EFFECTS OF ANONYMITY AND PEER ACCOUNTABILITY DURING PEER ASSESSMENT IN A GRADUATE WEB-BASED EDUCATION RESEARCH METHODS COURSE

GUNITA WADHWA

NOTE TO USERS

EFFECTS OF ANONYMITY AND PEER ACCOUNTABILTY DURING PEER ASSESSMENT IN A GRADUATE WEB-BASED EDUCATION RESEARCH METHODS COURSE

by
© Gunita Wadhwa

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirements for
the degree of Master of Education

Faculty of Education
Memorial University of Newfoundland

December 2003

St. John's, Newfoundland

Your file Votre référence
ISBN: 0-612-99126-1
Our file Notre référence
ISBN: 0-612-99126-1

ABSTRACT

The aim of this master's thesis was to explore the effects of anonymity and peer accountability on peer marking, and the criticality and quality of peer comments during online peer assessment. Thirty-six graduate students in a web-based education research methods course were asked to critique two research articles. Peer assessment was carried out on the students' first critique. Peer assessment involved the peer assessors assigning a numeric mark and qualitative comments on other students' critiques. An experiment was conducted to determine the effects of anonymity (anonymous vs. named) and peer-accountability (more-accountable vs. less-accountable) on peer over-marking, the number of critical, and quality comments made by the peer assessors during online peer assessment. The three main results were: First, significantly ($p < .04$) fewer peer assessors *over-marked* (i.e., peer assessors assigned a higher mark relative to the instructor) in the anonymous group, compared to the named group. Second, the peer assessors in the anonymous group provided a significantly ($p < .01$) higher number of *critical comments* (i.e., the number of negative comments or weaknesses), compared to the named group. Third, the peer assessors in the named group and the more-accountable group made a significantly ($p < .01$) higher number of *quality comments* (i.e., number of cognitive statements indicating strengths and weakness along with reasoned responses and suggestions for

improvement), compared to the peer assessors in the anonymous group and the less-accountable group. No conclusive results could be derived to indicate improvement in the students' performance in critiquing the research articles. However, students' responses to the questionnaire indicated that they found the peer assessment process helpful. This study suggests that in online peer assessment, the interaction of anonymity and the degree of peer accountability affects peer marking and peer comments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks, appreciation and gratitude to Dr. Bruce Mann, my thesis supervisor, for his continuous support, expert advice, timely guidance and helpful encouragement during the entire process of my master's thesis research and writing.

My deep appreciation goes to my co-supervisor, Dr. Henry Schulz, for his extraordinary support and advice on the statistical analysis of the data, selection of the participants, development of the instructional materials, and the methodological issues of this master's thesis study.

My sincere thanks to my family and friends for their continuous encouragement and support throughout my thesis study.

CHAPTER 1

Introduction

Chapter one is an introduction to the purpose of this master's thesis research on online peer assessment in a graduate web-based education research methods course. This thesis study examined ways to minimize two problems in peer assessment, namely: peer-assigned marks and peer comments. The concern regarding peer-assigned marks is that the peer assessors have a tendency to *over-mark* (i.e., peer assessors assign a higher mark relative to the instructor), thus affecting the validity of the peer assessment process (e.g., Boud & Homes, 1995; Falchikov, 1986, 1995; Kelmar, 1993; Pond, Rehan, & Wade, 1995). The concern regarding peer comments is that the peer assessors are reluctant to indicate weaknesses or provide critical comments in their assessment of other students' work (e.g., Falchikov, 1995, 1996; Fenwick & Parsons, 2000; Topping, Smith, Swanson, & Elliot, 2000). Since critical feedback is deemed to be important for learning, this inconsistency in peer comments may affect the learning benefits expected from the peer assessment process. This discrepancy in peer marking and peer comments may be due to loyalty towards friends and social pressure (Falchikov, 1995; Sluijsmans, Moerkerke, Dochy & Merriënboer, 2001;Topping et al., 2000).

In this master's thesis research, two variables: (1) anonymity and (2) peer accountability were examined to determine their effects on the peer-assigned marks and comments.

Anonymity. Since anonymity helps relieve social pressure and inhibition, it is expected to encourage accurate, honest and critical response. However, empirical evidence on the effects of anonymity is unclear. Some studies (e.g., Connolly, Jessup, & Valacich, 1990; Makkai & McAllister, 1992; McCollister, 1985), as expected, found that anonymity enhanced more accurate and critical response. However, other studies (e.g., Harkins & Petty, 1992; Zhao, 1998) concluded that while anonymity reduced social pressure, it also reduced responsibility resulting in careless and less concerned response by the participants. Therefore, interaction of anonymity with another variable, such as, accountability, may help in improving participants' response in a meaningful and constructive way.

Peer accountability. Studies (Gordon & Stuecher, 1992; Price, 1987) on the effect of accountability on student responses found that participants put more effort in their responses when they were asked to justify their comments or when others review their decision. These studies further suggest that varying the degree of accountability may also affect the quality of comments. Although there is no empirical evidence in online peer

assessment to indicate the effect of accountability on peer comments, some researchers (e.g., Topping et al., 2000; Zhao, 1998) suggest that incorporating peer accountability in peer assessment may improve the accuracy and quality of peer comments. Hence, the interaction of anonymity with accountability may affect participants' responses (Pinsonneault & Nelson, 1998; Zhao, 1998).

Therefore, the purpose of this master's thesis research was to determine the effects of anonymity and peer accountability on peer assigned-marks, and the criticality and quality of peer comments during peer assessment in a graduate web-based education research methods course. *Anonymity* (anonymous vs. named) was defined as the condition where the peer assessors and the students assessed were in the anonymous group or the named group. In the *anonymous group*, a number replaced the names of the peer assessors and the students assessed. In the *named group*, the peer assessors and the students assessed were identified by their names. *Peer accountability* (more-accountable vs. less-accountable) was defined as the condition where the peer assessors were in the more-accountable group or the less-accountable group. In the *more-accountable group*, the peer assessors were told that timely submissions of their assessment and the quality of their comments would contribute to their participation mark in the course. In the *less-accountable group* the peer assessors were only told about the timely submissions of their assessments contributing to their

participation mark for the course. The participants were randomly assigned to one of the four groups (i.e., named, more-accountable group; anonymous group, more-accountable group; named, less-accountable group; and anonymous, less-accountable group). To determine the effects of the two variables (anonymity and peer accountability), the marks and comments assigned by the peer assessors in all the four groups were examined and compared.

Chapter one was an introduction to the purpose of this master's thesis research. Chapter two will provide a review of the extant literature on the effect of friendship, anonymity and accountability on peer assessment.

CHAPTER 2

Literature Review

Chapter two is a review of the extant literature on the issues in peer assessment, the cognitive benefits of peer assessment, and the aim of this master's thesis research. In the first section, the issues in peer assessment related to peer-assigned marks and peer comments are discussed. In addition, the literature on the issues related to the two independent variables: (a) anonymity and (b) peer accountability examined in this study is summarized. In the second section, cognitive benefits of peer assessment with a focus on graduate education research methods course is reviewed. Finally, in the third section, the aim of the study and the research questions addressed in this master's thesis are stated. The questions addressed in this master's thesis determine how anonymity and peer accountability affect peer marking and peer comments during peer assessment in a graduate web-based education research methods course. A comprehensive review of the extant literature published in and before 2003, shows limited research on the effects of anonymity in online peer assessment. Further, the literature does not show empirical evidence on the effects of peer accountability in online peer assessment. Therefore, the hypotheses in this research were generated based on the literature on peer assessment in face-to-face environment.

2.1 Issues in Peer Assessment

Peer assessment is a process in which a group of individuals assess and rate each other's work (Falchikov, 1995; Topping, Smith, Swanson, & Elliot, 2000). Issues in peer assessment affecting the reliability and validity of peer ratings are manifold. Some of these issues identified in the literature on peer assessment are: *friendship marking*, where peer assessors tend to over-mark due to friendships and social pressure (Borman, White, & Dorsey, 1995; Dancer & Dancer, 1992; Falchikov, 1995; Falchikov & Goldfinch, 2000; Helmore & Magin, 1998; Magin, 2001; Pond, Rehan, & Wade, 1995; Slujismans, Moerkerke, Dochy, & Van Merriënboer, 2001;Topping et al., 2000); *raters style,* where peer assessors may differ in their severity or leniency in assigning marks on other students' work (Pond et al., 1995; Slujismans et al., 2001; Swanson, Case, & Vleuten, 1991); *marking criteria*, where different peer assessors may use different marking criteria to assess the same topic (Falchikov & Goldfinch, 2000; Orsmond, Merry, & Reiling, 2000; Stefani, 1994); *ability of the peer assessor*, where the ability of the peer assessors and raters knowledge of the content may affect peer marking (Jacobs, Briggs, & Whitney, 1975) ; *raters thinking styles*, where peers with different thinking styles (high-executive and low-executive thinking styles) may differ in their ratings (Lin, Liu, & Yuan, 2001); and *gender effects,* where peer ratings may differ due to gender bias (Falchikov & Goldfinch, 2000; Falchikov & Magin, 1997).

