

______ the online individual exam (where students individually design questions, answer other students' questions, and grade answers to questions they designed on Webboard or WebCT.)

______ the online small group exam (where students design questions and grade answers to their questions in small student groups, and answer other students' questions individually on Webboard or WebCT.)

1b. Why did you choose the particular exam above?

2. What did you like best about the exam that you participated in?

3. What do you think should be improved in the exam that you participated in?

4. Do you have any other comments about the exam that haven't been covered in this questionnaire?

Thank you for completing the post-exam questionnaire!

B.4 Post-exam Student Interview Guide

Appendix B.4 contains the post-exam student semi-structured interview guide used in the study. Before each interview, the interviewee's data were retrieved from the pre and post exam survey database. Student's background information and survey results were analyzed and noted on the interview sheet before the interview. The notes were used to customize questions in order to best explore each individual's experiences in the exam.

Background
How many years have you been in the graduate/undergraduate program? When will you graduate?

How many DL courses have you taken?

How many courses are you taking this semester?
  Are they all distance-learning courses?
  How many midterm exams did you have this semester?

I know you had the group assignment before the midterm exam. Did you enjoy working with other students on that assignment?

Have you had any previous online exam experience before this course?
  If so, when and which course? Was it similar or different from the cis675 online exam?

Questions
1. What's your impression of the quality of questions, answers, and grading posted by other students in the exam?

2. In terms of time and effort you put into this exam, would you say you spend more or less time, or put in more or less effort, in this exam compared with a traditional in-class exam?

  2.1 How much time, approximately, did you put into each of the main phase of the exam, including question design, answering, and grading?

3. Did you study for this exam the same way as you studied for other exams?
  3.1 What materials did you use in making up questions, answering, and grading? E.g. course textbook, web resources, etc.

  3.2 When you designed questions or answered questions, did you try to synthesize things you learned in this course or other related courses?

  3.3 Did you study with other students for this exam? Why? Was it helpful?

  3.4 To what extent did you use mobile devices such as PDAs, mobile phones, wireless networks and laptops in the exam process?
  (follow-up) What did you use them for? How did you use them?

  (follow-up) Did you find it beneficial to your study and your exam? Why?

4. Tell me about your group in the exam.
  4.1. Were you satisfied with your group? Why?

4.2 Did you learn anything from working with your group? What did you learn?

Overall

1. Overall, do you think the exam you participated is a good exam for this course? Why or why not?

2. What did you learn from the exam, if any, that you think you can apply in your profession in the future?

3. Do you have any suggestions to improve the process for the future?

4. Finally, is there anything else you want to tell me about the exam?

Closing

Than you very much for the interview. You have done a great job and I have learned a lot from you. We have prepared a small gift to you, which is a $10 Barnes and Noble gift card.

APPENDIX C

RESEARCH INSTRUMENTS – INSTRUCTORS

Appendix C contains research instruments to collect data from instructor subjects in this study. The documents include pre-exam instructor meeting notes (Appendix C.1), instructor exam log (Appendix C.2), and post-exam instructor interview guide (Appendix C.3).

C.1 Pre-exam Instructor Meeting Notes

The document included below is the issue list used for meetings with instructors before the exams.

This issue list is prepared before meeting with faculty in discussing incorporating collaborative / participatory exam into their existing courses. This is a list of things to consider in order to incorporate the collaborative / participatory exam in replacement of or in addition to the existing assessment methods. The overall objective of the initial and follow-up meetings with instructors is to discuss how a reasonable workload can be maintained for both faculty and students while maximizing students' learning through the collaborative online examination.

1. In blended courses, find out how much is the course online component, what has been done online before the exam, and the frequency of using the web.

2. Find out what the existing course exam/assignments/activities are. Pay special attention to activities that involve teamwork. Inquire the possibility to have 3 to 4 students in the projects or team assignments so the teams can be used later in the collaborative exams.

3. Adjust the existing timeline of the course to fit in the collaborative exam.

4. Adjust the percentage of the exam where the collaborative/participatory exam is adopted in the overall course grading.

5. In the collaborative/participatory exams, decide on question format, number of questions each student designs and answers.

6. In traditional exam, decide on the access to resources.

7. Revise and approve the student's instruction document provided by Jia Shen, including the exam scope, question design criteria, grading guidelines, etc.

8. Review the instructor's manual provided by Jia Shen before conducting the exam.

9. Discuss the alternative assignment to the experiment for students to get the extra credits.

C.2 Instructor Exam Log

Appendix C.2 contains the instructor's exam log. The log was given to the instructor as part of the instructor's guide document (see appendix E) before the exam started.

To help the research team investigate the time and effort instructors spend in conducting the online small group exam on WebCT, you are invited to record your exam related activities using the form below. Please record the date, and the duration (in hours) of your activities. Your activities are classified into the following five categories:

- **Administration** – this includes any effort in administering the exam. For example posting announcements, customizing guidelines for your course, assigning students with IDs, posting grades, etc. Notice this category refers to activities that only involve yourself, and it does not involve two-way communication with students.
- **Communication with students** – this refers to any activities that involve two-way communication with your students, such as answering students' questions, etc.
- **Question** – this refers any effort related to questions for the exam. For example, time spent reviewing all questions posted by students in online exams, etc.
- **Grade** – this refers to any effort related to grading students' answers. For example, time spent in reviewing and grading answers, and resolving disputes.
- **Other** – any other effort you put into the exam that is not listed in the categories above.

Your thorough completion of the log will greatly help us to investigate and reduce the workload for instructors in future studies. Thank you in advance!

Date	**Activity**				
	Admin.	**Communicate**	Question	**Grade**	**Other**
9/30/04		0.5 hour	1 hour		

The first line is filled in as an example.

C.3 Post-exam Instructor Interview Guide

Appendix C.3 contains the post exam instructor interview guide used in this study. Before each interview, the online exam WebCT/Webboard systems were carefully examined, which contains the record of the exam discussions, etc. General questions were customized for each instructor based on the individual course and the exam process to best explore the instructor's experiences in the exam.

Introduction

Thank you for agreeing to participate in this interview! This interview will last about 40 minutes. I will ask you a few questions about your experience as an instructor in the exam in _________ (course number and name). Please be completely open in your answers. Your comments will help us to understand how the exam works for instructors, and it will help us to improve examination design for future courses. This conversation is confidential.

Questions

Actual order and wording of the question may change depending on the interview.

1. Do you think the online exam as we designed is a good way to assess students' learning for this course? Why or why not?

2.

   a. On a 1 to 7 scale, how do you rate the overall quality of students' questions?

   b. On a 1 to 7 scale, how do you rate the overall quality of students' answers?

   c. On a 1 to 7 scale, how do you rate the overall quality of students' grading?

   d. What suggestions do you have to improve the quality of students' questions, answers, and grading? For exam, guidelines etc.

   e. What suggestions do you have to enhance students' learning/experience in the exam process?

3. In terms of grading:
   a. Was it more or less work in grading the online exam compared with your usual exam?

   b. How many students disputed their grades, roughly. How did you handle them? Was it time consuming?

4. Adaptation to the new exam:
   a. Student: What reactions did you get from your students when you introduce the exam methods to them?

   b. Instructor:
      a. Was it a lot of work to adapt the online exams into your existing course?

      b. Did you experience any difficulty in administering the online exams? (e.g., with the system)

5. How could the software better support either you or your students? If the software were improved in your way, would you more likely to use the online exam in the future in this or other courses? Why?

6. Did you learn more about your students and the course through the online exam, compared with the traditional exam?

7. Would you use the online exam in the future in this or other courses? Why?

Closing

Thank you very much for the interview! Here is a gift to show our appreciation to your participation in this research. Thank you and have a good day!

APPENDIX D

STUDENT EXAM INSTRUCTIONS

Appendix D contains the exam instructions given to student subjects in this study. Three versions of the instructions were designed and distributed to students depending on the exam mode they were in. The documents include student exam instructions for the traditional exam (Appendix D.1), the participatory exam (Appendix D.2), and the collaborative exam (Appendix D.3).

D.1 Student Exam Instructions – Traditional Exam

Appendix D.1 contains a document as the sample of student exam instructions in the traditional exam. The document was given to students in MIS635 in spring 2004, who participated in the traditional exam. Slight modifications were made in the instructions for other courses in the same exam mode, including exam scope, examples, etc.

MIS 635 In-class Final Exam

MIS 635 is participating in the Student Examination Research Project at NJIT. The in class section of MIS 635 will participate in the In-class Exam, which is a 3-hour proctored exam with open course materials on Thursday 4/29. This document describes the schedule of the exam, including detailed procedures, the grading scheme, and detailed grading guidelines.

Schedule

The final exam of MIS 635 will be conducted on Thursday 4/29. The following table lists dates and each step in the exam process. *Steps in italics (1, 3, and 7)* are optional participation in the research project. Your participation will be awarded with up to 12 extra credits towards the total exam grade. The alternative way to gain the extra credits is provided if you choose not to participate in the exam research, which is explained in the course WebCT.

1. Pre-exam survey	*Thursday 4/8 - Sunday 4/11*
2. Class members receive detailed exam instructions (this document)	By Wednesday 4/14
3. Class members obtain log template to record exam related activities	*Wednesday 4/14*
4. Class members review candidate exam questions	Thursday 4/22- Wednesday 4/28
5. Exam at NJIT or Mt. Laurel campuses or through proctor	Thursday 4/29
Professor posts final grades	By Monday 5/3
6. Class members may contest final grades	Tuesday 5/4
Professor resolves grade contests	Wednesday 5/5
7. Post-exam survey and return of log file	*Thursday 5/6 – Sunday 5/9*

1. Pre-exam survey: Thursday 4/8 - Sunday 4/11 (completed)

As the first step in the participation of the examination research project, you are invited to fill out a pre-exam survey. The survey is optional. You will receive 4 extra points towards your final exam grades for completing the survey. The pre-exam survey is conducted online and the URL has been posted by the professor on your course WebCT site before the exam.

Note that the survey data will be used strictly for research purpose and will be confidential. Your answers in the survey will not affect your course grades in any way.

2. Detailed Exam Instructions: by Wednesday 4/14

Detailed exam instructions are provided to you in this document.

3. Log template: by Wednesday 4/14

As the second step in the participation of the examination research project, you are invited to keep a record of your exam studying activities using a log file. Participation is optional. You will receive 4 extra points towards your final exam grades for filling the log file. The log file template will be posted on your course WebCT site by Wednesday 4/14.

4. Reviewing candidate questions: Thursday 4/22- Wednesday 4/28

The final exam mainly covers MIS 635 materials from lecture 7 to lecture 11, and sections in lecture 5 and 6 that are closely related to this set such as scale development and statistics. A list of candidate exam questions will be posted in the class final exam conference on the course WebCT site from Thursday 4/22 to Friday 4/30. You should review all questions as part of your preparation for the exam. Two questions from the list will be selected by the professor to appear on the final exam. You will notice that each question is associated with a level number ranging from 1 to 6. The number indicates the main **objective** of the question in terms of testing:

1. Knowledge of specifics- knowing major concepts, methods, and theories
2. Comprehension- understanding of major concepts, methods, and theories
3. Application - using theories, methods, and concepts in new context
4. Analysis – analyzing and solving problems
5. Synthesis – relating materials from several areas to make argument
6. Evaluation – judging the value of ideas and assessing quality of arguments

One of the two questions in the final exam will be on level 1 or 2, and another one will be on level 3 and up.

5. Proctored Exam: Thursday 4/29

The final exam will be offered on the NJIT campus during the class time on Thursday 4/29.

These are the instructions you will find on the final exam on 4/29:

- Answer each of the two questions. The question on level 1 or 2 counts 40 points. The question on level 3 or up counts 60 points.
- This exam is open book and open notes. You may bring the textbook, course notes, papers, and other materials.
- Be sure to incorporate the class readings and others whenever appropriate, stating the name of the author or title when you mention the reading.
- Take your time. You have 3 hours to answer the two questions.
- Please limit your answer to no more than one exam book per question for both questions.
- Be sure to read over your answers carefully.

Your professor will grade answers and post final exam grades on the course WebCT site by Monday 5/3.

6. Contesting the final grade: Tuesday 5/4

If you disagree with the final grade from your professor, then you may dispute it by Tuesday 5/4, after receiving the final grade from the professor. Note your dispute may result in **either increase or decrease** of your answer grades, when the instructor reevaluates all aspects of the question, answer, and grading. Notice:

1. Ensure that you are disputing a significant number of points (i.e. greater than 5 points total per answer disputed). The professor will not reconsider a small number of points.
2. Make sure that you provide compelling justification for the re-grade. Also state why you believe the grade was wrong in each category where you dispute the grade.

7. Post-exam survey and return of log files: Thursday 5/6 – Sunday 5/9

As part of your third and final step in the participation of the examination research project, you are invited to fill out a post-exam questionnaire. The survey is optional. You will receive 4 extra points towards your final exam grades for completing the survey. Similar to the pre-exam survey, the post-exam survey will be conducted online. The URL will be posted by the professor right after the exam. Please also return the log files by Sunday 5/9. Details of returning the log file will be posted on course WebCT site.

Note that the survey data will be used strictly for research purpose and will be confidential. Your answers in the survey will not affect your course grades in any way.

Grading Scheme

Your final exam grade will based on your answers to the two questions in the final exam with a total of 100 points. The grading guideline is further explained in the section below. You should read the grading guideline before the exam.

As pointed out before, 12 extra points towards your exam grade will be provided for your participation in the pre-exam survey (4 points), log record (4 points), and post-exam survey (4 points). As an alternative to participation in the research, you may write a research paper to gain the extra credits, which is further explained in the course WebCT. The maximum total of the final exam is 112.

Answer Grading Criteria– will be used by the professor

(Total 100 points)

For answers to questions on level 1 or 2 (40 points total):

- 25 points: The correctness and completeness of the answer, including citing most of all relevant course materials etc.
- 8 points: The quality and clarity of the writing.

- 5 points: Following editing guidelines including using citation (deduct up to 5 points), and length (deduct 2 points if answer exceeds one exam book per question).
- 2 points: Readability and clearness in hand writing.

For answers to questions on level 3 or up (60 points total):

- 40 points: The correctness and completeness of the answer, including considering all sides of issues, synthesizing material etc.
- 15 points: The clarity and quality of the writing, including having a framing paragraph to open, providing justification to points, etc.
- 3 points: Following editing guidelines including using citation (deduct up to 3 points), and length (deduct 2 points if answer exceeds one exam book per question).
- 2 points: Readability and clearness in hand writing.

D.2 Student Exam Instructions – Participatory Exam

Appendix D.2 contains a document as the sample of student exam instructions in the participatory online exam. The document was given to students in CIS433 in fall 2004, who participated in the participatory online exam which was conducted on Webboard with three questions required per student. Slight modifications were made in the instructions for other courses in the participatory exam, including exam scope, question type, examples, changing examples to WebCT when necessary, etc.