A concern indicated in most studies (Borman et al., 1995; Dancer & Dancer, 1992; Falchikov, 1995; Falchikov & Goldfinch, 2000; Helmore & Magin, 1998; Magin, 2001; Pond et al., 1995; Slujismans et al., 2001;Topping et al., 2000), is that of friendships, social relationships and loyalty towards friends affecting peer-assigned marks and peer comments.

Peer-assigned marks. A common concern with peer-assigned marks is that peer assessors have a tendency to over-mark: assign higher marks relative to the instructor (Boud & Homes, 1995; Falchikov, 1986, 1995; Kelmar, 1993; Pond et al., 1995; Mowl & Pain, 1995; Rushton, Ramsey & Rada, 1993; Sluijsmans et al., 2001). This inconsistency in peer marking may affect the validity of the peer assessment process.

Peer comments. A concern about peer comments is that peer assessors are reluctant to indicate weaknesses or provide critical comments in their assessment of other students' work (Falchikov, 1995, 1996; Fenwick & Parsons, 2000; Topping et al., 2000). Studies (Falchikov, 1996; Searby & Ewers, 1997; Topping et al., 2000) show that peers are capable of providing more detailed, timely and critical feedback. Further, research suggests that critical feedback is crucial for learning (Miyake, 1987; Zhao, 1998). Therefore, the peer assessors' reluctance in providing critical feedback may affect the learning benefit expected from the peer assessment process.

However, there is lack of empirical evidence to address the issue of friendship and social pressure affecting peer marking and peer comments. This thesis study examined the factors that may help in reducing peer over-marking and enhancing critical comments in peer feedback in a meaningful way. Two independent variables were considered important: (a) anonymity and (b) peer accountability.

Anonymity. The concept of anonymity has been experimented in various settings and context, such as, students' response to teacher evaluation form (e.g., McCollister 1985; Stone, Spool, & Robinowtz, 1977), group interaction using computer-mediated communication (e.g., Connolly, Jessup, & Valacich, 1990; Kahai, Avolio, & Sosik, 1998; Pinsonneault & Nelson, 1998; Zhao, 1998), and professional environment (e.g., Antonioni, 1994; Hiltz, Turoff, & Johnson, 1989). However, empirical evidence on the effects of anonymity on the participants' response is inconclusive.

Some studies (Antonioni, 1994; Davis, 2000; Falchikov, 1995; Haaga, 1993; Makkai & McAllister, 1992; McCollister, 1985; Stone et al., 1977; Tsai, Liu, Lin, & Yuan, 2001) indicate that anonymity breaks down social barriers, reduces inhibition, and promotes honest responses. In a study to determine a method to improve accuracy of response on sensitive questions, Makkai and McAllister (1992) found that the participants' response in anonymous condition (in a sealed booklet) was more accurate compared to an identifiable

condition (direct questions). Studies (McCollister 1985; Stone et al., 1977) on the students' response to teacher-rating form showed that the students who had to sign their names on the evaluation form gave more positive ratings to their instructors compared to anonymous student evaluations. McCollister's (1985) interview study with the medical students also revealed that the students felt that signing their names on teacher-evaluation form inhibited them from giving their honest response to overall quality of teaching. In another study on subordinate workers evaluation of their superiors, Antonioni (1994) found that anonymity provoked more critical feedback. In Antonioni's study workers were asked to provide evaluative feedback on their superiors. The study found that anonymous subordinates gave lower ratings to their superior's compared to the worker's who had to sign their names.

Yet, other studies (Hiltz et al., 1989; Ellis, 1984) found no difference in participants' response due to anonymity. In a study on teacher evaluation, Ellis (1984) found no significant difference in the ratings made by anonymous and identifiable students. Similarly, Hiltz et al. (1989) found that anonymity had no effect on inhibition and the number of comments made by the participants.

However, some other studies (Bostock, 2000; Harkins & Petty, 1992; McBeatry, 1982; Zhao, 1998) found that by reducing social pressure anonymity reduced responsibility. This lack of responsibility may induce social loafing that may result in careless and less concerned responses. Therefore,

critical comments may not necessarily be meaningful and constructive. In a case study with graduate students, Bostock (2000) found that anonymous peer assessors provided more "ruthless" feedback. McBeatry (1982) also suggested that anonymous evaluations foster lack of personal responsibility resulting in careless comments. In an attempt to determine the effect of anonymity on critical feedback, Zhao (1998) found anonymity to be a "double-edged sword in collaborative learning" (p. 311). The results of his study indicated that while anonymity enhanced critical and ruder feedback, participants in anonymous group exerted less effort in providing comments compared to the identifiable group. Further, the comments made by the participants in the anonymous group were considered less helpful and of lower quality than those by the identifiable group.

Hence, the effect of anonymity may depend on the context in which it is used. Some studies (e.g., Connolly et al., 1990; Kahai & Avolio, 1998; Pinsonneault & Nelson, 1998), suggest that the type of discussions and the type of anonymity may also affect the participants' response. Zhao (1998) summarized different types of anonymity as: *Complete anonymity*, where all participants are completely anonymous to each other; *One-way anonymity*, where one of the participant's identities is concealed; *Social anonymity,* where the absence of social presence depersonalizes individual's identity. Therefore individual's identity need not be concealed to create anonymity. Some researchers (Berge & Collins 1993; Brodia, 1997; Bump, 1990; Lin et

al., 2001; Jonassen & Kwon, 2001) believe that the electronic medium of communication causes social anonymity. Therefore, concealing or revealing participants identity should not affect their responses (Bump, 1990).

Empirical studies (Brodia, 1997; Jonassen & Kwon, 2001; Zhao, 1998) on student responses with electronic communication have also shown mixed results. Under certain conditions anonymity may encourage participants to be more honest whereas in other situations it may lead to social loafing. In eighteen experimental studies comparing face-to-face and computer-mediated communication (CMC), Brodia (1997) found that participants in the CMC environment felt less social pressure compared to face-to-face environment. However, Brodia also reported that discussion in the CMC environment was of poorer quality compared to the face-to-face environment. On the other hand, in a study with undergraduate engineering students, Jonassen et al. (2001) found that the participants in CMC were more tasks oriented and provided better comments than in face-to-face environment.

Anonymity in web-based educational courses may be viewed differently. Although the web-based learning environment causes social absence, this may not necessarily cause social anonymity. The students in a web-based course may know each other socially or from another face-to-face course. Therefore, concealing or revealing the participants' identity may affect their responses. This study attempted to examine the effects of anonymity on

peer-assigned marks and peer comments during peer assessment in a web-based course.

In order to preserve the positive effects of anonymity and to minimize the concern associated with it, this thesis study also examined the effect of peer accountability on online peer assessment.

Peer accountability. Tetlock (1983) defined accountability as "a special type of transmission set in which one anticipates the need not only to communicate one's opinions, but also to defend those opinions against possible counterarguments" (p. 75). Empirical studies (Gordon & Stuecher, 1992; Price, 1987) on the effect of accountability on student responses to teacher-evaluation questionnaire and group interactions found that participants put in more cognitive effort in their responses when they were asked to justify their comments or when they knew that others would be reviewing their responses. Gordon & Stuecher (1992) examined the differences in students' responses on teacher-evaluation questionnaire, based on degree of accountability (high and low accountability). In their study, students were asked to complete two closed-ended and one open-ended question evaluating their professor. Students were placed in *high accountability condition*, in which they were asked to submit their responses to the faculty, and *low accountability condition*, in which they were asked to submit their responses to their peers. The results of their study indicated that

the students in the high accountability condition framed their responses more careful, with increased linguistic complexity, compared to the students in the low accountability conditions. In another experiment with anonymity and accountability, Price (1987) found that anonymity / identifiability had no impact when the group members were accountable for their decisions. In their study, with a 2 x 2 (decision responsibility x identifiability) design, they found that the individual and group efforts were less when the participants were anonymous and they knew that no one was monitoring their decisions compared to the condition when the participants knew that their responses were being reviewed, irrespective of anonymity condition.

Although, there is no empirical evidence on the effects of peer accountability in online peer assessment, some researchers (Davis, 2000; Topping et al., 2000; Tsai et al., 2001; Zhao, 1998) suggest that incorporating peer accountability in peer assessment may affect accuracy in peer marking and quality of peer comments. One way of incorporating peer accountability in peer assessment could be, the instructor assessing the peer assessor's assessment. In a study on computerized peer assessment with undergraduate computer science students, Davis (2000) reported that peer assessors took greater care in marking, since they (peer assessors) knew that they were being assessed on their ability in marking other student's work. In a study on networked peer assessment (Tsai et al., 2001) quality of peer assessor's comments were reviewed and graded by the instructor. This was

done to encourage assessors to provide helpful comments. Therefore, this thesis study examined the effects anonymity and peer accountability on peer marking, critical peer comments, and the quality peer comments in online peer assessment.

Despite the issues in peer assessment that may affect the reliability and the validity of the process, the strength of peer assessment process seems to be related to student learning by means of reflection, analysis and diplomatic criticism (Boud & Homes, 1995; Falchikov, 1995, 1986, 2001; Searby & Ewers, 1997; Topping & Ehly, 1998). Peer assessment has been used extensively in higher educational settings in diverse fields such as science, teacher-education, writing, and medicine (Falchikov, 1995; Rada, 1998, Sluijsmans, Saskia, & Merriënboer, 2002; Topping et al., 2000). The following section summarizes the cognitive benefits of peer assessment in a graduate web-based education research methods course.