CIS 433 - 001 Fall '04
Midterm Schedule - *Student's Quick Overview*

Date		Student Activities		Date	Instructor Activities
Oct. 3	Sun		1. Pre-Exam Survey	Oct. 3	
4	Mon			4	
5	Tues			5	
6	Wed			6	
7	Thurs	2. Obtain exam instructions and IDs		7	Post exam instructions; Assign IDs
8	Fri			8	
9	Sat		3. Exam Log	9	
10	Sun			10	
11	Mon			11	
12	Tues			12	
13	Wed	4. Question Design		13	
14	Thurs			14	
15	Fri			15	
16	Sat	5. Question Review		16	Review/Assign questions
17	Sun			17	
18	Mon	6. Answering		18	
19	Tues			19	
20	Wed			20	
21	Thurs	7. Grading		21	
22	Fri			22	
23	Sat			23	
24	Sun			24	Review/Post final grades
25	Mon			25	
26	Tues	8. *Optional: Grade Dispute*		26	
27	Wed			27	Resolve disputes
28	Thurs		9. Post-Exam Survey	28	
29	Fri			29	
30	Sat			30	
31	Sun			31	

Note: Please use the number before each step to look for detailed instructions in the document.

CIS 433 - 001 Online Individual Midterm Exam
Student Instructions

CIS 433 is participating in the Student Examination Research Project at NJIT. The fall '04 class will participate in the Online Individual Exam during the midterm using the course Webboard. This document describes the schedule of the exam, including detailed procedures and grading guidelines.

Schedule

The main activities of the midterm will take place over a 2 week period from **Wednesday 10/13 to Tuesday 10/26**. The following table lists dates and each step in the exam process. *Steps in italics (1, 3, 9, and 10)* are optional participation in the research project. Your participation will be awarded with up to 12 extra credits towards the total exam grade.

1. Pre-exam online survey (approx. 15 minutes)	*Sunday 10/3 – Wed. 10/6*
2. Class members receive detailed exam instructions (this document) and question IDs for anonymity	By Saturday 10/9
3. Class members obtain online log URL to record exam related activities (approx. 5 minutes each time)	*Thursday 10/7*
4. Class members design and post questions in the Midterm Conference	Wednesday 10/13-Friday 10/15
Professor grades/edits and assigns questions	By Sunday 10/17
5. Class members review candidate exam questions	By Sunday 10/17
6. Class members answer questions and submit in the Midterm Conference on Webboard and on Turnitin	Monday 10/18- Wednesday 10/20
7. Class members grade answers to their questions and post grading in the Midterm Conference	Thursday 10/21- Saturday 10/23
Professor posts final answer grades	By Monday 10/25
8. Class members may contest final answer grades *Exam log activity ends*	Tuesday 10/26
Professor resolves contested grades	By Wednesday 10/27
9. Post-exam online survey (approx. 25 minutes)	*Thursday 10/28 – Sunday 10/31*
10 Selected student will be interviewed	*By Friday 11/12*

Webboard Conferences

The following conferences will be used for the midterm:

- Midterm Exam Announcements – public; the professor will post instructions, students may post questions regarding the exam procedure etc.
- Midterm Conference – public; where the questions, answers, and grading will be posted anonymously

See details on the use of these conferences in the detailed steps below.

Detailed Steps

1. Pre-exam survey (completed)

As the first step in the participation of the examination research project, you are invited to fill out a pre-exam survey. The survey is optional. You will receive 4 extra points towards your exam grade for completing the survey. Note all the research data collected in this study are confidential and will be used strictly for research only.

2. Detailed Exam Instructions and IDs for anonymity

Detailed exam instructions are provided to you in this document. Your professor will notify you of your three Question IDs (e.g. Q11, Q12, Q13) on Webboard. You will use the question IDs when posting questions in the public Midterm Conference so the process can remain anonymous.

3. Log template

As the second step in the participation of the examination research project, you are invited to keep a record of your exam studying activities using a log template online. Participation is optional. You will receive 4 extra points towards your exam grade for keeping a thorough log. The log URL will be announced on Webboard.

4. Posting questions

By 10/15, you should post three insightful questions in the Midterm Conference.

Questions must cover materials specific to CIS 433 from textbooks, lectures, and additional course readings. **Questions should be based on materials covered up to the lecture on 10/12**. Deign one question for each of the following three types:

- Coding or providing program result problem using HTML and/or JavaScript (similar to those appear in class quizzes)
- Coding or providing program result problem using ASP.net (similar to those appear in class quizzes)
- Short essay question on definitions of terminology or concepts of E-Commerce.

You should indicate the main **objective** of each question using the following categories in terms of testing:

1. Knowledge of specifics- knowing major concepts, methods, and theories
2. Comprehension- understanding of major concepts, methods, and theories
3. Application - using theories, methods, and concepts in new context
4. Analysis – analyzing and solving problems
5. Synthesis – relating materials from several areas to make argument
6. Evaluation – judging the value of ideas and assessing quality of arguments

You may have sub-questions in your questions, and you may indicate multiple objectives in each of your questions (e.g., Q1 on level 1 and 3 with sub q1 on level 1 and sub q2 on level 3).

If you write a scenario from a company or society as the context of your question, make sure it gives enough information for any class members to analyze the situation adequately. Please also carefully study the question grading criteria in the answer grading guidelines at the end of this document before designing your questions.

Make sure your question is different from the other ones posted already.

Posting Format in the Midterm Exam Conference:
Post each of your questions in the Midterm Conference by following the steps. Use the lowest question ID assigned to you for the coding question related to HTML/JavaScript (e.g. Q198); the second lowest question ID for the coding question related to ASP.net (e.g. Q199); and the highest question ID for the short essay question (e.g. Q200).

1. Fill the title/subject of the message with the question ID your are assigned. For example, assume you are assigned question ID 198. The title of your message should be:

   "Q198-Question"
2. In the main message box, cut and paste your question as plain text. Do not use attachment unless you want to include figures. The first two lines of the message should indicate the type of your question (HTMLT/JavaScript; ASP.net; essay), and the level of objective of your question. For example, assume your question is on level 1 and 2:

   "Short Essay Question
   Level 1 (knowledge of specifics), 2 (comprehension), and 3 (application)
   Question 198: *content goes here...*"
3. Post your question anonymously!

If you make a mistake in the topic or by not making it anonymous, use the "edit" command to fix it.

Professor grades, edits, and assigns questions.

The professor will look over each question and will grade based on the quality of the question (question grades will be kept in the professor's grade book). The professor may edit the question when necessary to ensure it is of sufficient quality. The revised question will be posted as a reply to the original question titled, e.g. "Q198-Revised Question". Each of you will be assigned to answer three questions by the instructor. You will be notified on Webboard of the question IDs that you are assigned to answer (e.g. Q14, Q43, Q68).

5. Reviewing questions

You should review all questions as part of your preparation for the exam when questions are posted in the Midterm Exam Conference.

6. Answering questions

You must answer the three questions assigned to you by the deadline. You should carefully study the answer grading criteria at the end of this document before answering. Note:

1. *Length restriction*: the answer to the short essay question should contain more than 350 words and less than 600 words (including tables but not including figures or bibliography section). Microsoft Word has a word count tool, so do other word processing packages.
2. *Few attachments*: please do NOT use attachments unless you need to post a figure as part of your answer. Use plain text or HTML instead.
3. *No Plagiarism will be tolerated*!! Make sure you have read and understood the grading policy on plagiarism before answering. All short essay answers must be submitted on Turnitin.com, in addition to Webboard, for plagiarism check. See details below.

Posting Format in the Public Midterm Conference:
Post your answer as a **reply** to the question you are answering. Post your reply **anonymously!** As an example, assume you are assigned to answer question Q198:

"Q198-Question
Q198 – Answer"

Double-check your answer **after your post it**. Make sure that Webboard posted it exactly what you expected. **Your may not change your answer after the answer posting deadline for any reason**.

Submitting answers on Turnitin:
To prevent and detect plagiarism, you must submit your answer to the short essay question on turnitin.com in addition to Webboard. Submit the answer under the "Exam Essay Answer" assignment link in the CIS 433-Fall 2004 class turnitin website. Use the format "Qxx-Answer" as the submission title. The instructor will do a plagiarism check as part of the final answer grading process.

7. Grading answers to your questions

You must grade answers to the questions you posted by 10/23.

You should pay attention to whether the instructor revised your original questions, and grade accordingly. As part of your grading, you must provide a full written explanation (justification) of your grading. Write at least 3 full sentences explanation for each of the grading categories, and give each category a grade. You should carefully study the answer grading criteria and the grade justification grading criteria at the end of this document before grading.

Posting Format in the Midterm Conference:
Post your grade as a **reply** to the answer you are grading. Post your reply anonymously! As an example, assume your question ID is Q198:

"Q198-Question
Q198 – Answer
Q198 – Answer Grade"

Professor posts final answer grades

The professor will then evaluate the questions, answers, grades, grade justifications and assign a final answer grade to each of your answers using the topic: "Q198 – final answer grade". All final answer grades will be posted on the Midterm Conference. Meanwhile, grading grade will be provided by the professor based on the quality of grading, including the grade justifications (grading grades will be kept in the professor's grade book).

8. Contesting the final grade

If you disagree with the final answer grade from your professor, you may dispute it by 10/26. Note your dispute may result in **either increase or decrease** of your answer grades, when the instructor reevaluates all aspects of the question, answer, and grading. Please follow the steps in dispute:

1. Ensure that you are disputing a significant number of points (i.e. greater than 4 points total per answer disputed). The professor will not reconsider a small number of points.
2. Re-grade your own answer fully, providing full justification using the grading guidelines in the grading section above.
3. Make sure that you provide compelling justification for the re-grade. Also state why you believe the grader was wrong in each category where you dispute the grade.

Do NOT dispute your grade until your professor has posted the final grade. Disputes that do not meet with ALL of the above requirements will be discarded.

<u>Posting Format in the Midterm Conference</u>:
Post your dispute entry as a reply to the final grade you are contesting. Post your reply anonymously! As an example, assume you answered question Q198 and you disagree with the final grade:

"Q198-Question
Q198 – Answer
Q198 – Answer Grade
Q198 – Final Answer Grade
Q198 - Dispute"

The online exam log will be open till the end of the grading dispute phase. You may keep the log to track your exam activities till the end of the grading dispute.

The professor will resolve grade contests by 10/27. At the end of the exam, the Exam Conference on Webboard will look like Figure 1.

9. Post-exam survey

As part of your third and final step in the participation of the examination research project, you are invited to fill out a post-exam questionnaire. The survey is optional. You will receive 4 extra points towards your midterm grade for completing the survey. The URL will be posted in the Midterm Announcement conference right after the exam.

10. Selected student interviews

The researchers of the examination project will invite some of you to post-exam interviews to further discuss your experience in the exam. Interviews will be either face-to-face or via the phone. Participation is optional. A small gift will be provided as appreciation of your participation.

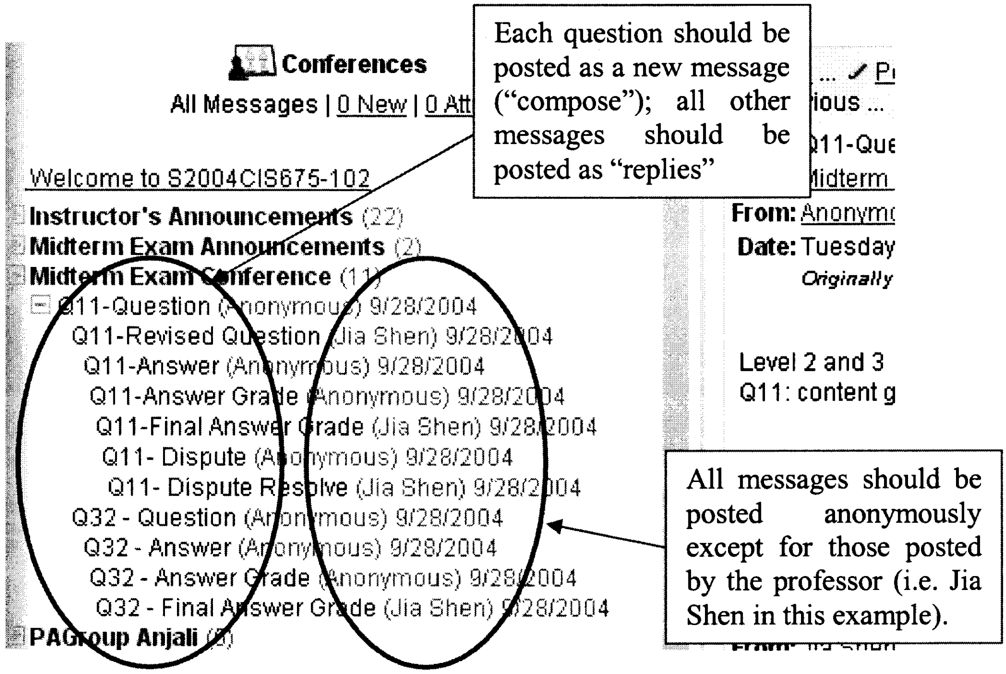


Figure 1 Webboard conference for the online individual exam.

Grading Scheme

Your exam grade will be based on the following three components (total 100 points):

- Question grading (15 points)
- Answer grading (70 points)
- Grade justification grading (15 points).

The professor will provide question grades and grade justification grades. Class member who designs the questions will grade your answers, and the professor will review and assign final answer grades. Grading guidelines for each of the three components are further explained in the section below.

12 extra points towards your exam grade will be provided for your participation in the pre-exam survey (4 points), log record (4 points), and post-exam survey (4 points). Participation is optional. As an alternative to participation in the research, you may write a research paper to gain the extra credits, which is further explained in the course Webboard. The maximum total of the exam is 112.

Grading guidelines

1. Question Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Questions cover different aspects of course material (3 points), are within the exam scope (3 points), and are identified with difficulty levels (3 points). • 4 points: Clarity of questions, including quality of the writing.
Following Directions	• 2 points: Submitting in the correct place and format in the WebCT conference. (Postings not submitted anonymously, or not submitted in the correct place, will be deducted 2 points) • Late submission: minus 4 points for submissions within 24 hours past the deadline; minus 8 points for submissions past 24 hours within 48 hours of the deadline; no submissions will be accepted after 48 hrs.

2. Answer Grading Criteria– used by students and the professor

(Total 70 points – Individual grade)

Q: Quality Grade; FD: Following Directions Grade

- Coding using HTML/JavaScript – total 20 points (Q: 17; FD: 3)
- Coding using ASP.net – total 20 points (Q: 17; FD: 3)
- Short essay question – total 30 points (Q: 26; FD:4)

Quality	• The correctness and completeness of the answer, and (whenever applicable) citing most of all relevant course materials, considering all sides of issues, the quality and clarity of writing, etc.
Following Directions	• Submission of the answer in the correct place and format • Late submission: minus 10 points for submissions within 24 hrs past the deadline; minus 20 points for submissions past 24 hrs within 48 hrs of the deadline; no submissions will be accepted after 48 hrs

Plagiarism: All answers must be examined for plagiarism by considering: 1) whether the answer gives proper citation to the source of information; 2) whether the writing shows the student's own understanding of the knowledge. Proper citation must be used every time sentences or paragraphs are copied from books, papers, or other resources. Students should demonstrate their own understanding of the knowledge by explaining in their own words concepts, theories, methods, and/or providing summaries, examples, etc. If you think parts of the answer are plagiarized, please post the supporting material in your answer grading and grade appropriately.

3. Grade justification Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Quality of grading, including providing a full written explanation (justification) of the grading with at least 3 full sentences explanation for each of the grading categories. • 4 points: Clarity of justifications.
Following Directions	• Same as the question grading criteria

D.3 Student Exam Instructions – Collaborative Exam

Appendix D.3 contains a document as the sample of student exam instructions in the collaborative online exam. The document was given to students in CIS675 in fall 2004, who participated in the collaborative online exam. The exam was conducted on WebCT, and two essay-type questions were required per student (or six in a three-person group). Slight modifications were made in the instructions for other courses in the collaborative exam, including exam scope, question type, examples, changing examples to Webboard when necessary, etc.