2.2 Cognitive Benefits of Peer assessment in a Graduate Web-based Education Research Methods Course

Graduate students taking online courses in education research methods are often asked to critique, evaluate and analyze, published research (e.g., Hittleman & Simon, 2002). Literature shows that these kinds of critiquing skills improve with practice (Anderson, Howe, Soden, Halliday &

Low, 2001; Kuhn, 1991; MacPherson, 1999; Sluijsmans et al., 2002). As an instructional method, peer assessment exercises can be incorporated in the curriculum to provide practice in critiquing skills. Peer assessment activities often involve having students' exchange assignments, discuss responses and rate each other's work (Falchikov, 1995; Topping et al., 2000). Studies (Anderson et al. 2001; Blumhof, & Stallibrass, 1994; Falchikov, 1986; Pond et al.,1995; Searby & Ewers, 1997; Topping et al., 2000; Towler & Broadfoot, 1992) indicate that these peer assessment activities can help learners develop critical, evaluative, and analytical skills. Peer assessment that involves students' ability to make value judgments, analyze responses and provide reasoned arguments on other students work is beneficial for both, the peer assessor and the student assessed. The peer assessor learns from critically analyzing and evaluating other students' work and the student assessed learns from peer feedback (Falchikov, 1995; Freeman, 1995; O'Donnell & Topping, 1998; O'Donnell & King, 1999; Searby & Ewers, 1997; Webb, 1989). However, different types of peer interactions might generate positive effects through different mechanisms (Topping et al., 2000).

Empirical studies (Anderson et al. 2001; Allen 1992; Falchikov, 1995) on the cognitive benefits from peer-based exercises have shown positive results. In a study on peer interaction with further education college students, Anderson et al. (2001) found that peer assessment exercise helped students in improving their critical thinking skills. In Anderson's study, students' were

asked to develop a project that required background research on the topic and justification for its use. Students took part in peer-based exercises in which they were asked to critique each others work. The exercises were developed based on Kuhn's (1991) four key argumentative critical thinking skills, namely, (1) providing evidence for one's own theory, (2) envisioning alternative theories, (3) providing counter arguments, and (4) rebutting, in the context of devising the design of their projects and writing their reports. Students' interactions were videotaped. Analysis of students' dialogues and written work indicated that the students who had participated in the peer-based critiquing exercises had learned the importance of justifying arguments and they engaged in justification of their arguments to significantly greater degree than the control group. In another study on peer assessment with multimedia learning environment, Allen (1992) employed peer critiquing in adults, giving students the task of producing compositions on their chosen topics. Students were required to present versions of their compositions to each other, and to discuss each other's work critically. Allen's data suggested that this peer critiquing exercise added value to what was produced. Falchikov (1986) reported that undergraduate science students involved in peer assessment found that the assessment process "made them think more, learn more and become more critical and structured" (p. 161). In another empirical study, Webb (1989) highlighted on the kinds of peer interaction that influenced student learning. Webb's review of 19 empirical studies on peer

interaction in small groups suggests that there is a positive correlation between giving an elaborate explanation and the learner's achievement. She found that when peers provide elaborate feedback by explaining their ideas, their own understanding of the subject matter improved. Therefore, the peer assessors benefited from providing high-level elaboration to other members of a group. However, Webb's study did not indicate the method of analyzing comments.

Analyzing peer comments. Over the last decade, several methods, models, and principles have been developed to analyze the content of online student interaction (e.g., Ahern, Peck & Laycock, 1992; Chi, 1996; Feenberg, 1987; Gunawardena, Lowe, & Anderson, 1997; Hara, Bonk, & Angel, 2000; Henri, 1992; Howell-Richardson & Mellar, 1996; Mowrer, 1996; Romiszowski & Mason, 1996; Rourke & Anderson, 2002; Walther & Tidewell, 1995; Zhu, 1998). In this thesis study, the content analysis framework for online discussions developed by Henri (1992) and modified by Hara et al. (2000) was employed to analyze qualitative peer comments. Henri (1992) developed a content analysis model in which she identified five key dimensions for analysis of online discussions, namely, (1) participation rate (e.g. raw number and timing of messages); (2) interaction type (e.g. direct response, "in response to the posting…"); (3) Social cues (e.g." Its my birthday today"); (4) cognitive skills (e.g. judgment "I disagree with….") and depth of processing

(surface level or deep level processing); and (5) Meta-cognitive skills and knowledge (e.g. providing examples and relating to situations). Hara et al. (2000) paralleled Henri's (1992) recommendations of content analysis in online discussions and proposed similar guidelines to analyze electronic conversations. Hara et al. examined student comments during online discussions as (1) student participation rate; (2) electronic interaction patterns; (3) social cues within student messages; (4) cognitive and metacognitive components of student messages; and (5) depth of processing - surface or deep - within message posting. Henri analyzed each idea unit within a message whereas Hara et al. analyzed each message for the level of processing. Based on Hara et al. (2000) and Henri's (1992) methods of content analysis, this study categorized peer comments in two categories, namely, social comments, and quality (cognitive) comments.

According to Hara et al. and Henri, *social messages* were statements not related to formal content of subject matter. Social messages included self-introduction, greetings, jokes, and complimentary messages. Similarly, in this study, *social comments* were statements made by the peer assessors that were not related to a specific content area. *Cognitive comments* as stated by Hara et al. and Henri, were statements related to student understanding, reasoning, and the development of critical thinking skills and problem solving skills. Similarly, in this study, *cognitive comments* were statements made by the peer assessors indicating strengths and weakness along with reasoned

responses and suggestions for improvement. In this study, the cognitive comments were labeled as *quality comments*. Cognitive comments were identified as surface level and in-depth level, as done by Henri and Hara et al. *Surface level comments* involved making judgments without justification. *In-depth level comments* involved indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions for improvement, and reasoned responses. The sum of surface level and in-depth level cognitive comments made by the peer assessors, determined the quality of comments. Figure 2.1 shows the method of categorizing peer comments in this study.

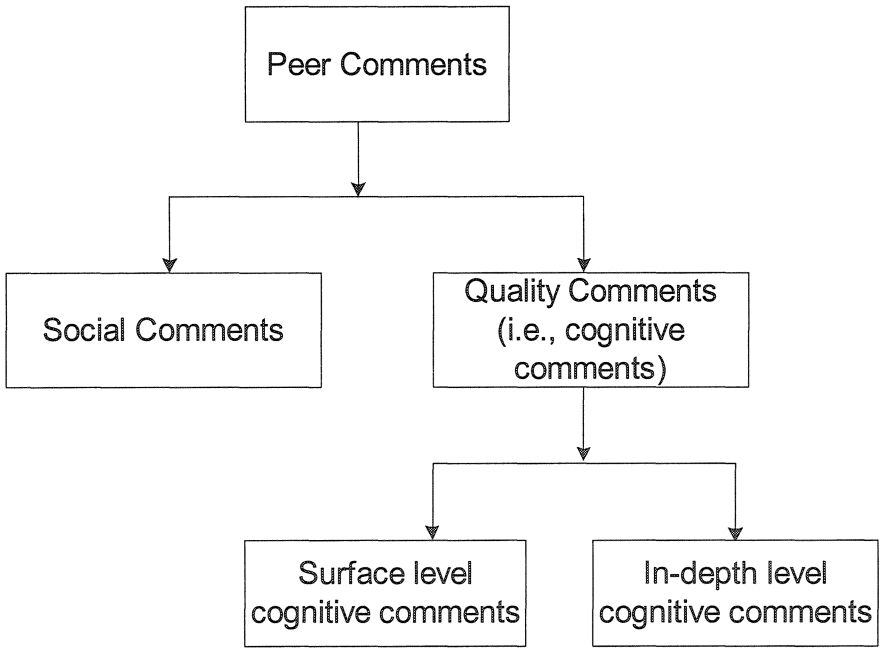


Figure 2.1. Method of categorizing peer comments.

WebCT. Some studies (Downing & Brown, 1997; Davis & Berrow, 1998; Zhao, 1998) on online peer assessment have been conducted using general-purpose Internet technologies, such as, commercial electronic communication applications or ftp. Web-based educational software, such as, MUCH, *NetPeas*, GSS and WebCT, provide tools that support peer assessment within the courseware system. Kwok & Ma (1999) used Group Support System (GSS) to support collaborative learning and peer assessment. Rada (1998) conducted peer assessment exercises with computer science students using a Many Using and Creating Hypermedia system (MUCH). Lin et al. (2001) used *NetPeas* to support web-based peer assessment. This thesis study used Web Course Management Tools (WebCT) to conduct online peer assessment. WebCT is one of the most popular educational web course management system in the world, with licenses sold to 800 institutions (2,600 Universities) in 81 countries serving 300, 400,500 courses, 9 million students (Mann, 2000).