CIS 675 Fall ‘04
Midterm Schedule - *Student’s Quick Overview*

Date		Student Activities		Date	Instructor Activities
Oct 1	Fri		**1. Pre-Exam Survey**	Oct 1	
2	Sat			2	
3	Sun			3	
4	Mon			4	
5	Tues	**2. Obtain exam instructions and IDs**		5	Post exam instructions; Assign groups and IDs
6	Wed			6	
7	Thurs		**3. Exam Log**	7	
8	Fri			8	
9	Sat			9	
10	Sun	**4. Select group leader**		10	
11	Mon	**5. Question Design (group)**		11	
12	Tues			12	
13	Wed			13	
14	Thurs	**6. Question Review**		14	Review/Assign questions
15	Fri			15	
16	Sat	**7. Answering (individual)**		16	
17	Sun			17	
18	Mon			18	
19	Tues	**8. Grading (group)**		19	
20	Wed			20	
21	Thurs			21	
22	Fri			22	Review/Post final grades
23	Sat			23	
24	Sun			24	
25	Mon	**9. *Optional: Grade Dispute***		25	
26	Tues		**10. Post-Exam Survey**	26	Resolve disputes
27	Wed			27	
28	Thurs			28	
29	Fri			29	
30	Sat			30	

Note: Please use the number before each step to look for detailed instructions in the document.

CIS 675 Online Small Group Midterm Exam
Student Instructions

CIS 675 is participating in the Student Examination Research Project at NJIT. The fall '04 class will participate in the Online Small Group Exam during the midterm using the course WebCT. This document describes the schedule of the exam, including detailed procedures and grading guidelines.

Schedule

The main activities of the midterm will take place over a 2 week period from **Sunday 10/10 to Tuesday 10/26**. The following table lists dates and each step in the exam process. *Steps in italics (1, 3, 10, and 11)* are optional participation in the research project. Your participation will be awarded with up to 12 extra credits towards the total exam grade.

1. Pre-exam online survey (approx. 15 minutes)	*Friday 10/1 - Monday 10/4*
2. Class members receive detailed exam instructions (this document), group assignment, and group question IDs for anonymity	By Wednesday 10/6
3. Class members obtain online log URL to record exam related activities (approx. 5 minutes each time)	*Wednesday 10/6*
4. Select group leader	By Sunday 10/10
5. Class members post questions first as individuals in the group conference, then reach consensus and post group results in the Midterm Conference	Monday 10/11 –Wednesday 10/13
Professor grades/edits and assigns questions	By Friday 10/15
6. Class members review candidate exam questions	By Friday 10/15
7. Class members answer questions individually, and submit in the Midterm Conference on WebCT and on Turnitin	Saturday 10/16-Monday 10/18
8. Class members grade answers to their questions first as individuals in the group conference, then reach consensus and post group results in the Midterm Conference	Tuesday 10/19 –Thursday 10/21
Professor posts final answer grades	By Sunday 10/24
9. Class members may contest final answer grades *Exam log activity ends*	Monday 10/25
Professor resolves contested grades	By Tuesday 10/26
10. Post-exam online survey (approx. 25 minutes)	*Tuesday 10/26 – Saturday 10/30*
11. Selected student will be interviewed	*By Friday 11/13*

WebCT Conferences

The following conferences will be used for the midterm:

- Midterm Announcement – public; the professor will post instructions, students may post questions regarding the exam procedure etc.
- Midterm Conference – public and anonymous; where the questions, answers, and grading will be posted
- Group conference – private; where each group will have private discussions

See details on the use of these conferences in the detailed steps below.

Detailed Steps

1. Pre-exam survey (completed)

As the first step in the participation of the examination research project, you are invited to fill out a pre-exam survey. The survey is optional. You will receive 4 extra points towards your exam grade for completing the survey. Note all the research data collected in this study are confidential and will be used strictly for research only.

2. Detailed Exam Instructions and IDs for anonymity

Detailed exam instructions are provided to you in this document. Your professor will notify you of your group assignment (e.g. G1). **The group will be the same as the protocol analysis group, unless being instructed by the professor otherwise**. Each student will be assigned with two Question IDs (e.g. GQ11, GQ12). You will be notified via the grade book tool on WebCT of your group assignment and the question IDs. Each group will use the question IDs when posting questions in the public Midterm Conference so the process can remain anonymous. Discussions in the private group conference are not required to be anonymous.

3. Log template

As the second step in the participation of the examination research project, you are invited to keep a record of your exam studying activities using a log template online. Participation is optional. You will receive 4 extra points towards your exam grade for keeping a thorough log. The log URL will be on WebCT.

4. Selecting group leader

Each group needs one group leader. The first member in the group who volunteers before the deadline will be the group leader. If no one volunteers by the deadline, the professor will appoint a group leader. The group leader will be responsible for copying group discussion results from the private group conference to the public Midterm Conference, which will be explained in the following steps. The group leader will also be responsible for contacting group members to make sure every member is following the steps on schedule. Group leader will receive up to 4 extra points towards the exam grade for fulfilling these responsibilities.

5. Posting questions

By 10/11, each group member should post two insightful questions in the private group conference individually using their question IDs (e.g. GQ11- Question). By 10/12, the group should review, comment, and revise questions to make sure of the quality of the questions. By 10/13, the group should reach consensus of the final questions, and the group leader should post all group questions in the public Midterm Conference using the Group Question IDs assigned to group members.

You should indicate the main **objective** of each question using the following categories in terms of testing:

1. Knowledge of specifics- knowing major concepts, methods, and theories
2. Comprehension- understanding of major concepts, methods, and theories
3. Application - using theories, methods, and concepts in new context
4. Analysis – analyzing and solving problems
5. Synthesis – relating materials from several areas to make argument
6. Evaluation – judging the value of ideas and assessing quality of arguments

Each group member should design both questions at level 2 (comprehension) and up. You may have sub-questions in your questions, and you may indicate multiple objectives in each of your questions (e.g., Q1 on level 1 and 3 with sub q1 on level 1 and sub q2 on level 3). Do not write questions that are on level 1 (knowledge of specifics) only. Your question should take into consideration the word limits of the expected answer, which should be more than 750 words but less than 1,700 words. Please carefully study the question grading criteria in the grading guidelines at the end of this document before designing your questions.

Questions should be based on materials up to week 6 of the course, including articles, lectures, online discussions, etc. Each of your two questions should address a **different aspect** of the course materials. Do not write, for example, two questions on protocol analysis or two questions on scales. The answers you anticipate for your two questions can overlap to a limited degree, but write questions where you expect the answers will primarily cover different materials. If you write a scenario from a company or society as the context of your question, make sure it gives enough information for any class members to analyze the situation adequately.

Make sure your question is different from the other ones posted already.

Posting in Private Group Conference:
Each group member should post the questions individually in the private group conference using the question IDs they are assigned as the message title (e.g. GQ11 - Question). Then the group should review, comment, and revise questions to enhance the quality of the questions. After the group decides on the final questions, the group leader should copy all the questions from the private group conference to the public Midterm Conference, as follows.

Posting Format in the Public Midterm Conference:
Post each of the group questions in the Midterm Conference by following the steps.

> Note 1: To display and post messages correctly throughout the exam on WebCT, always set the display of the conference messages as "Threaded" by clicking on the "Threaded" button on top of the conference display (see figure 1).

4. Fill the title/subject of the message with the question ID the group is assigned. For example, assume the group is assigned group question ID GQ11. The title of your message should be:

   "GQ11-Question"

5. In the main message box, cut and paste your question as plain text. Do not use attachment unless you want to include figures. The first line of the message should indicate the level of objective of your question. For example, assume your question is on level 1 and 2:

   "Level 1 (knowledge of specifics), 2 (comprehension), and 3 (application)

   GQ11: *content goes here...*"

6. Post your question anonymously!
7. Repeat steps 1-3 until you finish posting all the questions.

> Note 2: Since WebCT does not allow you to "edit" or "delete" your message once you have posted it, please double check the message before you post, including the topic, message body, and the anonymity check. If you do make a mistake, use the "reply" command to post the correct message and email the professor on WebCT, who will delete the wrong message for you.

Professor grades, edits, and assigns questions.

The professor will look over each question and will grade based on the quality of the question (question grades will be kept in the professor's grade book). Group members will receive the same grade on questions. The professor may also edit the question when necessary to ensure it is of sufficient quality. The revised question will be posted as a reply to the original question titled, e.g. "GQ11-Revised Question". Each of you will be assigned to answer two questions by the instructor. You will be notified via the grade book tool on WebCT of the question IDs that you are assigned to answer (e.g. GQ14, GQ43).

6. Reviewing questions

You should review all questions as part of your preparation for the exam when questions are posted in the Midterm Conference.

7. Answering questions

You must answer the two questions assigned to you by the deadline. This is an individual activity and you are not allowed to discuss with other students. You should carefully study the answer grading criteria at the end of this document before answering. Note:

1. Length restriction: each of your answers should contain more than 750 words and less than 1,700 words (including tables but not including figures or bibliography section).

2. *Bibliography*: Be sure to include a bibliography section in your answer. Every time you reference any class material, put a citation marker such as [1] and then put the full bibliographic citation including page number in the bibliography, e.g.

[1] Spark, J. S., Glow, J. P. and Twinkle, L., (1994). APA format for journal articles. *Management Science*, **28**(10), 1187-1197.

For details on the proper citation format for books, journal articles, book chapters, etc., visit: http://owl.english.purdue.edu/handouts/research/r_apa.html#Examples Again, bibliographic references do not count towards your word length restriction.

3. *Few attachments*: please do NOT use attachments unless you need to post a figure as part of your answer. Use plain text or HTML instead.

4. *No Plagiarism will be tolerated*!! Make sure you have read and understood the grading policy on plagiarism before answering. All answers must be submitted on Turnitin.com, in addition to WebCT, for plagiarism check. See details below.

Posting Format in the Public Midterm Conference:
Post your answer as a **reply** to the question you are answering. Post your reply **anonymously!** As an example, assume you are assigned to answer question GQ11:

"GQ11-Question
GQ11 – Answer"

Double-check your answer **after your post it**. Make sure that WebCT posted it exactly what you expected. See note 1 and note 2 on the use of WebCT. **Your may not change your answer after the answer posting deadline for any reason**.

Submitting answers on Turnitin:

To prevent and detect plagiarism, you must submit your answers on turnitin.com in addition to WebCT. Turnitin is a web-based service that compares and reports similarities of the submitted document with published literature and online resources. Student Quick Start Guide is located at:
http://www.turnitin.com/static/training_support/tii_student_quickstart.html

Here are the steps for students to use turnitin.com:
1. Go to http://www.turnitin.com/. Create a user profile (use your real name when filling out the profile)
2. Use the "class enrollment wizard" to join the class (see class info below)
3. Find the two exam links on the class assignment list.
4. Submit each of your two answers under each of the exam links (details below).

The class information that you need after you create your profile are:
Class ID: **1184278**
Enrollment password: **675851**

After you successfully create your profile and join the class, you should be able to see "CIS 675-851 Fall 04" class. Find the "Exam Answer 1" and "Exam Answer 2" assignments on the assignment list. Submit your answer to the question with lower ID number to "Exam Answer 1", and the answer to the question with higher ID number to "Exam Answer 2" (e.g. if you are assigned to answer questions GQ 21 and GQ22, submit your answer to GQ21 under "Exam Answer 1" and the answer to GQ22 under "Exam Answer 2"). Use the format "GQxx-Answer" as the submission title. The instructor will do a plagiarism check as part of the final answer grading process.

8. Grading answers to your questions

The group leader should copy the answers to the group questions from the public Midterm Conference to the private group conference by the morning of 10/19. By 10/19, each group member should post grading of answers to their questions in the private group conference. By 10/20, the group should review, comment, and revise grading to ensure of the quality of the grading. By 10/21, the group should reach consensus of the final grading, and the group leader should post all the grading in the public Midterm Conference following the format below. The group should pay attention if the instructor revised the group's original questions, and grade accordingly.

As part of your grading, you must provide a full written explanation (justification) of your grading. Write at least 3 full sentences explanation for each of the grading categories, and give each category a grade. You should carefully study the answer grading criteria and the grade justification grading criteria at the end of this document before grading.

Posting in Private Group Conference:
Each group member should post grading and justifications individually in the private group conference for each of the answers to your questions. Then the group should review, comment, and revise grading to ensure the quality of the grading. After the group decides on the final grading, the group leader will post grading in the Midterm Conference **separately** for each of the answers, as follows.

Posting Format in the Public Midterm Conference:
Post your grade as a **reply** to the answer you are grading. Post your reply anonymously! As an example, assume your group question ID is GQ11:

"GQ11-Question
GQ11 – Answer
GQ11 – Answer Grade"

Professor posts final answer grades

The professor will then evaluate the questions, answers, grades, grade justifications and assign a final answer grade to each of your answers using the topic: "GQ11 – final answer grade". All final answer grades will be posted on the public Midterm Conference. Meanwhile, the professor will provide a grading justification grade to each team based on the quality of grading, and keep the scores in the grade book.

9. Contesting the final grade

If you disagree with the final answer grade from your professor, you may dispute it by 10/25. Note your dispute may result in **either increase or decrease** of your answer grades, when the instructor reevaluates all aspects of the question, answer, and grading. Please follow the steps in dispute:

1. Ensure that you are disputing a significant number of points (i.e. greater than 5 points total per answer disputed). The professor will not reconsider a small number of points.
2. Re-grade your own answer fully, providing full justification using the grading guidelines in the grading section above.
3. Make sure that you provide compelling justification for the re-grade. Also state why you believe the grader was wrong in each category where you dispute the grade.

Do NOT dispute your grade until your professor has posted the final grade. Disputes that do not meet with ALL of the above requirements will be discarded.

<u>Posting Format in the Public Midterm Conference</u>:
Post your dispute entry as a reply to the final grade you are contesting. Post your reply anonymously! As an example, assume you answered question GQ11 and you disagree with the final grade:

"GQ11-Question
GQ11 – Answer
GQ11 – Answer Grade
GQ11 – Final Answer Grade
GQ11 - Dispute"

The online exam log will be open till the end of the grading dispute phase. You may keep the log to track your exam activities till then.

The professor will resolve grade contests by 10/26. At the end of the exam, the Midterm Conference on WebCT will look like Figure 1.

10. Post-exam survey

As part of your third and final step in the participation of the examination research project, you are invited to fill out a post-exam questionnaire. The survey is optional. You will receive 4 extra points towards your midterm grade for completing the survey. The URL will be posted in the Midterm Announcement conference right after the exam.

11. Selected student interviews

The researchers of the examination project will invite some of you to post-exam interviews to further discuss your experience in the midterm. Interviews will be either face-to-face or via the phone. Participation is optional. A small gift will be provided as appreciation of your participation.