2.3 Summary

Graduate web-based education research methods courses aim to promote critical, evaluative and analytical skills. These skills improve with practice. Peer assessment activities can be helpful in providing practice for developing such critiquing skills. Peer assessment activities often involve

having peer assessors assign a numeric mark and qualitative comments on other students' work. However, within the current literature there are some concerns about the marks and comments provided by the peer assessors during peer assessment. The concern about peer-assigned marks is that peers have a tendency to *over-mark* (i.e., peer assessors assign a mark higher relative to the instructor). The concern about peer comments is that peers are reluctant to provide *critical comments* (i.e., indicate weaknesses) in their assessment of other students' work. This inconsistency in peer over-marking and reluctance in providing critical comments may be due to friendships and social pressure. In this master's thesis research, two variables: (a) anonymity and (b) peer accountability were considered important to overcome the effect of friendship and social pressure on peer-assigned marks and peer comments in online peer assessment. Some studies indicate that anonymity can help to relieve social pressure and allow individuals to behave more freely to provide more critical feedback. While other studies found that anonymity also reduces responsibility thus affecting the quality of the participants work. Incorporating peer accountability variable in peer assessment may help in reducing the dangers associated with anonymity, and improving the quality of response. Although there is no empirical evidence in online peer assessment to determine the effect of peer accountability on the quality of peer comments, manipulating accountability in teacher-evaluation and group interaction have shown positive results.

Therefore, the aim of this study was to examine the effects of anonymity and peer accountability on peer marking, and the criticality and the quality peer comments during online peer assessment. The following section elaborates on the aim of the study and the research questions addressed in this master' thesis.

2.4 Aim of the Study

The aim of this thesis study was to determine the effects of anonymity and peer accountability on peer over-marking, critical peer comments, and the quality of comments provided by the peer assessors during peer assessment in graduate web-based research methods course. Students in a graduate web-based education research methods course were asked to critique two research articles (critique 1 and critique 2). Peer assessment was carried out on the students' first critique (critique 1). *Peer assessment* involved the peer assessors assigning a numeric mark and providing qualitative comments on other students' critiques. *Peers* were other students in the same course. The instructor assessed both students' critiques (critique 1 and critique 2). The instructor's assessment was independent of the peers' assessment. The instructor's assessment only involved assigning a numeric mark.

Two variables: *anonymity* and *peer accountability* were manipulated to determine their effects on peer-assigned marks and peer comments. *Anonymity* was defined as the condition where the peer assessors and the students assessed were either an anonymous group or a named group. In the *anonymous group*, a number replaced the names of the peer assessors and students assessed. In the *named group*, the peer assessors and the students assessed were identified by their names. *Peer accountability* was defined as the condition where the peer assessors were either in a more-accountable group or a less-accountable group. In the *more-accountable group*, the peer assessors were told that timely submissions of their assessment and quality of their comments would contribute to their participation mark for the course. In the *less-accountable* group the peer assessors were only told about timely submissions of their assessments contributing to their participation mark for the course. The peer assessors and the students' assessed were from the same groups.

Based on the above-mentioned conditions of anonymity and peer accountability, following four research questions were examined:

Question 1. Does anonymous online peer assessment affect peer over-marking? It was hypothesized that fewer peer assessors would over-mark in the anonymous group compared to the named group. *Peer over-*

marking was operationalized as the peer assessors assigning a higher mark relative to the instructor, on a student's critique.

Previous studies (Boud & Homes, 1995; Falchikov, 1986, 1995; Kelmar, 1993; Pond et al., 1995; Mowl & Pain, 1995; Rushton et al., 1993; Sluijsmans et al., 2001) suggest that due to loyalty towards friends and social pressure, peers have a tendency to over-mark. This discrepancy in peer marking may affect the validity of peer assessment. Therefore, this study examined whether anonymous peer assessment reduced peer over-marking.

To determine the effect of anonymity on peer over-marking, participants' were randomly assigned to one of the two groups: anonymous or named. Within each group, peer assessors were asked to assess the critiques of other students' in their group. The instructor randomly assigned three student critiques to each peer assessor. Peer assessment involved the peer assessors assigning a numeric mark and providing qualitative comments on the assessed critiques. The instructor also assessed each student critique. The instructor's assessment was independent of the peers' assessment. The instructor's assessment only involved assigning a numeric mark on the student's critique. The average of the three peer-assigned numeric mark on a student's critique was compared with the instructor's mark. The number of peer assessors who over-marked in the anonymous group and the named group were compared.

Question 2. Does anonymous online peer assessment facilitate critical comments in peer feedback? It was hypothesized that the peer assessors in anonymous group would provide more critical comments or indicate more weaknesses in their assessment of other students' critique compared to the peer assessors in the named group.

Literature shows that critical feedback is crucial for learning (Miyake, 1987; Webb, 1982; Zhao, 1998). However, due to social pressure, peer assessors are reluctant to indicate weaknesses or negative comments in their assessment of other students' work (Falchikov, 1995; Fenwick & Parsons, 2000; Topping et al., 2000). This discrepancy in peer comments may affect the learning benefits expected from the process. Therefore, this study examined whether anonymity in peer assessment affected the number of critical comments provided by the peer assessors.

All comments made by the peers were placed in one of the two categories: critical or positive. *Critical comments* were operationalized as the negative comments or weaknesses indicated by the peer assessor on other students' critiques (e.g., "the statement of purpose is vaguely stated"). *Positive comments* were the strengths indicated by the peer assessor on other students' critiques (e.g., "very good points on threats to external validity"). This method of categorizing peer comments as positive and critical (or negative) was based on Falchikov's (1995) peer feedback marking scheme. As explained in question 1, participants were randomly assigned to

an anonymous group or a named group. To determine the effect of anonymity on critical comments in peer feedback, the number of critical comments made by the peer assessors in the anonymous group and the named group were compared.

Question 3. In online peer assessment, how does a more-accountable or less-accountable peer assessor affect the quality of peer comments? It was hypothesized that peer assessors in the more-accountable group would provide a higher number of quality comments compared to the peer assessors in the less-accountable group. *Quality comments* were operationalized as the number of cognitive comments made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement.

Some researchers (Davis, 2000; Topping et al., 2000; Tsai et al., 2001) suggest that incorporating peer accountability condition in the peer assessment process may help in improving quality of peer comments. However, there is no empirical evidence on the effects of peer accountability on online peer assessment to support this claim. Studies on the effect of accountability on the students' response to teacher-evaluation questionnaire (Gordon & Stuecher, 1992) and group interactions (Price, 1987) have shown positive results. These studies further suggest that varying the degree of accountability also affect the complexity of participants' response. Therefore,

this study examined whether the degree (more-accountable or less-accountable) of peer accountability affects the quality of peer comments.

In this study, each comment made by the peer assessors was categorized as either social comment or quality comment (also called cognitive comment) see Figure 2.1. The method of identifying peer comments as social or quality (cognitive) was same as the social and cognitive categories in the online content analysis model developed by Henri (1992) and modified by Hara et al. (2000). In this study the *social comments* were operationalized as general statements made by the peer assessors that were not related to any specific content area. However, the statements were in reference to the context and the content being assessed. For example

> "The assignment I printed had some spelling errors and unfinished sentences. Not sure if it was my printer of the paper. However I know quality is better than quantity."

Cognitive comments, labeled as quality comments, were identified as either surface level or in-depth level comments. *Surface level comments* involved making judgments without justification. For example,

> "Independent and dependent variables were well defined. However there was no mention of other variables."

In-depth level comments involved indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions for improvement, and reasoned responses. For example,

> "The introductory paragraph was very concise and clear and outlined the research design used. My only suggestion here would be to further illustrate the specific type of research design by saying it was true experiment design, comprised of a two-group, post-test only, randomized experiment."

The sum of surface level and in-depth level cognitive comments made by the peer assessors indicated the quality of peer comments. *Comments* (i.e., social or quality) were the written statements provided by the peer assessors on the assessed critique.

To determine the effect of peer accountability on the quality of peer comments, peer assessors were randomly assigned to one of the two groups: more-accountable group or less-accountable group. Each comment (statement) made by the peer assessors was categorized as either social or quality. The number of quality comments (sum of surface level and in-depth level cognitive comments) made by the peer assessors in each group was compared.

Question 4. How does peer assessment in a graduate web-based education research methods course affect student performance in critiquing research articles? It was hypothesized that after participating in the peer assessment exercise for the first critique (critique 1), the students' performance would improve in the second critique (critique 2). *Student performance* was opertaionalized as the student's ability to critique published education research articles.

One of the objectives of graduate web-based education research methods course was to help students develop critiquing skills. Studies (Allen, 1992; Anderson et al., 2001; Howe et al., 2001; Kuhn, 1991; Miyake, 1987; Sluijsmans et al., 2002) indicate that these skills improve with practice and engaging students in critiquing exercises involving peer interaction would provide them with such practice. Literature also shows that the peer assessment process benefits both, the peer assessor and the student assessed (Anderson et al., 2001; Falchikov, 1986; Pond et al., 1995; Searby & Ewers, 1997; Topping et al., 2000; Towler & Broadfoot, 1992). The peer assessors' benefit from analyzing and assessing other students' work. The students assessed benefit from receiving timely and detailed peer feedback. Therefore, this study examined whether the students' performance in critiquing research articles improved after participating in the peer assessment exercise.

To determine students' performance, the instructor-assigned marks on the students' critique 1 and critique 2 were compared. The difference in the two indicated the improvement in student performance in critiquing research articles.

Questionnaire Analysis. In addition to determining improvement in student performance by comparing instructor-assigned marks on students' critique 1 and 2, students' perceptions on learning benefits from the peer assessment process were collected through a questionnaire. It was also hypothesized that the students would perceive the benefit of peer assessment from (a) assessing other students' work, and (b) receiving peer feedback.

Chapter two was a review of the extant literature on issues related to online peer assessment and the aim of the study. Chapter three provides a report on design and methodology of the experiment conducted to determine the effects of anonymity and peer accountability on peer assessment in a graduate web-based education research methods course.

CHAPTER 3

Methods

Chapter three is a report on the participants, materials, and the instruments used in the experiment to determine the effects of anonymity and peer accountability during peer assessment in a graduate web-based education research methods course.

3.1 Participants

Thirty-six graduate students enrolled in a web-based education research methods course, for spring semester 2003, at Memorial University agreed to participate in the experiment. The participants' were 22 females, 14 males. Their ages ranged from 26-55 years. Since these students were enrolled in a web-based course it was assumed that they had the computer skills required for this study. The computer skills required in this study were: ability to use World Wide Web as a research and information tool; use a word processing software package; copy/paste text between the web and a word processor; use electronic mail communications including attachments; and ability to open and read PDF files. The students came from various educational and professional backgrounds including K-12 teachers, school

administrators, those in the post-secondary system, business and industry, as well as other adult learning situations.

Context for the experiment. The research methods course was a compulsory course for all education graduate students at Memorial University. The course was offered online through web course management tool (WebCT). During this study WebCT version 4.0 was being used for all the online courses at this university. WebCT is an educational web course management system to support web-based courses. The content of a WebCT course is provided in HTML pages designed by the instructor or support person (Mann, 2000). WebCT provides variety of tools that can be used by the students and the instructors for content presentation, group interactions and collaboration, monitoring student progress and managing files. Some of the participants in the experiment had used WebCT for other online courses. Prior to this study, however, students in this course had only used WebCT for reading course material posted by the instructor, and contributed their views on the WebCT bulletin board. Online peer assessment was introduced for the first time in this graduate online education research methods course. The instructor and the students had no prior experience of the online peer assessment process introduced in the experiment. The peer assessment process was integrated into the course curriculum.

3.2 Materials

This section provides a detailed description of the experimental platform created in a WebCT system shell, namely: the WebCT tools and the instructional material used in the experiment. Prior to this experiment, the system attributes, WebCT tools and the instructional material were tried and tested with another group of graduate students from a different web-based education course. The WebCT system attributes were tested to understand best practice, workload on the students and instructor, and the timeliness of the assignments. The WebCT tools were tested to check their effectiveness, appropriateness, functionality, and response time. The instructional materials were reviewed for language, conciseness and clarity. A log was maintained to record various technical difficulties experienced during the trials. Sixteen graduate education students participated in the trials. The online peer assessment process in the trials was based on the "post and vote" online peer assessment system developed by Mann (2000).

The following section provides a detailed description of the learning environment created in the WebCT system, the WebCT tools, and the instructional materials used in the experiment. Revisions made based on the trials are also noted in the following sections.

The WebCT system. WebCT is a management system for supporting web-based courses (Mann, 2000). The web-based education research

methods course was conducted on WebCT. The peer assessment process followed in the course was within the WebCT course shell. Figure 3.1 shows the peer assessment process in a WebCT course.

In the experiment, the instructor posted the course information on the WebCT course homepage. The information on the course homepage included links to the course syllabus, course goals, information about the instructor, information about the textbooks, the course requirements, and the information about the course assignments. Appendix A shows a view of the course homepage in WebCT for the education research methods course. For the course assignments, students were asked to critique published research articles. As shown in Figure 3.1, the instructor posted the assignments online. The students viewed the assignments and submitted (uploaded) their critiques into the WebCT. The uploaded students' critiques were available for other students in the course (peers) to view and assess. Peers assessment meant that the students assigned a numeric mark and qualitative comments on the assessed critiques. The instructor compiled the peers' assessments and e-mailed it to individual student. The Instructor also assessed each student's critique, independent of the peers' assessments. The instructor's assessment involved only assigning a numeric mark on the students' critique. The instructor updated and posted the instructor-assigned marks on the student's critique, on WebCT.

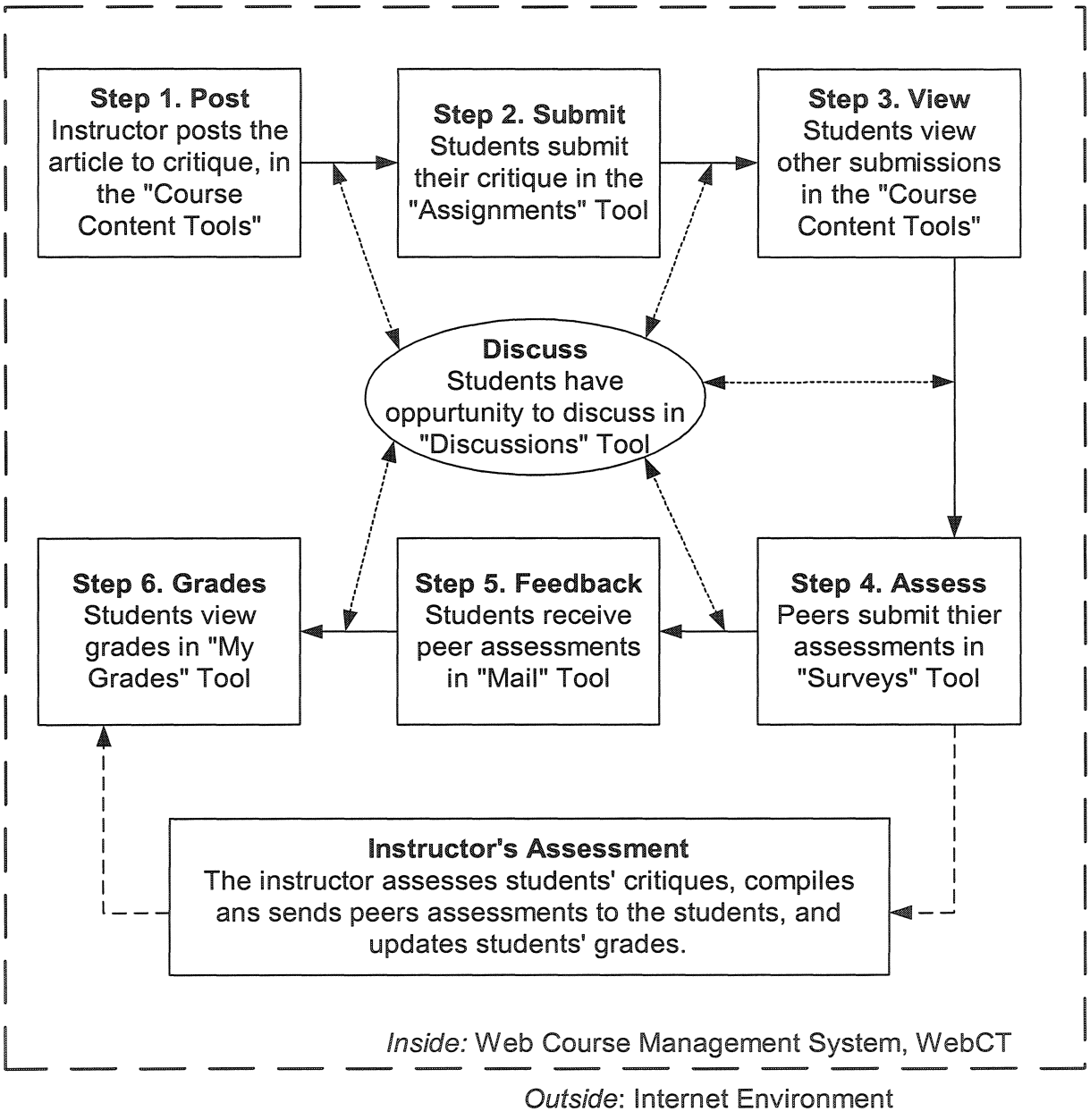


Figure 3.1. The peer assessment process in a WebCT course.

The students viewed their updated grades in the WebCT course shell. On completion of the first assignment, the instructor posted the second assignment. Peer assessment was not carried out on the second assignment.

In developing this model of peer assessment, the initial plan was to conduct peer assessment on both assignments. However, during the trials, students evaluating the tasks found the process of peer assessing two assignments time consuming and suggested that on assignments with the similar content, peer assessment on one assignment was sufficient. Therefore, in the experiment, although the participants were asked to complete two assignments (critique two research articles), peer assessment was carried out only on the first assignment. The details on the assignments are provided in the *Instructional materials* section of this chapter.