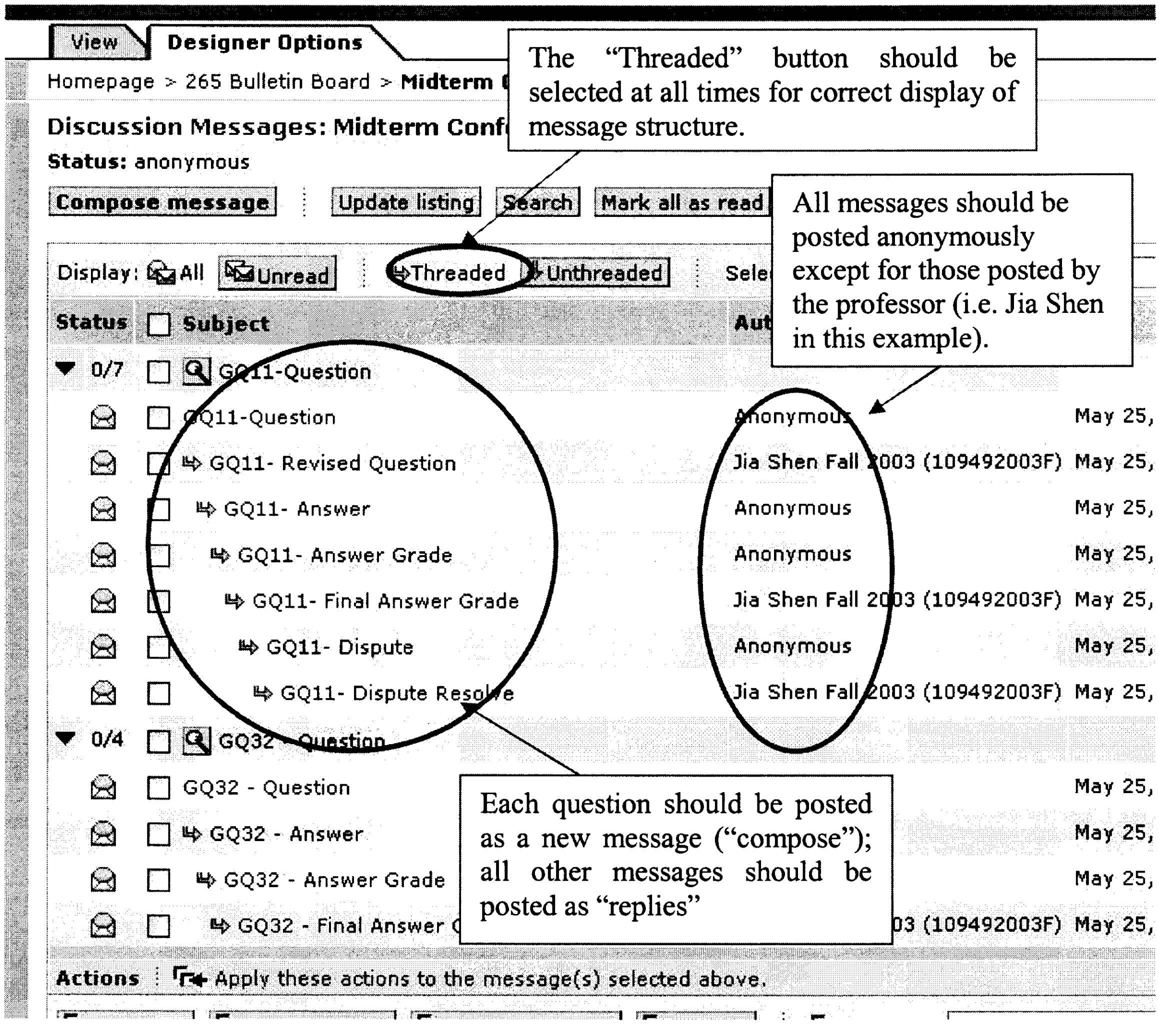


Figure 1 WebCT conference for the online small group exam

Grading Scheme

Your midterm grade will be based on the following three components (total 100 points):

- Question grading (15 points)
- Answer grading (70 points)
- Grade justification grading (15 points).

The professor will provide question grades and grade justification grades. The group who designs the questions will grade your answers, and the professor will review and assign final answer grades. Grading guidelines for each of the three components are further explained in the section below.

12 extra points towards your exam grade will be provided for your participation in the pre-exam survey (4 points), log record (4 points), and post-exam survey (4 points). The group leader will be awarded with up to 4 extra credits for fulfilling various group responsibilities. The maximum total of the exam is 116.

Grading guidelines

1. Question Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Questions cover different aspects of course material (3 points), are within the exam scope (3 points), and are identified with difficulty levels (3 points). • 4 points: Clarity of questions, including quality of the writing.
Following Directions	• 2 points: Submitting in the correct place and format in the WebCT conference. (Postings not submitted anonymously, or not submitted in the correct place, will be deducted 2 points) • Late submission: minus 4 points for submissions within 24 hours past the deadline; minus 8 points for submissions past 24 hours within 48 hours of the deadline; no submissions will be accepted after 48 hrs.

2. Answer Grading Criteria– used by students and the professor

(Total 70 points (35 points each answer)– Individual grade)

Quality	• 23 points: The correctness and completeness of the answer, including citing most of all relevant course materials, considering all sides of issues, synthesizing etc. • 7 points: The quality and clarity of writing, including providing justification to points, etc.
Following Directions	• 3 points: Following editing guidelines including correct citation format (deduct up to 3 points), and length (deduct 3 points if the answer is under 750 words or exceeds the 1,700 words limit.) • 2 points: Submission of the answer in the correct place and format • Late submission: minus 10 points for submissions within 24 hrs past the deadline; minus 20 points for submissions past 24 hrs within 48 hrs of the deadline; no submissions will be accepted after 48 hrs

Plagiarism: All answers must be examined for plagiarism by considering: 1) whether the answer gives proper citation to the source of information; 2) whether the writing shows the student's own understanding of the knowledge. Proper citation must be used every time sentences or paragraphs are copied from books, papers, or other resources. Students should demonstrate their own understanding of the knowledge by explaining in their own words concepts, theories, methods, and/or providing summaries, examples, etc. If you think parts of the answer are plagiarized, please post the supporting material in your answer grading and grade appropriately.

3. Grade justification Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Quality of grading, including providing a full written explanation (justification) of the grading with at least 3 full sentences explanation for each of the grading categories. • 4 points: Clarity of justifications.
Following Directions	• Same as the question grading criteria

APPENDIX E

INSTRUCTOR'S EXAM GUIDE

Appendix E contains samples of exam guidelines for instructors in this study. The documents include instructor's guide for the participatory exam (Appendix E.1), and the collaborative exam (Appendix E.2).

E.1 Instructor's Exam Guide – Participatory Exam

Appendix E.1 contains a sample document of the instructor's exam guide used for the participatory exam. The document was used in CIS433 in fall 2004, which participated in the participatory online exam. The exam was conducted on Webboard, and three questions were required per student. Slight modifications were made in the documents for other courses in the participatory exam, including exam scope, question type, examples, changing examples to Webboard when necessary, etc.

CIS 433 Fall '04
Midterm Exam Schedule - *Instructor's Quick Overview*

Date		Student Activities		Date	Instructor Activities
Oct. 3	Sun		Pre-Exam Survey	Oct. 3	
4	Mon			4	
5	Tues			5	
6	Wed			6	
7	Thurs	Obtain exam instructions and IDs		7	**1. Post exam instructions; Assign IDs**
8	Fri			8	
9	Sat		Exam Log	9	
10	Sun			10	
11	Mon			11	
12	Tues			12	
13	Wed	Question Design		13	
14	Thurs			14	
15	Fri			15	
16	Sat	Question Review		16	**2.Review/Assign questions**
17	Sun			17	
18	Mon	Answering		18	
19	Tues			19	
20	Wed			20	
21	Thurs	Grading		21	
22	Fri			22	
23	Sat			23	
24	Sun			24	**3. Review/Post final grades**
25	Mon			25	
26	Tues	*Optional: Grade Dispute*		26	
27	Wed			27	**4. Resolve disputes**
28	Thurs		Post-Exam Survey	28	
29	Fri			29	
30	Sat			30	
31	Sun			31	

Note: Please use the number before each step to look for detailed procedures in the document.

Online Individual Midterm Exam - Instructor's Guide

Thank you for your participation in the Examination Research Project at NJIT! This document provides the detailed guidelines for conducting the online individual exam on Webboard, including Webboard tools for the exam, detailed procedures for instructors, and total exam grade calculation. The appendix includes the grading guidelines, Excel template, and the instructor's log form for research purpose. Please carefully review this document, and the instruction document to students, before the exam starts. Your adherence to the suggested guidelines is critical to the success of the research. Please feel free to contact Jia Shen (jxs1866@njit.edu) at any time if you have any questions, or if you need any assistance. Your effort will be greatly appreciated by the research team!

Summary of Webboard Tools for the exam:

The following tools will be used in the exam:

- Webboard Conferences
- Excel spreadsheets for the instructor to keep track of grades
- Turnitin to help the instructor detect plagiarism

Details will be described in the steps below.

Detailed Steps:

1. Post exam instructions; Assign IDs

Step 1: Review student exam instructions provided by Jia Shen

Please make sure you review the student instructions document to understand the steps your students will go through in this exam. Also customize the sections highlighted blue for your course, such as exam scope, etc.. Please save the document after your review and revision for step 2.

Step 2: Create Two Exam Conferences on Webboard

- Midterm Exam Announcement – public. This is where you will post announcements, exam instructions, and answer students' questions about the exam, etc. Survey announcements will also be posted in this conference.
    - o Please post the student exam instruction document you reviewed and revised in step 1 in the Final Exam Announcement conference.
- Midterm Exam Conference – public. This is the main conference where students will post questions, answers, and grading anonymously, and where you will post final grades etc. (see Figure 1).

Optional Step if you haven't done so in your class: Assign students with Pen Names

Assign each student with a pen name before the exam starts so that you can inform students of their question IDs, etc. by posting one message instead of sending out individual emails to each student. You may use existing pen names if you already have them. Otherwise, please assign pen names and inform each student individually via email.

Figure 1 Creating Webboard conferences for the online individual exam.

Step 3: Assign Question IDs
Each student will design three questions, and should be assigned three question IDs (e.g. Q197, Q198, Q199). Please record the question IDs in the excel file (see app 2). Record the question IDs in the excel file (see appendix 2). Post a message in the Midterm Exam Announcement conference to inform students of their question IDs (see Figure 2).

The above three steps should be completed by 10/9. Then students are ready to start the exam!

2. Review/assign questions

Step 4. Review/Edit Questions
Once students have posted questions in the Midterm Exam Conference, please review all questions to make sure they are of sufficient quality. Take into consideration:

1. Each student should deign three questions, one for each of the following types:
   - Coding or providing program result problem using HTML and/or JavaScript (similar to those appear in class quizzes)
   - Coding or providing program result problem using ASP.net (similar to those appear in class quizzes)
   - Short essay question on definitions of terminology or concepts of E-Commerce.
2. All questions should be identified with difficulty level from level 1 to level 6.
3. The expected answer to the essay question should be more than 750 words (1.5 half pages, single space) and less than 1700 words (3.5 pages, single space)

If you need to edit the question, please post a reply to the original question on Webboard (see Figure 3). The title of the message should be the question ID with "Revised Question" (e.g. "**Q198- Revised Question**"). In the message box, please put in your revised question.

While you review the questions, you are suggested to record the "topic area" for each question using key words (e.g. IP telephony, LEC), and the difficulty level, e.g. (low, medium, high) for better question assignment in step 6 (see appendix 2).

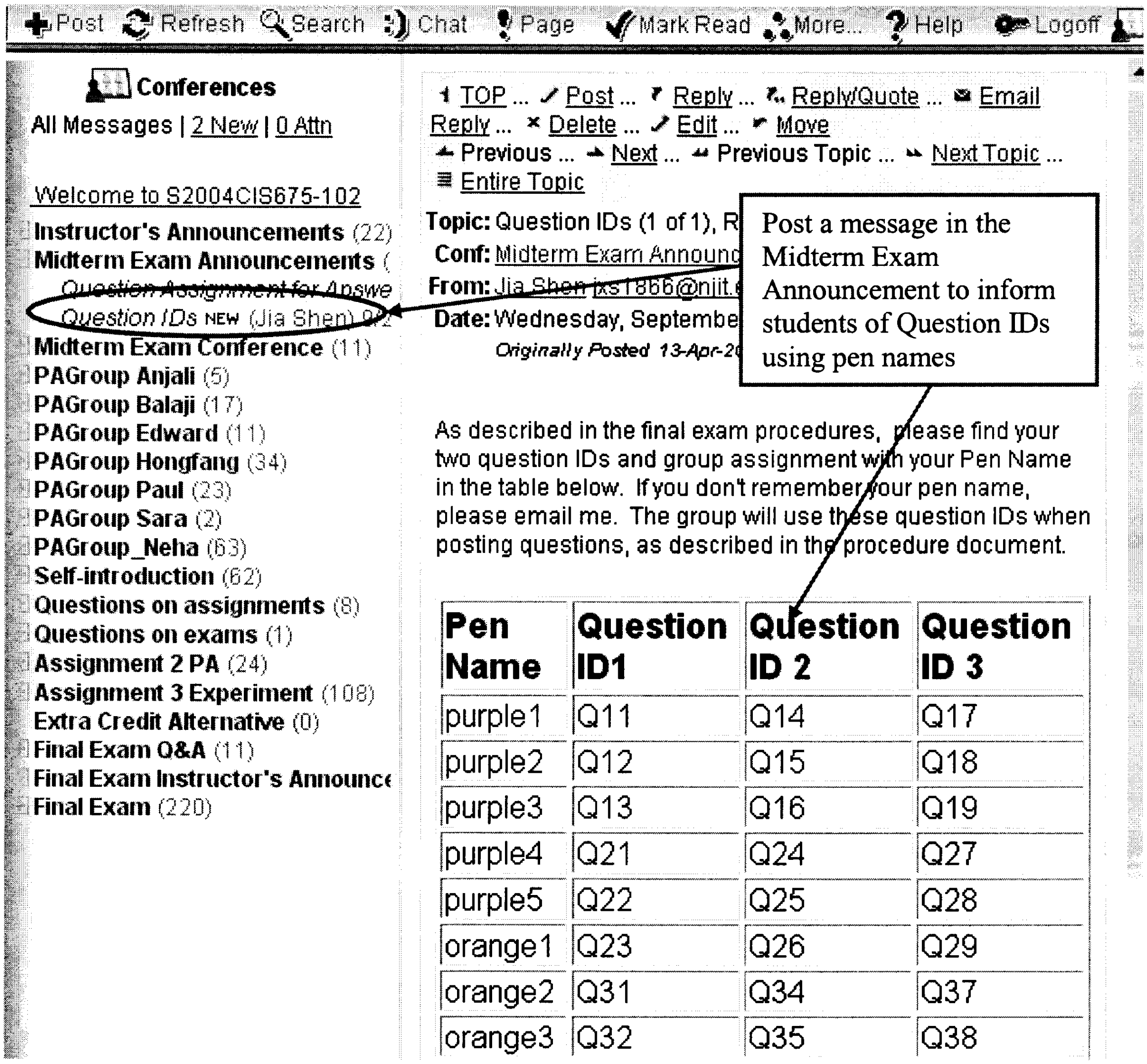


Figure 2 Using pen names to inform students of question IDs.

Step 5. Grade Question Quality

You will give each student a "question grade" based on the quality of questions with a maximum of 15 points. It is suggested that you provide question quality grades while you are reviewing questions in step 4. All question grades should be kept in the excel file (see appendix 2). See the grading criteria at the end of this document for details.