WebCT tools. This section provides details on the WebCT tools used for peer assessment activities followed in the experiment. WebCT tools can be categorized in six categories: (i) Pages (ii) Course Content Tools (iii) Communication Tools (iv) Evaluation and Activity Tools (v) Student Tools and (vi) Content Utilities. Each tool category has various options (Goldberg, 1997). In the experiment, the category of tools and options were selected based on the requirements. Table 3.1 provides a summary of WebCT tools used for peer assessment activities in the experiment.

Table 3.1

Summary of WebCT Tools Used for Peer Assessment Activities in the Experiment

Activity and Purpose		WebCT Tools Used		Tools Labeled On The Course Homepage As
		Category	Option	
Post	The instructor posts the assignment (details of the articles) on the course homepage.	Course Content Tools	Content Module	Quantitative Assignments
Submit	Students submit their assignments (critiques) as per instructions.	Evaluation and Activity Tools	Assignments	SubmitYourAssignment
View	Students (peers) view other students' critiques.	Course Content Tools and Pages	Content Module & Single page	ViewPeerAssignment
Assess	Peers assess other students' critiques and submit their assessments.	Evaluation and Activity Tools	Surveys	SubmitYourAssessment
Feedback	The instructor compiles peers assessments for each student and mails it to individual student.	Communication Tools	Mail	Course Mail
Discussions	Students and instructor post messages and discuss course related issues.		Discussions	Discussions
Grades	The instructor assesses each student critique and updates students' grades.	Student Tools	My Grades	My Grades
Questionnaire	Students respond to the questionnaire on the peer assessment process followed in this study.	Evaluation and Activity Tools	Quizzes	Questionnaire

The following section provides details on each activity and the tools used in the experiment.

1. To post the assignment: WebCT's *Course Content Tools* was used to post the details of the research article for the students to critique. The course content tools had five options: (i) Syllabus (ii) Content Module (iii) Glossary (iv) Image Database (v) Index. The *Content Module* option was used to post the article and instructions for the students. The content module organizes the content in a sequence that makes it easier for students to find specific information. The details of the research articles were hyperlinked on the course homepage and labeled as "Quantitative Assignment". Prior to the experiment, the functionality of the content module was tested with the students in the trial group.

2. To submit student critiques: WebCT's *Evaluation and Activity Tools* was used for students to submit their critiques. The evaluation and activity tools had five options: (i) Quizzes and Surveys (ii) Self Test (iii) Assignments (iv) Student Presentations (v) Student Homepages. During the trials, students uploaded their assignment files using the *Student Presentations* and the *Assignments* options. In the student presentation option, files could only be uploaded in the HTML format. In the assignments option, the files could be saved and uploaded in any format (e.g. MS word, Word perfect, HTML etc.).

During the trials, students expressed difficulty in uploading their html'd files in the student presentation option. Further, during the trials it was

observed that the student files uploaded in the student presentation option were immediately available for other students viewing, where as the student files uploaded using the assignment option were only available to the instructor. The design of the experiment required all student files (critiques) to be first submitted to the instructor. The design of this experiment is explained in detail later in this chapter.

Therefore, in view of the (a) difficulty expressed by the students in the trials, to upload the files in the student presentation option, and (b) the design of the experiment, students in the experiment used only the assignments option to submit their critiques. The assignments option was labeled on the WebCT course homepage as "SubmitYourAssignment".

3. To view other students' critiques: WebCT's *Pages Tools* and *Course Content Tools* were selected for peers to view other students' critiques. *Single Page* option of the pages tools and the *Content Module* option of the course content tools, were used. Content module option was setup for viewing other students' critiques. With the single page option a separate web page was created for each student's critique. These web pages were then added to the content module. Features in the content module organize the single web pages in a form of table of contents that makes it easier for students to find the required information. In the experiment, each student's critique was uploaded as a single page. Each critique was assigned a number. The assigned student numbers were indicated in the content module (see

Appendix D). Student peers could view other critiques by clicking on the assigned student number. This option was labeled as "ViewPeerAssignment" on the course homepage. Prior the experiment, the students in the trial group used these tools and options, to view submissions.

4. To submit peer assessments: The *Survey* option of the *Evaluation and Activity Tools* was used for student peers to submit the peer assessments. The features in the survey option allowed student responses to be automatically compiled and tabulated. Student responses were saved without the user's names thus maintaining anonymity. In the experiment, the survey option was used to compile peers assessments on a student's critique. This option was labeled as "SubmitPeerAssessment", on the course homepage. This tool was tested prior to the experiment.

5. To send and receive peer feedback: WebCT's *Communication Tools* were used to communicate peer feedback to the students. The communication tools had six options: (i) Discussions (ii) Mail (iii) Chat (iv) Whiteboard (v) Calendar (vi) Student Tips. The *Mail* option was used to send peers feedback to each student individually. The features in the mail option allowed private messaging to individual students, like any other electronic mail. This option also allows attaching documents with the message. In the experiment, the compiled peer assessments were concatenated and saved in a word processor, as suggested in Mann (2000). This word document was then attached to the mail and sent to each student. Appendix I show a sample

of compiled peer assessment (peer assigned marks and qualitative comments) sent to each student. The mail option was labeled as "Course Mail" on the course homepage. The functionality of this tool was tested prior to the experiment.

6. Online discussions: The *Discussions* option of the *Communications Tools* was also used for student-student interaction and student-instructor interaction online. The messages posted in the discussions area were accessible to all users in the course. The discussion option was labeled as "Discussions" on the course homepage. The functionality of this option was tested prior to the experiment.

7. To update and view student grades: The instructor used WebCT's *Student Tools* to update students' grades. The student tools had three options (i) My Progress (ii) My Grades (iii) Language Selector. Features in *My Grades* option allowed students to see updated grades on their assignments. With this option the students could view only their own grades. In this study, the instructor and the peers independently assessed students' critique. The peer assessments were sent to each student through the course mail. However, the instructor updated the instructor-assigned marks on the students' critiques using my grades option. This was labeled as "My Grades" on the course homepage. To ensure that the instructor could update student grades and the students could access and view their own grades, the option was tested with the students in the trial group.

8. To view and submit questionnaire: In the experiment, the students' response to the questionnaire on the perceived benefits of the peer assessment process followed the experiment were collected. The *Quizzes* option of the *Evaluation and Activity Tools* was used to construct the questionnaire and collect the responses. The features in quizzes option allowed generating multiple-choice questions. During the experiment, students were given a set of statements and asked to select a response from one of five choices. The quizzes option was labeled as "Questionnaire" on the course homepage. This option was tried and tested prior to the experiment.

Instructional materials: During the quality review of the instructional materials (as suggested in Alessi & Trollip, 2001) with the trial group, students were asked to complete two assignments. For the first assignment, the participants wrote a "synopsis" of a research proposal on a topic of their own choice. The synopsis included a brief outline of the components of a research proposal. The second assignment was a "complete research proposal" on the chosen topic. For the complete research proposal the participants submitted a detailed document, elaborating on each component of the proposal. The instructor and the peers assessed the students' first assignment (synopsis) independently. The instructor's assessment and the peers' assessment involved assigning a numeric mark and qualitative comments on the students' assignments. The instructor's assessment and the peers' assessments were

e-mailed to each student. Based on the instructor and peers assessments, students' submitted their second assignment (a complete research proposal). Again the peers and the instructor assessed the students' research proposal. Throughout the duration of the trial, the participants had the option to discuss their assignments, peer assessments and any other course related activities with the instructor and other students.

However, in the experiment some revisions were made to the instructional materials. In the trials, students submitted the assignment on their own chosen topics. As a result the content assessed was different for each peer assessor. Some students (peer assessors) expressed difficulty in assessing different topics due to lack of knowledge on the topics. Topping et al. (2000) also reported that in his study with postgraduate students, peers expressed difficulty in assessing different topics due to lack of knowledge on some topics. Therefore, in the experiment all participants were given the same articles to critique and assess. Details on the first research article and the criteria for critiquing the article is in Appendix B.

In the experiment, the instructor and peers independently assessed the students' critique. The independent initial trials using this method with students had indicated that only the assigned marks and justifying comments of the peer assessors, and the instructor's numeric mark would be required for the experiment. Therefore, qualitative feedback on student critiques was not solicited from the instructor. The main reason for this was that during the

trials when the instructor provided qualitative feedback, the students did not take peer comments seriously. This attitude is consistent with the literature that indicates that students trust the instructor's feedback more than peers' feedback (e.g. Davis, 2000; Falchikov 2001; Pond et al., 1995; Sluijsmans et al., 2001; Topping et al., 2000; Zhao, 1998). A second reason for eliminating instructor's qualitative feedback on the students' critiques was that the initial trials had revealed how difficult it was to determine whether students had followed advice of their instructor or their peers.

Another observation made during the trials was that the peers did not submit their assessments as scheduled. As a result all students could not get timely peer feedback. To ensure timely submissions of the peers' assessments, all participants in the experiment were told that timely submission of the peer assessments would contribute to their participation grade for the course.

This section provided a detailed description of the learning environment created in the WebCT system, the WebCT tools and the instructional material used in the experiment. The following section elaborates on the design of the study.