Step 6. Assign Questions to Each Student to Answer

Each student should be assigned with three questions to answer, which are not composed by the student him/herself. Please assign each student one question from each of the three question types discussed above. Please assign student with varying levels of difficulties (e.g. one easy, one medium, and one difficult, or three medium).

Record the question assignment in the excel file (see appendix 2). Inform students of their three questions to answer using pennames on Webboard (see Figure 3).

Creating the assignment on Turnitin.com
Please create one assignment "Exam Essay Answer" on turnitin.com so students can submit their answers to the essay questions for plagiarism check.
The above three steps should be completed by 10/17. Then students are ready to answer the questions!

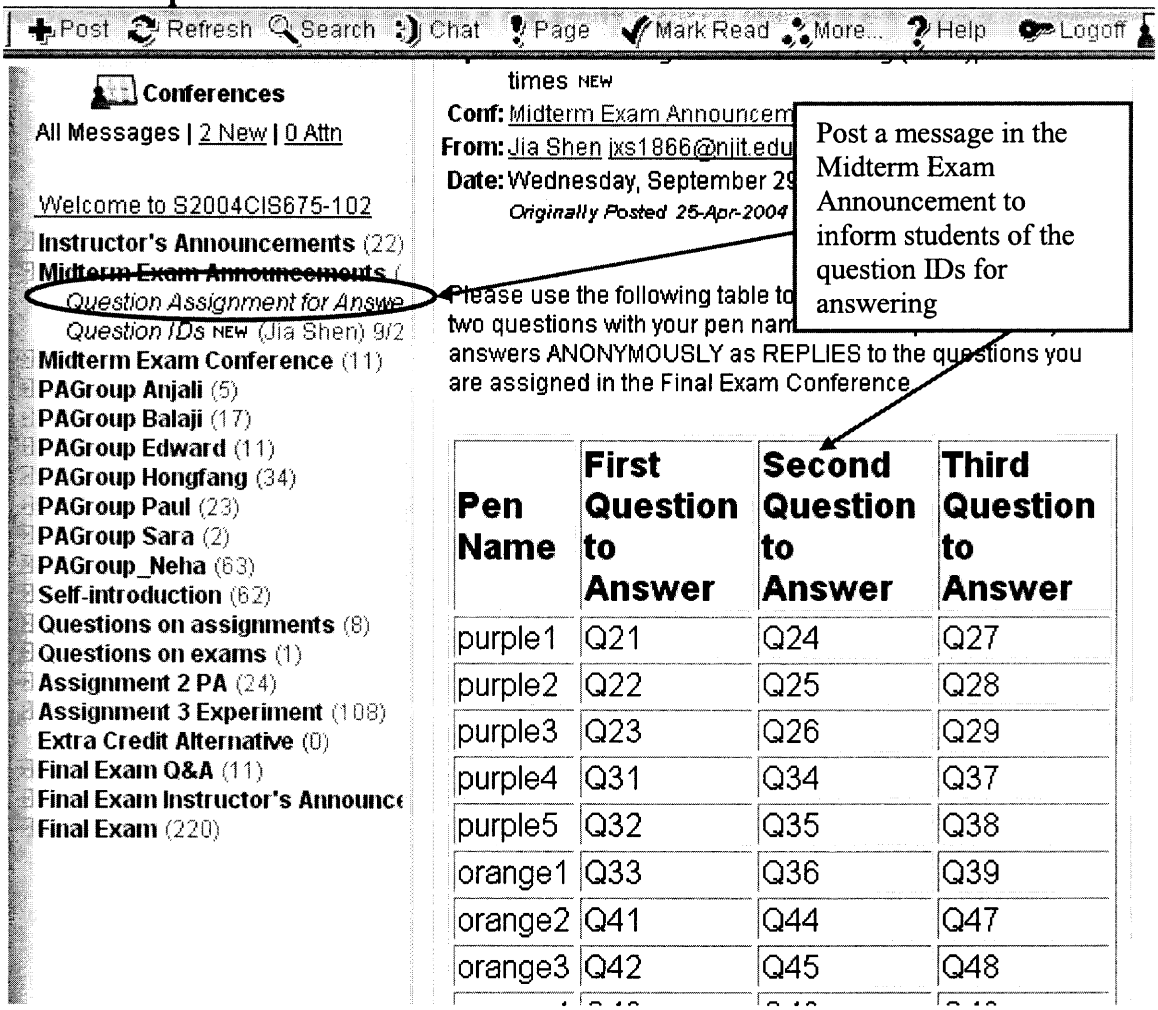


Figure 3 Using Webboard to assign questions for answering.

3. Review/Post final grades

Step 7. Review/Post Final Answer Grades

Once students have posted the answer grades, please review the question, answer, and grading justification, and provide a final answer grade. You should NOT grade answers from scratch. Please use the grading justification provided by the student grader to help you expedite your grading. Please remember to check Turnitin for plagiarism check of the answers to short essay questions.

When you finish determining the grade, please post your final answer grading as a reply to the grading message posted by the student on Webboard. Your message should be titled in the format of "**Qxx - Final Answer Grade**" (see Figure 4). If you agree with the student grading, you may simply put in the message box "Agree with student grading. The final answer grade is: XX/30 points". If you need to grade on your own, please follow the answer grading guideline (see appendix 1). You should also update the "answer grade" columns in the Excel file (see appendix 2).

Step 8. Grade Grading Justification Quality

You will give each student a "grading grade" based on the quality of grading and grading justifications with a maximum of 15 points. It is suggested that you provide grade justification grades while you are reviewing grades in step 7. All grade justification grades should be kept in the excel file (see appendix 2). See the grading criteria at the end of this document for details.

4. Resolve disputes

Step 9: Re-evaluation of the Grade

Students are given one day to dispute your final answer grades. If anyone disputes the grade, please first examine whether the dispute meets with all of the three criteria in the student instruction document: disputing a significant number of points (more than 5 points total per grading disputed), re-grade one's own answer fully, and providing compelling justification using the grading guideline (see appendix 1). Dispute that does not meet with all the requirements should be discarded.

If the dispute meets the requirements, you should re-evaluate all aspects of the grading, including the question, answer, grading justification, and especially the self-evaluation and justification provided in the dispute. You may either increase or decrease the original final answer grade through the re-evaluation. Please post your resolution of the dispute as a reply to the dispute message. The title of the message should be question ID with "Dispute Resolve" (e.g. "**Q198- Dispute Resolve**") (see Figure 4). Please update the excel file with the dispute results (see appendix 2).

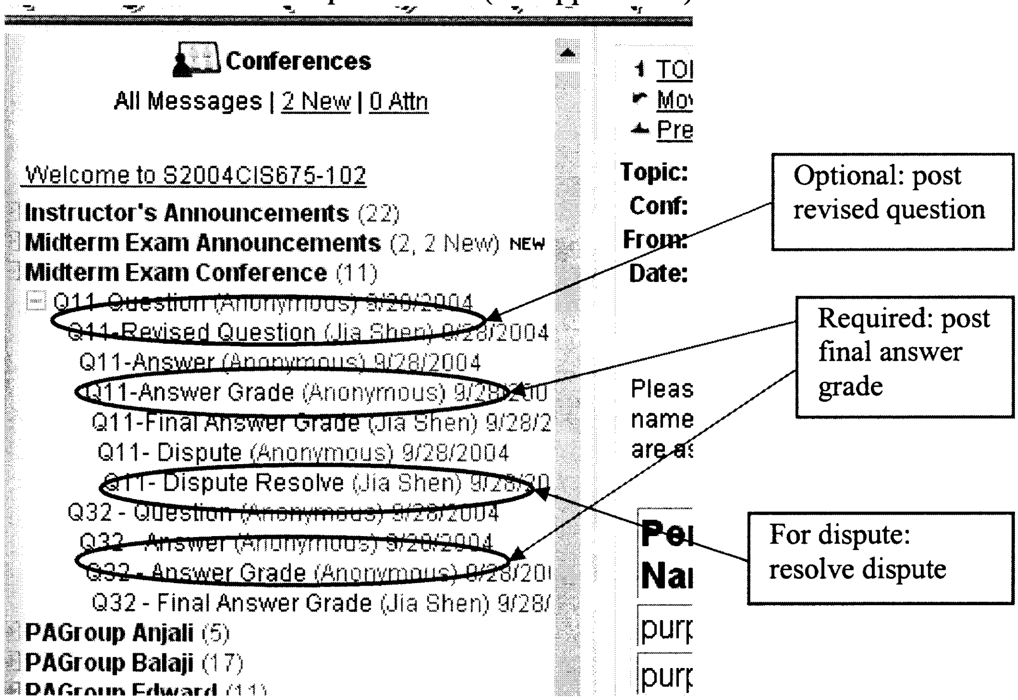


Figure 4 Using Webboard conference for the online individual exam.

You have now completed the online exam!

Final Step (step 10): Total exam grade calculation

The total exam grade should be calculated as the sum of the following four grades:

- Question grade (15 points) + Answer grades (70 points) + Grade justification grade (15 points) + Extra credits total (12 points maximum)

Appendix 1 - Grading guidelines

1. Question Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Questions cover different aspects of course material (3 points), are within the exam scope (3 points), and are identified with difficulty levels (3 points). • 4 points: Clarity of questions, including quality of the writing.
Following Directions	• 2 points: Submitting in the correct place and format in the WebCT conference. (Postings not submitted anonymously, or not submitted in the correct place, will be deducted 2 points) • Late submission: minus 4 points for submissions within 24 hours past the deadline; minus 8 points for submissions past 24 hours within 48 hours of the deadline; no submissions will be accepted after 48 hrs.

2. Answer Grading Criteria– used by students and the professor

(Total 70 points – Individual grade)

Q: Quality Grade; FD: Following Directions Grade

- Coding using HTML/JavaScript – total 20 points (Q: 17; FD: 3)
- Coding using ASP.net – total 20 points (Q: 17; FD: 3)
- Short essay question – total 30 points (Q: 26; FD:4)

Quality	• The correctness and completeness of the answer, and (whenever applicable) citing most of all relevant course materials, considering all sides of issues, the quality and clarity of writing, etc.
Following Directions	• Submission of the answer in the correct place and format • Late submission: minus 10 points for submissions within 24 hrs past the deadline; minus 20 points for submissions past 24 hrs within 48 hrs of the deadline; no submissions will be accepted after 48 hrs

Plagiarism: All answers must be examined for plagiarism by considering: 1) whether the answer gives proper citation to the source of information; 2) whether the writing shows the student's own understanding of the knowledge. Proper citation must be used every time sentences or paragraphs are copied from books, papers, or other resources. Students should demonstrate their own understanding of the knowledge by explaining in their own words concepts, theories, methods, and/or providing summaries, examples, etc. If you think parts of the answer are plagiarized, please post the supporting material in your answer grading and grade appropriately.

3. Grade justification Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Quality of grading, including providing a full written explanation (justification) of the grading with at least 3 full sentences explanation for each of the grading categories. • 4 points: Clarity of justifications.
Following Directions	• Same as the question grading criteria

Appendix 2- Using Excel Spreadsheet to Manage ID Assignment and Grades

Last name	First name	Q1 ID	Q1 Notes	Q2 ID	Q2 Notes	Ques. Grade	Grad. Grade	A1 QID	A1 Grade	A 1 Disp.	A 2 QID	A 2 Grade	A 2 Disp.	Extra Total	Sum
Smith	John	Q11	DFD/E	Q12	T-test/E	15	15	Q85	26		Q86	34		0	90
Bush	Kenny	Q13	Case/H	Q14	DFD/M	15	15	q63	31		q64	35		4	100
Manuel	Loren	Q15	Telecom /H	Q16	Case/M	15	14	q45	35		q46	35		9	108
Clinton	Tracy	Q31	BPR/H	Q32	Telecom/E	2	15	q15	30		q62	35		8	90
Patel	Bob	Q33	DFD/H	Q34	Case/H	15	14	q61	32		g66	35		12	108
Wang	Bill	Q35	T-test/M	Q36	yes	14	15	q65	29		q96	33		4	95
O'Brien	Will	Q41	BPR/E	Q42	Telecom/M	0	15	q95	34		q16	30		8	87
Khan	Ray	Q43	Case/M	Q44	DFD/E	15	14	q83	34		q84	31		6	100
Han	Adam	Q45	Telecom /E	Q46	T-test/H	15	15	q31	32		q32	34		12	108
....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....

Step 3 Step 4* Step 5 Step 8 Step 6 Step 7 Step 9 Step 10

The above example shows the situation when students are assigned to design/answer TWO questions. Please add columns for additional questions/answers.

*The Notes field for each question (Q1 Notes, Q2 Notes) may be used to record the topic area of each question (e.g. DFD, Telecom, T-test) and the difficulty level for better question assignment (Easy, Medium, Hard).

Instead of starting from scratches to build the spreadsheet, you may download the grade book from Webboard and import it into Excel. In this way you will have students' names in the spreadsheet instead of manually input them.

Appendix 3 - Instructor Exam Log Template

To help the research team investigate the time and effort instructors spend in conducting the online individual exam on Webboard, you are invited to record your exam related activities using the form below. Please record the date, and the duration (in hours) of your activities. Your activities are classified into the following five categories:

- **Administration** – this includes any effort in administering the exam. For example posting announcements, customizing guidelines for your course, assigning students with IDs, posting grades, etc. Notice this category refers to activities that only involve yourself, and it does not involve two-way communication with students.
- **Communication with students** – this refers to any activities that involve two-way communication with your students, such as answering students' questions, etc.
- **Question** – this refers any effort related to questions for the exam. For example, time spent reviewing all questions posted by students in online exams, etc.
- **Grade** – this refers to any effort related to grading students' answers. For example, time spent in reviewing and grading answers, and resolving disputes.
- **Other** – any other effort you put into the exam that is not listed in the categories above.

Your thorough completion of the log will greatly help us to investigate and reduce the workload for instructors in future studies. Thank you in advance!

Date	**Activity**				
	Admin.	**Communicate**	Question	**Grade**	**Other**
9/30/04		0.5 hour	1 hour		

The first line is filled in as an example

E.2 Instructor's Exam Guide – Collaborative Exam

Appendix E.2 contains a sample document of the instructor's exam guide used for the collaborative exam. The document was used in CIS675 in fall 2004, which participated in the collaborative online exam. The exam was conducted on WebCT, and two essay-type questions were required per student (or six in a three-person group). Slight modifications were made in the documents for other courses in the collaborative exam, including exam scope, question type, examples, changing examples to Webboard when necessary, etc.

CIS 675-851 Fall '04

Midterm Schedule- *Instructor's Quick Overview*

Date		Student Activities		Date	Instructor Activities
Oct				Oct	
1	Fri		Pre-Exam Survey	1	
2	Sat			2	
3	Sun			3	
4	Mon			4	
5	Tues	Obtain exam instructions and IDs		5	**1. Post exam instructions; Assign groups and IDs**
6	Wed			6	
7	Thurs		Exam Log	7	
8	Fri			8	
9	Sat			9	
10	Sun	Select group leader		10	
11	Mon	Question Design (group)		11	
12	Tues			12	
13	Wed			13	
14	Thurs	Question Review		14	**2. Review/Assign questions**
15	Fri			15	
16	Sat	Answering (individual)		16	
17	Sun			17	
18	Mon			18	
19	Tues	Grading (group)		19	
20	Wed			20	
21	Thurs			21	
22	Fri			22	**3. Review/Post final grades**
23	Sat			23	
24	Sun			24	
25	Mon	*Optional: Grade Dispute*		25	
26	Tues		Post-Exam Survey	26	**4. Resolve disputes**
27	Wed			27	
28	Thurs			28	
29	Fri			29	
30	Sat			30	

Note: Please use the number before each step to look for detailed procedures in the document.

Online Small Group Exam - Instructor's Guide

Thank you for your participation in the Examination Research Project at NJIT! This document provides the detailed guidelines for conducting the online small group exam on WebCT, including WebCT tools for the exam, detailed procedures for instructors, and total exam grade calculation. The appendix includes the grading guidelines, Excel template, and the instructor's log form for research purpose. Please carefully review this document, and the instruction document to students, before the exam starts. Your adherence to the suggested guidelines is critical to the success of the research. Please feel free to contact Jia Shen (jxs1866@njit.edu) at any time if you have any questions, or if you need any assistance. Your effort will be greatly appreciated by the research team!

Summary of WebCT Tools for the exam:

The following tools will be used in the exam:

- WebCT Conference
- WebCT Grade book
- WebCT email
- WebCT Calendar (optional)

In addition, Turnitin will be used to help the instructor detect plagiarism. Details will be described in the steps below.

Detailed Steps:

1. Post exam instructions; Assign groups and IDs

Step 1: Create Two Exam Conferences on WebCT

- Midterm Announcement – public. This is where you will post announcements and answer students' questions about the exam, etc. Survey announcements will also be posted in this conference.
- Midterm Conference – public and anonymous (please check the "anonymous" box on the conference property page, see Figure 1). This is the main conference where students will post questions, answers, and grading anonymously, and where you will post final grades etc.

Step 2: Form Groups

The online small group exam is conducted using student groups with 3 students per group. Occasionally, 4 or 2 people groups can be formed if the class size does not divide by 3. It is preferred to use existing groups that have already been working together through the semester, e.g. on group assignments, etc. If there are more than 3 people in the existing group, please randomly select people out of the group to form new groups. Please use an excel spreadsheet to keep track of group assignment, and name the groups as G1, G2, G3, etc (see appendix 2).

Step 3: Create Private Group Conferences on WebCT

In the question design and grading phases of the exam, students will discuss with other group members in the private group conferences on WebCT. Please create private group conferences for each of the groups you formed. Please make sure you set the groups as

"Private", and add group members into each of the corresponding groups. The conference names should be the same as the group name (see Figure 1).

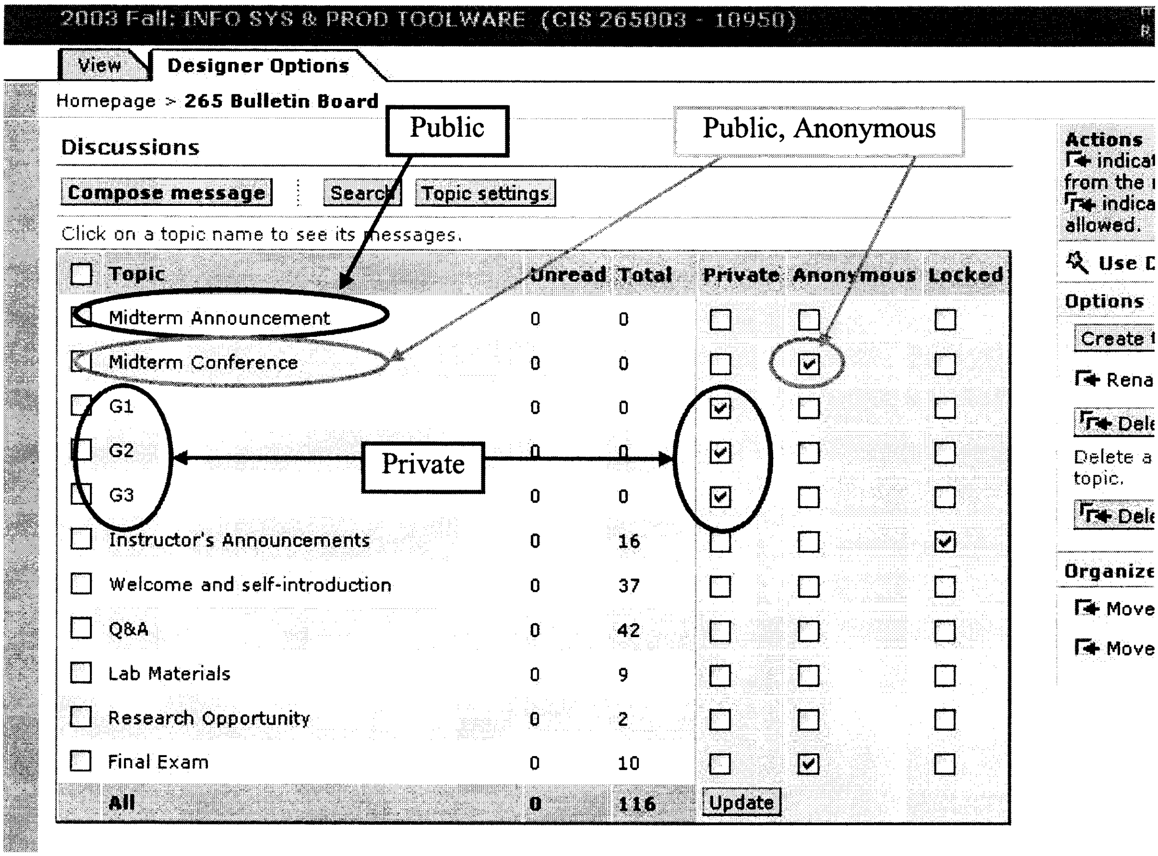


Figure 1 Creating WebCT conferences for the online small group exam.

Step 4: Inform Students of Group IDs
Use the WebCT grade book tool to inform each students of their group assignment (see Figure 2).

Step 5: Assign Question IDs
Each student will design two questions, and should be assigned two question IDs. For example, the first member in group 1 should be given question IDs GQ11 and GQ14, the second member in group 1 should be given question IDs GQ12 and GQ15, and the third member in group 1 should be given question IDs GQ13 and GQ16.

Record the question IDs in the excel file (see appendix 2). Use the WebCT grade book tool to inform students of their question IDs (see Figure 2).

The above five steps should be completed by Oct. 6th. Please also review the instructions to student document provided by Jia Shen, and revise/customize the document (e.g. exam scope) when necessary. Please upload the instruction document in the Midterm Announcement Conference by Oct. 6th. Then students are ready to start the exam!

Note: Since WebCT does not allow students to "edit" or "delete" messages once posted, students are instructed to carefully examine the message before posting. However, if an error is made, the student is instructed to post the correct message as a reply to the original message, and email the instructor on WebCT. Please delete the wrong message for students upon receipt of the email, if any.

2. Review/assign questions

Step 6. Review/Edit Questions

Once students have posted questions in the Midterm Conference, please review all questions to make sure they are of sufficient quality. Take into consideration:

1. All questions should be on level 2 (comprehension) and up in terms of question objectives. The question should denote the objectives in the first line of the question message, e.g.
   "Level 1 (knowledge of specifics), 2 (comprehension), and 3 (application)
   GQ11: *content goes here...*"
2. The expected answer to the question should be more than 750 words (1.5 half pages, single space) and less than 1700 words (3.5 pages, single space)
3. Two questions each cover different aspects of course material. Both questions are relevant to the course material within the exam scope.

If you need to edit the question, please post a reply to the original question on WebCT (see Figure 3). The title of the message should be the question ID with "Revised Question" (e.g. "**GQ11- Revised Question**"). In the message box, please put in your revised question.

While you review the questions, you are suggested to record the topic area for each question using key words (e.g. use case, DFD, experiment design), and the difficulty level, e.g. (low, medium, high) for better question assignment in step 8 (see appendix 2).

Step 7. Grade Question Quality

You will give each group a "question grade" based on the quality of questions with a maximum of 15 points. It is suggested that you provide question quality grades while you are reviewing questions in step 6. All question grades should be kept in the excel file (see appendix 2). Group members should receive the same question grade, except for situations described below.

In case that some group member does not submit any question in the private group conference, please inform the group to submit the group results before the deadline without the member. Also remind the group leader to keep an eye on any late postings within 48 hours past the deadline, and copy the message to the public exam conference. The late submissions will be graded separately according to the grading criteria. See the grading criteria at the end of this document for details.

Step 8. Assign Questions to Each Student to Answer

Each student should be assigned with two questions to answer, which are not composed by the student him/herself, or any other members on the same group. It is preferred to assign questions from two topic areas (e.g. one question on DFD, and another one on use case) with varying levels of difficulties (e.g. one easy and one difficult, or two medium).

Record the question assignment in the excel file (see appendix 2). Inform students of their two questions to answer using the grade book tool on WebCT (see Figure 2).

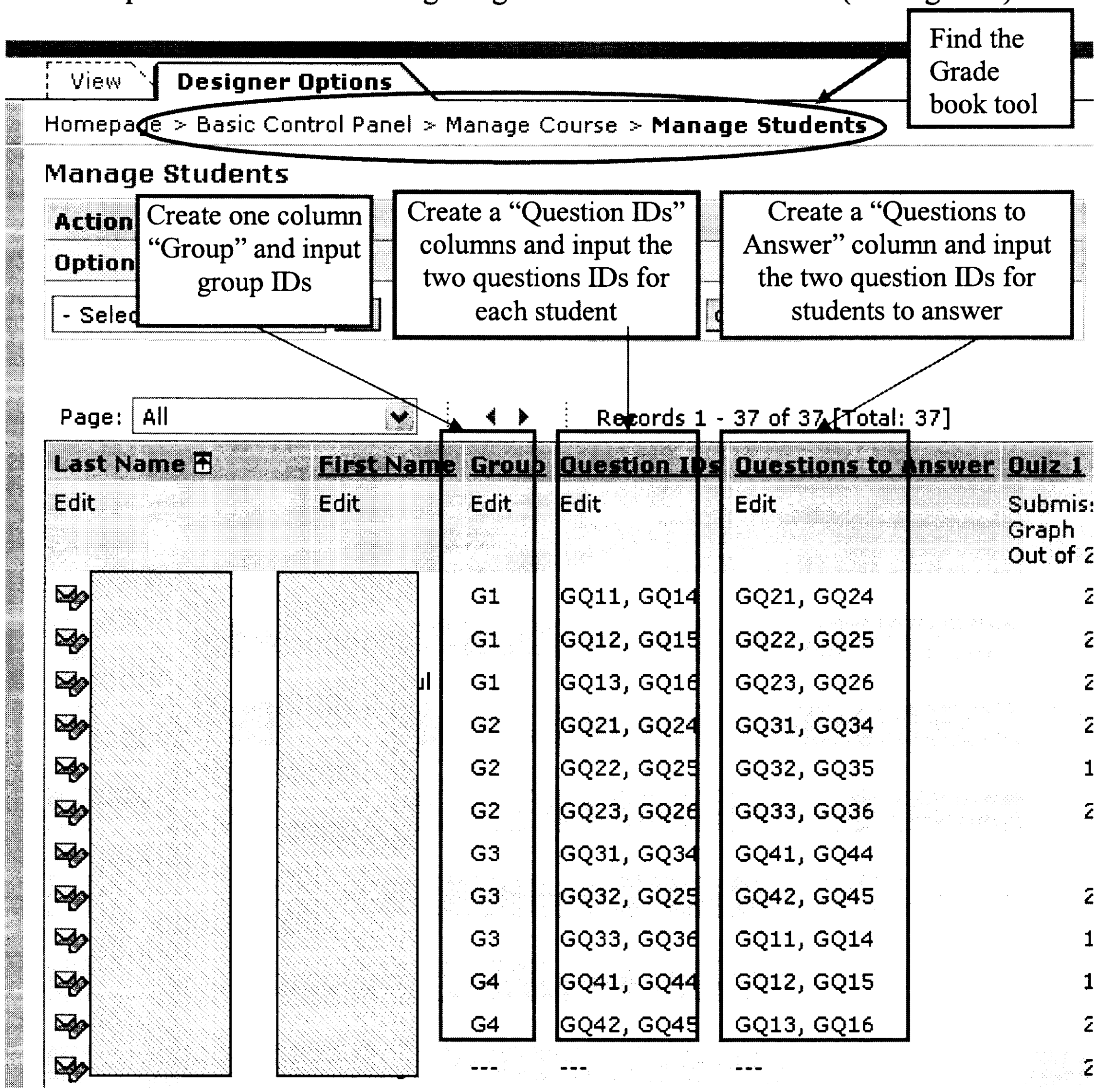


Figure 2 Using WebCT grade book to assign IDs.

The above three steps should be completed by Oct 15th. Then students are ready to answer the questions!

3. Review/Post final grades

Step 9. Review/Post Final Answer Grades

Once students have posted the answer grades, please review the question, answer, and grading justification, and provide a final answer grade. You should NOT grade answers from scratch. Please use the grading justification provided by the student grader to help you expedite your grading. To prevent and help to detect plagiarism, Jia has set up your course on Turnitin.com, and students were instructed to submit their answers there as well as on WebCT. Jia will check the report produced on Turnitin and inform you of any plagiarism cases, if any.

When you finish determining the grade, please post your final answer grading as a reply to the grading message posted by the student on WebCT. Your message should be titled in the format of "**GQxx- Final Answer Grade**" (see Figure 3). If you agree with the student grading, you may simply put in the message box "Agree with student grading. The final answer grade is: XX/35 points". If you need to grade on your own, please follow the answer grading guideline (see appendix 1). You should also update the two "answer grade" columns in the Excel file (see appendix 2).

Step 10. Grade Grading Justification Quality

You will give each group a "grading grade" based on the quality of grading and grading justifications with a maximum of 15 points. It is suggested that you provide grade justification grades while you are reviewing grades in step 9. All grade justification grades should be kept in the excel file (see appendix 2). Group members should receive the same grading grade, except for situations described below.

In case that some group member does not submit any grading in the private group conference, please inform the group to submit the group results before the deadline without the member. Also remind the group leader to keep an eye on any late postings within 48 hours past the deadline, and copy the message to the public exam conference. The late submissions will be graded separately according to the grading criteria. See the grading criteria at the end of this document for details.

Please complete the above two steps by Oct 24th.

4. Resolve disputes

Step 11: Re-evaluation of the Grade

Students are given one day to dispute your final answer grades. If anyone disputes the grade, please first examine whether the dispute meets with all of the three criteria in the student instruction document: disputing a significant number of points (more than 5 points total per grading disputed), re-grade one's own answer fully, and providing compelling justification using the grading guideline (see appendix 1). Dispute that does not meet with all the requirements should be discarded.

If the dispute meets the requirements, you should re-evaluate all aspects of the grading, including the question, answer, grading justification, and especially the self-evaluation and justification provided in the dispute. You may either increase or decrease the original final answer grade through the re-evaluation. Please post your resolution of the dispute as a reply to the dispute message. The title of the message should be question ID with "Dispute Resolve" (e.g. "**GQ11- Dispute Resolve**") (see Figure 3). Please update the excel file with the dispute results (see appendix 2).