3.3 Design of the Experiment

In this experiment, a 2 x 2 (anonymous versus named x more-accountable versus less-accountable) factorial design was used (Leedy & Ormrod, 2001). The two independent variables were:

Anonymity (anonymous vs. named). *Anonymity* was defined as the condition where the participants were in the *anonymous group* or the *named group*. In the *anonymous group* a number replaced the names of the peer assessors and the students being assessed. In the *named group* the peer assessors and the students being assessed were identified by their names.

Peer accountability (more-accountable vs. less-accountable). *Peer accountability* was defined as the peer assessors being in the *more-accountable group* or the *less-accountable group*. Peer assessors in the *more-accountable group* were told that timely submission of peer assessment and quality of their feedback would contribute to their participation mark for the course. The peers in the *less-accountable group* were only told about timely submission of their peer assessment as being a part of their participation mark, and not about the quality of their feedback being considered in forming the participation mark.

Students (N = 36) were randomly assigned to one of the four groups (i.e., anonymous, more-accountable group; named, more-accountable group; anonymous, less-accountable group; and named, less-accountable group). Each group (n = 9) received the same research article to critique. Figure 3.2 shows the design of the experiment.

Variable 1 Anonymity	Variable 2 Peer accountability	
	More-accountable	Less-accountable
Anonymous	Group 1	Group 3
Named	Group 2	Group 4

Figure 3.2. The design of the experiment.

The following section describes the instruments and data collection methods used to answer each research question in the experiment.

3.4 Instruments and Data Collection

The 4 research questions in this thesis study were:

1. Does anonymous online peer assessment affect peer over-marking?

2. Does anonymous online peer assessment facilitate critical comments in peer feedback?
3. In online peer assessment, how does a more-accountable or less-accountable peer assessor affect the quality of peer comments?
4. How does peer assessment in a graduate web-based education research methods course affect student performance in critiquing research articles?

The following section presents the instruments developed to answer the four research questions in this thesis experiment.

Question 1. Does anonymous online peer assessment affect peer over-marking? As explained in chapter 2, previous studies suggest that due to loyalty towards friends and social pressure, peers have a tendency to over-mark. This inconsistency in peer marking may affect the validity of peer assessment. Therefore, this research question examined whether anonymous online peer assessment affects peers over-marking. *Peer over-marking* was operationalized as the peer assessors assigning a higher mark relative to the instructor. *Peers* were other students in the same course. Peer assessment involved the peer assessors assigning a numeric mark, on a scale of 1 to 10, and providing qualitative comments on other students' critiques. The

instructor randomly assigned three student critiques to each peer assessor. Three peer assessors assessed each student critique. The average of the marks assigned by the three peer assessors on a student's critique was the peer-assigned mark. The instructor also assessed each student critique. The instructor's assessment involved assigning a numeric mark on a scale of 1 to 10. The instructor's assessment was independent of the peers' assessment.

To determine peer over-marking, the peer-assigned mark (average of the three peers' marks) was compared with the instructor-assigned mark. Figure 3.3 show the instrument developed to compare the peer-assigned mark with the instructor-assigned mark on the students' critique. The number of peer assessors who over-marked in the anonymous group and the named group were compared. It was expected that fewer peer assessors would over-mark in the anonymous group, compared to the named group.

Peers submitted their assessments using the *survey* option of the WebCT tools. WebCT tools were also used to collect data on students' marks. The *management analysis* feature was used to collect data on the peer-assigned marks. The *manage student* option was used to collect the data on the instructor-assigned marks.

Anonymity groups	Peer accountable groups	Peers marks			Peer-assigned mark (P1+P2+P3 /3)	Instructor-assigned mark
		Peer 1 (P1)	Peer 2 (P2)	Peer 3 (P3)		
Anonymous	More-accountable					
	Less-accountable					
Named	More-accountable					
	Less-accountable					

Figure 3.3. The instrument developed to compare the peer-assigned mark with the instructor-assigned mark on the students' critique.

Question 2: Does anonymous online peer assessment facilitate critical comments in peer feedback? As explained in chapter 2, the literature indicates that due to social pressure, peers are reluctant to indicate weaknesses or provide critical (negative) comments in their assessment of other students' work. Since critical feedback is crucial for learning, this discrepancy in peer comments may affect the learning benefits expected from the peer assessment process.

Components	Sub-components	Categories	
		Positive	Critical
Problem	Problem statement & Research question		
	Literature review		
	Hypothesis		
Method	Sampling procedure & subjects characteristics		
	Variables		
	Validity – internal		
	Validity – external		
	Reliability		
	Research design		
Number of positive comments			
Number of critical comments			
Total comments (positive and critical)			

Figure 3.4. The instrument developed to measure the number of positive and critical comments made by the peer assessors.

Therefore, this research question examined whether anonymous online peer assessment enhances the number of critical comments in peer feedback.

To determine the number of critical comments, peer comments were placed in one of the two categories: positive or critical. *Comments* (positive and critical) were the written statements provided by the peer assessors on other students' critiques. Each statement was analyzed as being positive or critical. Figure 3.4 shows the instrument developed to measure the number of positive and critical comments made by the peer assessors. *Critical comments* were defined as weaknesses (or negative comments) indicated by the peer assessor in their assessment of other students' critiques. *Positive comments* were the statements identifying strengths indicated by the peer assessors on other students' critiques. Figure 3.5 shows examples of positive and critical peer comments provided by the peer assessors. The method of identifying peer comments as positive or critical was based on peer feedback marking scheme developed by Falchikov (1995).

In the experiment, students' were asked to critique two research articles. Peer assessment was carried out on students' first critique (critique 1). For critique 1, students were asked to critique only two components of the article, namely: the problem and the method sections. The instructor provided the criteria for critiquing the research articles (see Appendix B). The problem and the method sections were further broken down into 10 sub-components (see Figure 3.4).

Figure 3.5. Examples of the positive and critical comments provided by the peer assessors.

Component	Sub- Component	Positive Comment	Critical Comment
Problem	Problem Statement & Research Question	"The problem statement was clearly defined."	"The statement of purpose is vaguely stated."
	Literature Review	"I felt you gave a good summary of the literature review."	"Your introduction should have focused on the literature review."
	Hypothesis	"You also gave a very thorough overview of the hypothesis."	"You should to discuss whether the hypothesis was in testable form."
Method	Sampling Procedure & Subjects characteristics	"You raised a valid point when discussing the methodology of how participants were selected."	"The information on selection of the participants was provided but you mentioned otherwise."
	Variables	"The choice and explanation of your variables were accurate and well written."	"Control or intervening variables were not identified."
	Validity – Internal	"Issues concerning the internal validity were covered in detail."	"In the 'Internal Validity' section, I differ with you on maturation."
	Validity - External	"Very good points on threats to external validity."	"There was no mention made to external validity."
	Reliability	"You raised some valid points concerning the operations trainer and reliability."	"Analysis on the reliability of these measures would help to strength the critique."
	Research Design	"You correctly identified the research design as posttest only control design."	"Disagree with authors critique of research design i.e. pretest-posttest".
	Overall Presentation	"Overall you presented a very clear and effective critique."	"There are many typo errors and vague sentences."

Since the students were asked to critique only two components (problem and methods) of the article, peer comments only on those sections were considered in the analysis. The number of critical comments made by the peer assessors in the anonymous group and the named group were counted and compared. It was expected that the peer assessors in the anonymous group would provide a higher number of critical comments compared to the peer assessors in the named group.

Peers submitted their assessments using the *survey* option of the WebCT tools. The *management analysis* feature was used to collect data on the peer comments.

Question 3. In online peer assessment, how does a more-accountable or less-accountable peer assessor affect the quality of peer comments? As mentioned in chapter 2, studies (e.g., Gordon & Stuecher, 1992; Price, 1987) on the effect of accountability on student responses showed that participants put in more cognitive effort in their responses when they were asked to justify their comments or when others reviewed their decision. These studies also showed that the degree (more or less) of accountability also affected student responses. Therefore, this research question examined whether varying the degree of peer accountability affected the quality of peer comments. All comments made by the peer assessors were placed in one of the two categories: social or quality (cognitive).

Components	Sub-components	Category of Comments		
		Social	Quality (cognitive)	
			Surface	In-depth
Problem	Problem statement & research question			
	Literature review			
	Hypothesis			
Method	Sampling procedure & subjects characteristics			
	Variables			
	Validity – internal			
	Validity – external			
	Reliability			
	Research design			
Number of social comments				
Total number of quality (cognitive) comments				
Total number of comments (social and quality)				

Figure 3.6. The instrument developed to measure the number of social and quality comments made by the peer assessors.

Figure 3.6 shows the instrument developed to measure the number of social and quality comments made by the peer assessors. Figure 2.1 shows the method of categorizing peer comments. *Comments* (social and quality) were the written statements made by the peer assessors on the assessed critique. *Social comments* were general statements made by the peer assessors that were not related to any specific content area. However, the statements were with reference to the context and the content assessed. *Quality comments* (also called cognitive comments) were statements made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement. Cognitive comments were identified as either surface level or in-depth level cognitive comments. *Surface level cognitive comments* were statements indicating the strengths and weaknesses in students' work without any suggestion, justification and elaboration. *In-depth level cognitive comments* were statements indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions for improvement, and reasoned responses. Figure 3.7 shows examples of the social and quality comments made by the peer assessors. This method of identifying peer comments as social or quality (cognitive) was the same as done by Henri (1992) and Hara et al. (2000).