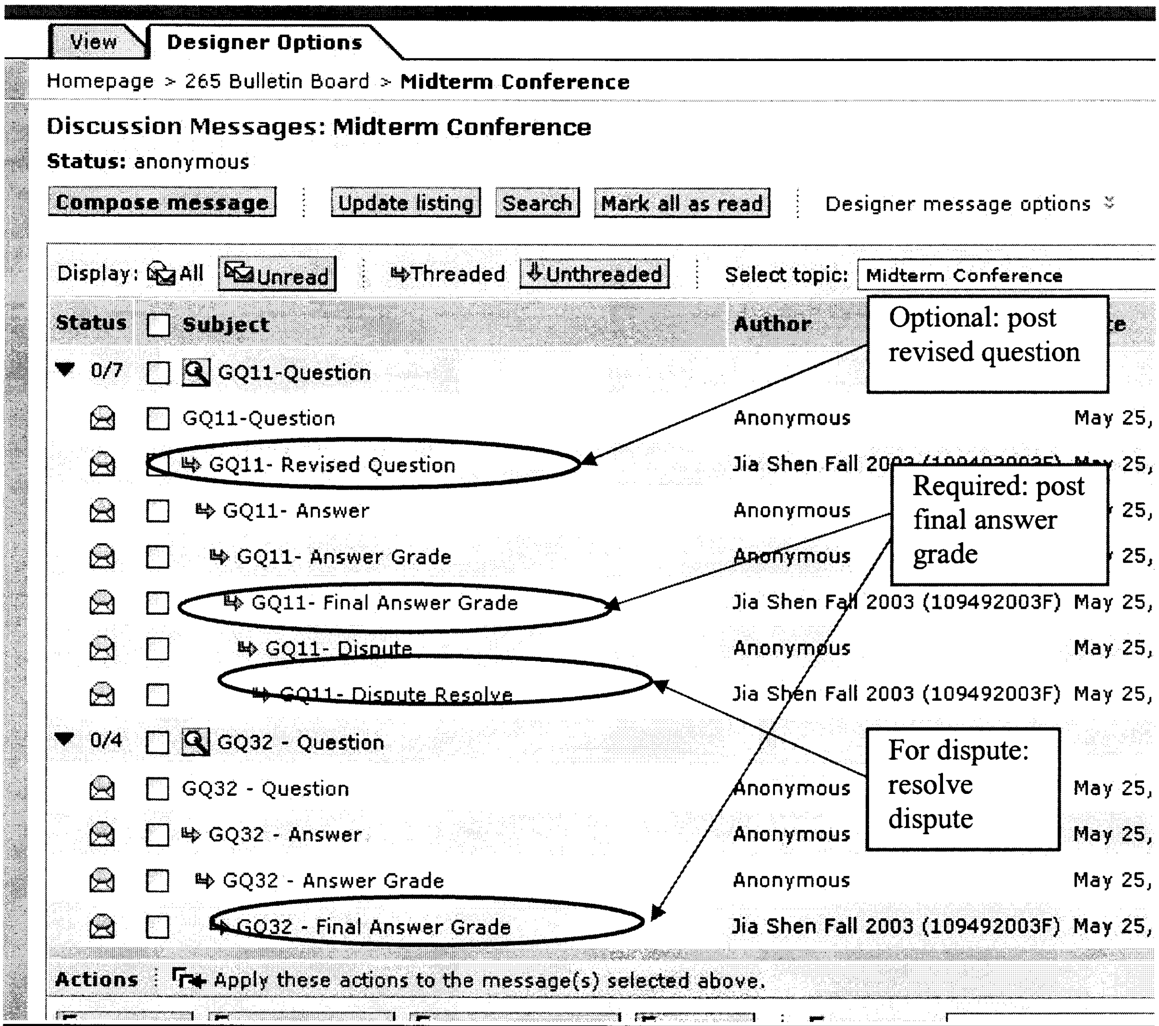


Figure 3 Using WebCT conference for the online small group exam.

You have now completed the online exam!

Final Step: Total exam grade calculation

The total exam grade should be calculated as the sum of the following four grades:

Question grade (15 points) + Answer grades (70 points) + Grade justification grade (15 points) + Extra credits total (16 points maximum)

You should have given the first three grades and filled out the Excel spreadsheet during the exam. Jia will provide you with the extra credits within 1 week after the exam based on the results of the pre and post exam surveys, exam log, and group leader information.

Appendix 1 - Grading guidelines

1. Question Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Questions cover different aspects of course material (3 points), are within the exam scope (3 points), and are identified with difficulty levels (3 points). • 4 points: Clarity of questions, including quality of the writing.
Following Directions	• 2 points: Submitting in the correct place and format in the WebCT conference. (Postings not submitted anonymously, or not submitted in the correct place, will be deducted 2 points) • Late submission: minus 4 points for submissions within 24 hrs past the deadline; minus 8 points for submissions past 24 hours within 48 hrs of the deadline; no submissions will be accepted after 48 hrs.

2. Answer Grading Criteria– used by students and the professor

(Total 70 points (35 points each answer)– Individual grade)

Quality	• 23 points: The correctness and completeness of the answer, including citing most of all relevant course materials, considering all sides of issues, synthesizing etc. • 7 points: The quality and clarity of writing, including providing justification to points, etc.
Following Directions	• 3 points: Following editing guidelines including correct citation format (deduct up to 3 points), and length (deduct 3 points if the answer is under 750 words or exceeds the 1,700 words limit.) • 2 points: Submission of the answer in the correct place and format • Late submission: minus 10 points for answers submitted within 24 hrs past the deadline; minus 20 points for answers submitted after 24 hrs within 48 hours of the deadline; no answer will be accepted after 48 hrs.

Plagiarism: All answers must be examined for plagiarism by considering: 1) whether the answer gives proper citation to the source of information; 2) whether the writing shows the student's own understanding of the knowledge. Proper citation must be used every time sentences or paragraphs are copied from books, papers, or other resources. Students should demonstrate their own understanding of the knowledge by explaining in their own words concepts, theories, methods, and/or providing summaries, examples, etc. If you think parts of the answer are plagiarized, please post the supporting material in your answer grading and grade appropriately.

3. Grade justification Grading Criteria– used by the professor

(Total 15 points– group grade, except specially advised by the instructor)

Quality	• 9 points: Quality of grading, including providing a full written explanation (justification) of the grading with at least 3 full sentences explanation for each of the grading categories. • 4 points: Clarity of justifications.
Following Directions	• Same as the question grading criteria

Appendix 2- Using Excel Spreadsheet to Manage ID Assignment and Grades

Last name	First name	Gp. ID	Q1 ID	Q1 Notes	Q2 ID	Q2 Notes	Ques. Grade	Grad. Grade	A1 QID	A1 Grade	A 1 Disp.	A 2 QID	A 2 Grade	A 2 Disp.	Extra Total	Sum
Smith	John	G1	GQ11	DFD/E	GQ14	T-test/E	15	13	GQ85	26		GQ86	34		0	90
Bush	Kenny	G1	GQ12	Case/H	GQ15	DFD/M	15	13	gq63	31		gq64	35		4	100
Manuel	Loren	G1	GQ13	Telecom/H	GQ16	Case/M	15	13	gq45	35		gq46	35		9	108
Clinton	Tracy	G3	GQ31	BPR/H	GQ34	Telecom/E	13	15	gq15	30		gq62	35		8	90
Patel	Bob	G3	GQ32	DFD/H	GQ35	Case/H	13	15	gq61	32		qg66	35		13	109
Wang	Bill	G3	GQ33	T-test/M	GQ36	yes	13	15	gq65	29		gq96	33		4	95
O'Brien	Will	G4	GQ41	BPR/E	GQ44	Telecom/M	10	15	gq95	34		gq16	30		8	87
Khan	Ray	G4	GQ42	Case/M	GQ45	DFD/E	10	0	gq83	34		gq84	31		6	100
Han	Adam	G4	GQ43	Telecom/E	GQ46	T-test/H	10	15	gq31	32		gq32	34		12	108
....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....

Step 2 Step 5 Step 6* Step 7 Step 10 Step 8 Step 9 Step 11 Step 12

Group grade Group Grade Individual Grade

*The Notes field for each question (Q1 Notes, Q2 Notes) may be used to record the topic area of each question (e.g. DFD, Telecom, T-test) and the difficulty level for better question assignment (Easy, Medium, Hard).

Appendix 3 - Instructor Exam Log Template

To help the research team investigate the time and effort instructors spend in conducting the online small group exam on WebCT, you are invited to record your exam related activities using the form below. Please record the date, and the duration (in hours) of your activities. Your activities are classified into the following five categories:

- **Administration** – this includes any effort in administering the exam. For example posting announcements, customizing guidelines for your course, assigning students with IDs, posting grades, etc. Notice this category refers to activities that only involve yourself, and it does not involve two-way communication with students.
- **Communication with students** – this refers to any activities that involve two-way communication with your students, such as answering students' questions, etc.
- **Question** – this refers any effort related to questions for the exam. For example, time spent reviewing all questions posted by students in online exams, etc.
- **Grade** – this refers to any effort related to grading students' answers. For example, time spent in reviewing and grading answers, and resolving disputes.
- **Other** – any other effort you put into the exam that is not listed in the categories above.

Your thorough completion of the log will greatly help us to investigate and reduce the workload for instructors in future studies. Thank you in advance!

Date	**Activity**				
	Admin.	**Communicate**	Question	**Grade**	**Other**
9/30/04		0.5 hour	1 hour		

The first line is filled in as an example

REFERENCES

Abowd, G. D. (1999). "Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment." IBM Systems Journal, special issue on pervasive computing 38 (4): 508-530.

Alavi, M. (1994). "Computer-Mediated Collaborative Learning: An Empirical Evaluation." MIS Quarterly 18 (2): 159-174.

Alavi, M. and Dufner, D. (2004). "Technology-Mediated Collaborative Learning: A Research Perspective". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Alavi, M. and Leider, D. (1999). "Knowledge Management Systems: Emerging Views and Practices from the Field." HICSS 1999.

Angelo, T. A. (1995). "Definition of Assessment." AAHE Bulletin, November: 7.

Angelo, T. A. and Cross, K. P. (1993). Classroom Assessment Techniques: A Handbook for College Teachers, Jossey-Bass.

Arbaugh, J. B. B. and Benbunan-Fich, R. (2004). "Contextual Factors That Influence ALN Effectiveness". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Artman, H. and Warn, Y. (1995). "Theoretical Perspectives on and Pragmatic Concerns for Situated Action in Relation to Information Technology." Prerequisites and Consequences of the Use of Information Technology.

Association of American Colleges (1985). Integrity in the College Curriculum: A Report to the Academic Community.

Bannon, L. J. and Kuutti, K. (1996). "Shifting Perspectives on Organizational Memory: From Storage to Active Remembering." HICSS 1996.

Bardram, J. E. (1997). "Plans as Situated Action: An Activity Theory Approach to Workflow Systems." Proceedings of ECSCW'97 Conference, Lancaster UK, September 1997.

Bartlett, F. C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge, England, The University Press.

Benbunan-Fich, R., Hiltz, S. R. and Harasim, L. (2004). "The Online Interaction Learning Model: An Integrated Theoretical Framework for Learning Networks".

Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Bennett, R. E. (Retrieved in November 2003). "Reinventing Assessment - Speculations on the Future of Large-Scale Educational Testing." ETS - Policy Information Center, Research Division ftp://ftp.ets.org/pub/res/reinvent.pdf.

Bennett, W. J. (1984). To Reclaim a Legacy: A Report on the Humanities in Higher Education. Washington, National Endowment for the Humanities.

Berge, Z. L. (1995). "Facilitating Computer Conferencing: Recommendations from the Field." Education Technology 35 (1): 22-30.

Biggs, J., Kember, D. and Leung, D. Y. P. (2001). "The Revised Two-Factor Study Process Questionnaire: R-Spq-2f." British Journal of Educational Psychology 71 (1): 133-149.

Biggs, J. B. (1987a). "Student Approaches to Learning and Studying." Camberwell, Vic.: Australian Council for Educational Research.

Biggs, J. B. (1987b). "The Study Process Questionnaire (SPQ): Manual." Hawthorn, Vic.: Australian council for educational research.

Biggs, J. B. (1994). "Student Learning Theory and Research: Where Do We Currently Stand?" Improving Student Learning: Theory and Practice. Gibbs, G. Oxford, Oxford brooks university.

Bloom, B. (1956). A Taxonomy of Educational Objectives: The Classification of Educational Goals. New York, D. McKay Co.

Boston, C. (2002). "The Concept of Formative Assessment." Practical Research, Assessment & Evaluation 8 (9).

Boverie, P., Nagel, L., McGee, M. and Garcia, S. (1998). "Predictors of Satisfaction for Distance Learners: A Study of Variable Conditions." SIGCUE Outlook 26 (2).

Bradford, K. (Retrieved in March 2004). "Deep and Surface Approaches to Learning and the Strategic Approach to Study in Higher Education; Based on Phenomenographic Research." http://www.arasite.org/guestkb.htm

Bridgwater, W. and Kurtz, S. (1963). The Columbia Encyclopedia. New York & London, Columbia University Press.

Brown, J. S., Collins, A. and Duguid, P. (1989). "Situated Cognition and the Culture of Learning." Educational Researcher 18: 32-42.

Brown, J. S. and Duguid, P. (1991). "Organizational Learning and Communities of Practice: Toward an Unified View of Working, Learning and Innovation." Organization science 2 (1): 40-57.

Bruffee, K. A. (1999). "Collaborative Learning: Higher Education". Interdependence and the Authority of Knowledge. Baltimore, MD, John Hopkins University Press.

Cole, M. (1991). "On Socially Shared Cognitions". Socially Shared Cognitions. Resnick, L., Levine, J. and Behrend, S. Hillsdale, NJ, Lawrence Erlbaum Associates: 398-417.

Coppola, N., Hiltz, S. R. and Rotter, N. (2001). "Becoming a Virtual Professor: Pedagogical Roles and ALN." Proceedings of the 34th Hawaii International Conference on Systems Sciences, IEEE Computer Society Press, Los Alamitos, CA, (Maui, January 3-6, 2001).

Corbitt, G., Wright, L. and Martz, B. (1999). "Addressing the Challenges of the Future: Implementing a Collaborative Student Environment at a University Business School." Proceedings of the 32nd Hawaii International Conference on System Science.

Creswell, J. W. (1994). Research Design: Qualitative & Quantitative Approaches, Sage Publications.

Daft, R. L. and Lengel, R. H. (1986). "Organizational Information Requirements, Media Richness and Structural Design." Management Science 32 (5): 554-571.

Deming, W. E. (1986). Out of the Crisis. Cambridge, Massachusetts, Massachusetts Institute of Technology.

Dennis, A. R., Jay F. Nunamaker, J. and Vogel, D. R. (Winter1990/91). "A Comparison of Laboratory and Field Research in the Study of Electronic Meeting Systems." Journal of Management Information Systems 7 (3).

Dewey, J. (1897). "My Pedagogic Creed." School Journal LIV (3): 77-80.

Dewey, J. (1963). Experience and Education. New York, MacMillan Press.

Drucker, P. (1998). "Harvard Business Review on Knowledge Management." Harvard Business School Press.

Dzuiban, C., Shea, P. J. and Arbaugh, J. B. B. (2004). "Faculty Roles and Satisfaction in Asynchronous Learning Networks". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

EdSource (Retrieved in December 2003). "Assessment Overview." http://www.edsource.org/edu_ass.cfm.

Entwistle, N. (2000). "Promoting Deep Learning through Teaching and Assessment". AAHE's 2000 Assessment Conference: Rising expectations for assessment: can we deliver? American Association for Higher Education.

Entwistle, N. and Ramsden, P. (1983). Understanding Student Learning. London, Croom Helm.

Erhaut, M. (1994). Developing Professional Knowledge and Competence. London, The Falmer Press.

Erwin, T. D. (1991). "Assessing Student Learning and Development", Jossey-Bass: 14-19.

Fagrell, H., Ljungberg, F. and Kristoffersen, S. (1999). "Exploring Support for Knowledge Management in Mobile Work." ECSCW 1999.

Falchikov, N. and Goldfinch, J. (2000). "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks." Review of Educational Research 70 (3): 287-322.

Fjermestad, J., Hiltz, S. R. and Zhang, Y. (2004). "Effectiveness for Students: Comparisons of "in-Seat" and ALN Courses". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Flynn, J. L. (1992). "Cooperative Learning and Gagne's Events of Instruction: A Syncretic View." Educational Technology October: 53-60.

Gardner, H. (1991). The Unschooled Mind: How Children Think and How Schools Should Teach. New York: Basic Books.

Gardner, H. (Retrieved in Oct. 2003). "Reinventing Our Schools: A Conversation with Howard Gardner." http://www.ed.psu.edu/insys/ESD/gardner/Assess.html.

Garrison, D. R., Anderson, T. and Archer, W. (2000). "Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education." Unpublished manuscript.

Gordon, C. and Debus, R. (2002). "Developing Deep Learning Approaches and Personal Teaching Efficacy within a Pre-service Teacher Education Context." British Journal of Educational Psychology 72 (4): 483-511.

Gunawardena, C. and Zittle, F. (1997). "Social Presence as a Predictor of Satisfaction within a Computer Mediated Conferencing Environment." American journal of Distance Learning 11 (3): 8-26.

Harasim, L. (1990). On-Line Education: Perspectives on a New Environment. New York, Praeger.

Hathorn, L. G. and Ingram, A. L. (2002). "Cooperation and Collaboration Using Computer-Mediated Communication." Journal of Educational Computing Research 26 (3): 325-347.

Hawkins, J. (May 1993). "Technology and the Organization of Schooling." communications of the ACM 36 (5): 30-34.

Hiltz, S. R. (1986). "The Virtual Classroom: Using Computer-Mediated Communication for University Teaching." Journal of Communication 36 (2): 95-104.

Hiltz, S. R. (1988). "Productivity Enhancement from Computer-Mediated Communication: A System Contingency Approach." CACM 31 (12): 1438-1454.

Hiltz, S. R. (1994). The Virtual Classroom: Learning without Limits Via Computer Networks. Norwood, NJ, Ablex.

Hiltz, S. R. (1996). "Post-Course Questionnaire for Students - Virtual University Project."

Hiltz, S. R., Coppola, N., Rotter, N., Turoff, M. and Benbunan-Fich, R. (2000). "Measuring the Importance of Collaborative Learning for the Effectiveness of ALN: A Multi-Measure, Multi-Method Approach." Journal of Asynchronous Learning Networks 4 (2).

Hiltz, S. R. and Shea, P. J. (2004). "The Student in the Online Classroom". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Hollan, J. D., Hutchins, E. and Kirsh, D. (1999). "Distributed Cognition: A New Foundation for Human-Computer Interaction Research." TOCHI special issue on human-computer interaction in the new millennium.

Huba, M. E. and Freed, J. E. (1999). Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning, Pearson Allyn & Bacon.

Hutchins, E. (1995). Cognition in the Wild, MIT Press.

Johnson, D. W. and Johnson, R. T. (1996). "Cooperation and the Use of Technology". Handbook of Research for Educational Communications and Technology. Honassen, D. H. New York, Simon and Schuster Macmillan: 1017-1044.

Jonassen, D. H. (1993). "Thinking Technology." Educational Technology January (35-37).

Kember, D., Charlesworth, M., Davies, H., McKay, J. and Stott, V. (1997). "Evaluating the Effectiveness of Educational Innovations: Using the Study Process Questionnaire to Show That Meaningful Learning Occurs." Studies In Educational Evaluation 23 (2): 141-157.

Kwok, R. C. W. and Ma, J. (1999). "Use of a Group Support System for Collaborative Assessment." Computers & Education 32 (2): 109-125.

Ledlow, S. (Retrieved in Nov. 2003). "Group Grades in Cooperative Learning Classes." http://www.public.asu.edu/~ledlow/sledlow/group.htm.

Leidner, D. and Fuller, M. (1997). "Improving Student Learning of Conceptual Information: GSS Supported Collaborative Learning Vs. Individual Constructive Learning." Decision Support Systems 20: 149-163.

Leidner, D. E. and Jarvenpaa, S. L. (1995). "The Use of Information Technology to Enhance Management School Education: A Theoretical View." MIS Quarterly 19 (3): 265-291.

Liu, C.-C., Liu, B.-J., Hui, T.-A. and Horng, J.-T. (2002). "Web Based Peer Assessment Using Knowledge Acquisition Techniques: Tools for Supporting Contexture Awareness". Computers in Education, IEEE.

Lorenzo, G. and Moore, J. (2002). "Five Pillars of Quality Online Education." The Sloan Consortium Report to the Nation.

Marton, F. and Saljo, R. (1976). "On Qualitative Differences in Learning, Outcome and Process." British Journal of Educational Psychology 46 (4-11).

McLoughlin, C. and Luca, J. (2002). "A Learner-Centered Approach to Developing Team Skills through Web-Based Learning and Assessment." British Journal of Educational Technology 33 (5).

McVarish, J. and Solloway, S. (2002). "Self-Evaluation: Creating a Classroom without Unhealthy Competitiveness." Educational Forum 66 (3): 253-260.

Mennecke, B. E., Hoffer, J. A. and Valacich, J. S. (1995). "An Experimental Examination of Group History and Group Support System Use on Information Sharing

Performance and User Perceptions." 28th Hawaii International Conference on System Sciences.

National Commission on Excellence in Education (1983). A Nation at Risk: The Imperative for Educational Reform. Washington, U.S. Department of Education.

National Institute of Education (1984). Involvement in Learning: Realizing the Potential of American Higher Education. Washington, US Department of Education.

Nonaka, O. and Takeuchi, H. (1995). The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York, Oxford university press.

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D. R. and George, J. F. (1991). "Electronic Meeting Systems to Support Group Work." CACM 34 (7): 40-61.

O' Loughlin, M. (1992). "Rethinking Science Education: Beyond Piagetian Constructivism toward a Sociocultural Model of Teaching and Learning." Journal of Research in Science Teaching 29 (8): 791-820.

Okada, M., Tarumi, H., Yoshimura, T. and Moriya, K. (April 2001). "Special Issue on Sac 2001 Best Papers: Collaborative Environmental Education Using Distributed Virtual Environment Accessible from Real and Virtual Worlds." ACM SIGAPP Applied Computing Review 9 (1).

Olson, L. (1992). Quality-Management Movement Spurs Interest in New Awards for Education. Education Week. 11: 8.

Paulsen, D. (October 2004). "Leadership Essentials: Racilitation Skills for Improving Group Effectiveness." Proceedings of the 32nd annual ACM SIGUCCS conference on User services.

Perry, W. G. (1970). Forms of Intellectual and Ethical Development in the College Years. New York, Holt, Rinehart and Winston.

Piaget, J. (1969). The Mechanisms of Perception. London, Rutledge & Kegan Paul.

Piaget, J. (1970). Science of Education and the Psychology of the Child. NY, Grossman.

Picciano, A. G. (2002). "Beyond Student Perceptions: Issues of Interaction, Presence, and Performance in an Online Course." Journal of Asynchronous Learning Networks 6 (1): 21-40.

Pimmel, R. L. (2003). "A Practical Approach for Converting Group Assignments into Team Projects." IEEE Transactions on Education 46 (2): 273- 282.

Polanyi, M. (1967). The Tacit Dimension. New York, Anchor.

Ramsden, P. and Entwistle, N. (1981). "Effects of Academic Departments on Students' Approaches to Studying." British Journal of Educational Psychology 51 (368-383).

Resnick, L. (1987). Education and Learning to Think. Washington D. C., National Academy Press.

Rhodes, L. A. (1992). On the Road to Quality. Educational leadership. 49: 76-80.

Rice, R. E., Hiltz, S. R. and Spencer, D. S. (2004). "Media Mixes and Learning Networks". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Richardson, J. and Swan, K. (2003). "An Examination of Social Presence in Online Learning: Students' Perceived Learning and Satisfaction." Journal of Asynchronous Learning Networks 7 (1).

Richardson, J. T. E. (1994). "Cultural Specificity of Approaches to Studying in Higher Education: A Literature Survey." Higher Education 27 (449-468).

Rosenthal, R. and Rosnow, R. (1991). Essentials of Behavioral Research: Methods and Data Analysis, McGraw-Hill.

Rößling, G., Bär, H., Trompler, C. and Choi, C.-M. (June 2003). "Mobile Computing in Education." ACM SIGCSE Bulletin, Proceedings of the 8th annual conference on Innovation and technology in computer science education 35 (3).

Rudner, L. and Boston, C. (Winter 1994). "Performance Assessment." ERIC Review 3, no. 1: 2-12

Rushton, C. (Sum 1993). "Peer Assessment in a Collaborative Hypermedia Environment: A Case Study." Journal of Computer-Based Instruction 20 (3): 75-80.

Schlechter, T. M. (1990). "The Relative Instructional Efficiency of Small Group Computer-Based Training." Journal of educational Computing Research 6 (3): 329-341.

Shen, J., Bieber, M. and Hiltz, S. R. (2004b). "Collaborative Examinations for Asynchronous Learning Networks: Three Evaluation Results." submitted to the Journal of ALN.

Shen, J., Cheng, K.-E., Bieber, M. and Hiltz, S. R. (2004). "Traditional in-Class Examination Vs. Collaborative Online Examination in Asynchronous Learning

Networks: Field Evaluation Results". Americas Conference on Information Systems 2004, New York City, NY.

Shen, J., Cheng, K.-E., Cho, Y., Hiltz, S. R. and Bieber, M. (2000). "Evaluation of an on-Line Collaborative Examination Process". Proceedings of the 2000 Americas Conference on Information Systems, California.

Shen, J., Hiltz, S. R., Cheng, K.-E., Cho, Y. and Bieber, M. (2001). "Collaborative Examinations for Asynchronous Learning Networks: Evaluation Results". Proceedings of the 34th Hawaii International Conference on Systems Sciences, Maui, Hawaii, IEEE Computer Society Press.

Shen, J. and Patten, K. (April 2005). "Faculty Perspectives of Collaborative Online Examinations: Initial Report." ISOneWorld Conference.

Shindler, J. (2003). "Examining the Soundness of Collaborative Essay Exams in Teacher Education Courses." National Forum of Teacher Education Journal 12 (3).

Shindler, J. and Yang, H. (April, 2000). "Authentic Assessment in an Educational Technology Course: Examining the Soundness of a Process-Based System of Assessment." Annual Meeting of the American Educational Research Association, New Orleans, LA.

Short, J., Williams, E. and Christie, B. (1976). The Social Psychology of Telecommunication. London, John Wiley & Sons.

Shuell, T. J. (1986). "Cognitive Conceptions of Learning." Review of Educational Research Winter: 411-436.

Slavin, R. E. (1990). Cooperative Learning: Theory, Research and Practice. Englewood Cliffs, NJ, Prentice Hall.

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

Sluijsmans, D., Dochy, F. and Moerkerke, G. (1999). "Creating a Learning Environment by Using Self-, Peer- and Co-Assessment." Learning Environments Research 1: 293-319.

Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merriënboer, J. and Bastiaens, T. R. (2003). "The Training of Peer Assessment Skills to Promote the Development of Reflection Skills in Teacher Education." Studies in Educational Evaluation 29 (1): 23-42.

Southern Regional Education Board (1985). Access to Quality Undergraduate Education. The Chronicle of Higher Education.

Steinkuehler, C. A. and Derry, S. J. (Retrieved in Oct. 2003). "Measuring Learning Effectiveness." http://www.alnresearch.org/JSP/Tutorials/tutorial_frame.jsp.

Stevens, S. S. (1951). "Mathematics, Measurement, and Psychophysics". Handbook of Experimental Psychology. Stevens, S. S. New York, Wiley: 1-49.

Stewart, T. A. (1997). Intellectual Capital: The New Wealth of Organizations.

Suchman, L. (1987). Plans and Situated Actions. Cambridge, Cambridge University Press.

Sunny San-Ju Lin Liu, E. Z.-F. S.-M. Y. (2001). "Web Based Peer Assessment: Attitude and Achievement." IEEE Transactions on Education 44 (2): 13.

Suskie, L. (2000). "Assessment to Promote Deep Learning: Foreword to the Aahe's Conferences on Assessment, 2000 and 1999." AAHE Bulletin http://www.aahe.org/pubs/Sample%20Chapters/Suskie%20foreword.htm.

Swan, K. (2004). "Relationships between Interactions and Learning in Online Environments." The Sloan Consortium.

Swan, K. and Shea, P. J. (2004). "The Development of Virtual Learning Communities". Learning Together Online: Research on Asynchronous Learning Networks. Hiltz, S. R. and Goldman, R. Mahwah, NJ, Lawrence Erlbaum Associates (In Press).

Tang, C. (1994). "Effects of Modes of Assessment on Students' Preparation Strategies". Improving Student Learning - Theory and Practice. Gibbs, G. Oxford, Oxford Center for Staff Development.

Tang, K. C. C. (1991). "Effects of Different Assessment Procedures on Tertiary Students' Approaches to Studying." Unpublished PhD Thesis. The University of Hong Kong.

Tiwari, A. and Tang, C. (2001). "Does Portfolio Assessment Encourage Students to Adopt Different Assessment Preparation Strategies?" The Scholarship of Teaching & Learning: the second Hong Kong Conference on Quality in Teaching and Learning in Higher Education, Hong Kong.

Topping, K. J. (1998). "Peer Assessment between Students in College and University." Review of Educational Research 68 (3): 249-276.

Vogel, D., Wagner, C. and Ma, L. (1999). "Student-Directed Learning: Hong Kong Experiences." Thirty-Second Annual Hawaii International Conference on System Sciences.

Vygotsky, L. (1962). Thought and Language. Cambridge, MA, MIT Press.

Weaver, T. (1992). "Total Quality Management, Eric Digest Number 73." ERIC Clearinghouse on Educational Management.

Weiner, M. and Mehrabian, A. (1968). Language within Language: Immediacy, a Channel in Verbal Communication. New York, Appleton Century Crofts.

Weiser, M. (1998). "The Future of Ubiquitous Computing on Campus." Communication of the ACM 41 (1): 41-42.

Wenger, E. (1998). Communities of Practice: Learning, Meaning, and Identity, Cambridge University Press.

Wilson, E. V. (2004). "Examnet Asynchronous Learning Network: Augmenting Face-to-Face Course with Student-Developed Exam Questions." Computers & Education 42: 87-107.

Wright, B. D. (2003). "More Art Than Science: The Postsecondary Assessment Movement Today." Sloan-C Consortium Workshop.

Wu, D., Bieber, M., Hiltz, S. R. and Han, H. (2004). "Constructivist Learning with Participatory Examinations". Proceedings of the HICSS 37th Hawaii International Conference on Systems Sciences, Big Island, Hawaii, IEEE Computer Society Press, CD-Rom.

Yarusso, L. (1992). "Constructivism Vs. Objectivism." Performance and Instruction Journal April (7-9).