Category of Comments	Definition	Indicator
Social	Statements general in nature and not related to any specific subject matter.	"The assignment I printed had some spelling errors and unfinished sentences. Not sure if it was my printer of the paper. However I know quality is better than quantity."
		"The student has shown clear understanding of the concepts. She has raised many points from the text. Good job."
Quality (surface level cognitive comments)	Statements indicating the strengths and weaknesses in students' work	"Independent and dependent variables were well defined. However there was no mention of other variables."
		"The purpose of the experiment, type of hypothesis and the variables were well defined. The section on research design and validity needs elaboration."
Quality (in-depth level cognitive comments)	Statements indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions for improvement, and reasoned responses	"There was a concise and elaborative content relating to the breakdown and distinction of independent and dependent variables. However, I think you should have taken it a little further by elaborating the results based on the reliability of the study."
		"The introductory paragraph was very concise and clear and outlined the research design used. My only suggestion here would be to further illustrate the specific type of research design by saying it was true experiment design, comprised of a two-group, post-test only, randomized experiment."

Figure 3.7. Examples of social and quality comments made by the peer assessors.

To determine the effect of varying degree of peer accountability on the quality of peer comments, the peer assessors were randomly assigned to one of the two peer accountable groups: *more-accountable* or *less-accountable.* Peer assessors in the *more-accountable group* were told that timely submission of their assessment and quality of their comments would contribute to their participation grade for the course. Peer assessors in the *less-accountable group* were only told about timely submission of their assessment contributing to their participation grade for the course. Peer assessors in the less-accountable group were not told about the quality of their comments being a part of the participation grade. Quality of comments was the sum of surface level and in-depth level cognitive comments made by the peer assessors. Comments only related to the problem and the method sections of the critique were considered, since the students were asked to critique only those two components of the research article (see Figure 3.6). Each comment (statement) was analyzed for the level of processing, as done by Hara et al. (2000).

It was expected that peer assessors in the more-accountable group would provide more quality comments (sum of surface level and in-depth level cognitive comments) compared to the peer assessors in the less-accountable group.

Peers submitted their assessments using the survey option of the WebCT tools. As mentioned in research question 2, the *management analysis* option was used to collect the data on the peer comments.

Question 4. How does peer assessment in a graduate web-based education research methods course affect student performance in critiquing research articles? As indicated in chapter 2, one of the objectives of graduate web-based education research methods course is to help students develop critiquing abilities. Kuhn (1991) suggests that these skills improve with practice. Engaging students in critiquing exercises involving peer interaction, provides them practice and helps them develop insight into the nature of thinking. Therefore, this research question examined whether students' performance improved after engaging in peer assessment exercise. *Student performance* was defined as the student's ability to critique published education research articles.

To determine improvement in the students' performance in critiquing research articles, the students were asked to critique two published education research articles. For the first article (critique 1) students were asked to critique two components, namely: the problem and the method sections, of the research article (see Figure 3.5 for details on the two components). The instructor provided the components and the criteria to critique (see Appendix B). Peer assessment was carried out for critique 1. Peer assessment included

the peers assigning a numeric mark, on a scale of 1 to 10, and qualitative comments on the students' critique. The instructor also assessed the students' critiques. The instructor's assessment involved only assigning a numeric mark, on a scale of 1 to 10. The instructor's assessment was independent of the peers' assessment. As explained in the *materials section* of this chapter, there were two reasons for the instructor assigning only numeric marks without any qualitative comments on the students critiques: (1) During the trials, both the instructor and the peers provided numeric marks and qualitative comments on students assignments. It was observed that since the instructor provided qualitative feedback, the students did not take peer comments seriously. This attitude was consistent with previous research that indicated that students trust the instructor's feedback more than peer feedback (e.g. Davis, 2000; Falchikov 2001; Pond et al., 1995; Sluijsmans et al., 2001; Topping et al., 2000; Zhao, 1998). (2) During the trials, it was noted that when both the instructor and the peers provided qualitative feedback, it was difficult to determine whether the student benefited from the instructor's feedback or peer feedback.

After the students received the instructor-assigned marks and the peers assessments on critique 1, they were asked to critique the second research article (critique 2). For the second research article students' were asked to critique four components of the research article, namely: problem, method, results, and conclusion sections of the research article. The

instructor provided the components and the criteria for critiquing the second research article, as done for critique 1 (see Appendix B). Critique 2 was assessed on a scale of 1 to 20. Only the instructor assessed students' critique 2. The reason for only the instructor assessing critique 2, as explained in the *materials section* of this chapter, was that during the trials it was noted that the students found the process of assessing two assignments with similar content time consuming. Since, in the experiment both critiques were on similar content (quantitative research methods), peer assessment on the first critique (critique 1) was considered sufficient.

To determine the difference in students' performance, the instructor-assigned marks on students' critique 1 and 2 were compared. Figure 3.8 shows the instrument developed to calculate the differences in the instructor-assigned marks on the students' critiques. Since the students' critique 1 was on a scale of 1 to 10, the instructor-assigned mark on critique 2 was also equated to a 10-point scale. It was expected that students' performance on critique 2 would be better than critique 1. The data on the instructor-assigned marks was collected using the *manage student* option of the *manage course* feature in WebCT.

Anonymity groups	Peer accountable groups	Instructor-assigned marks		Difference (Critique 2 (–) Critique 1)
		Critique (1)	Critique (2)	
Anonymous	More-accountable			
	Less-accountable			
Named	More-accountable			
	Less-accountable			

Figure 3.8. The instrument developed to calculate the differences in the instructor-assigned marks on the students' critiques.

Questionnaire Analysis. In addition to determining the improvement in the student performance based on the instructor-assigned marks on the students' critiques, the students were also asked to respond to a questionnaire. The questionnaire was constructed to determine whether the students' perceived the peer assessment exercise beneficial and easy to follow. The questions (statements) in the questionnaire were aimed to determine whether the students' (a) learned from assessing other students' work (b) learned from receiving peer feedback, and (c) found the peer

assessment procedure easy to follow. The questionnaire on students' perception is placed at Appendix K. Students were asked to respond to the statements by selecting one of the five categories graded 'strongly agree' to 'strongly disagree' with 'undecided' as a middle category. The questionnaire was constructed using the quiz tool in WebCT. The data on students' responses were collected from the manage student option in WebCT.

3.5 Procedure

The peer assessment process followed in this experiment was adopted from previous studies (Falchikov, 1995,1996; Mann, 2000; Slujismans et al., 2002; Topping et al., 2000). WebCT, web course management system was used for the course and the peer assessment process. The WebCT course shell was password protected. The course homepage interface for the student and the instructor was the same (see Appendix A). However, the instructor also had the designer option. The online activities involved in this study included, the instructor posting the information about the research articles to be critiqued, the students submitting their critiques, peers viewing student submissions and submitting their assessments, and the instructor updating student marks and forwarding peers assessments to each student. All activities were within the WebCT environment. Throughout the experiment, the participants had the option to discuss their assignments, peer assessments and any other course related

activities with the instructor and other students. Figure 3.1 shows the online peer assessment process followed in this experiment.

The schedule of activities and the timelines followed in this experiment were negotiated and developed by the researcher and the course instructor. The entire process lasted six weeks. The activities and procedure followed during the six weeks were:

Week 1: During the first week of the semester, the course related reading material was posted on the WebCT course homepage (see Appendix A). The details of the first assignment, a research article to be critiqued by the students, were also hyperlinked as "Quantitative Assignments" on the WebCT course homepage. All students were asked to critique the same article. Since this was the first assignment for the course, a simple article was chosen for students to critique. The article focused on quantitative research methods. Students were asked to critique only two components of the article, namely, the problem and the method sections. The instructor provided the criteria for critiquing the research article. Students were told their critique should be brief (approximately 6 double spaced pages) and should focus on critique not the description. Details on the first research article and the criteria for critiquing the article is in Appendix B. Since online peer assessment was introduced for first time in this course, the instructor posted the purpose of peer assessment on the WebCT discussion board for all students to view. The instructor's message posted on the discussion board regarding the purpose of peer

assessment in the course is in Appendix C. The discussion board was accessible to the students throughout the experiment to enable engage instructor-student and student-student interaction in asynchronous online discussions.

Prior to the study, approval from the interdisciplinary committee on ethics in human research (see Appendix P) and students' consent for data collection and analysis was obtained. Thirty-six students consented to participate. One student did not agree to participate. Therefore, that student's data was excluded from the experiment. Students were given two weeks to submit their first assignment (critique 1).

Week 3: In the third week, the students submitted their first critique online using the *assignment tool* in WebCT. After submitting the critique, if the students wanted to make certain changes and re-submit their critique, they had the option to retrieve the submitted critique and reload it before the due date. Students were informed that late submissions would not be accepted. To upload the submission in the designated area, students were told to save their assignment file on the computer, log on to WebCT, click on the "SubmitYourAssignment" link on the course menu (left navigation bar), and follow the directions on the WebCT screen. All students indicated their name and student number on the submitted critique. The instructor randomly assigned all 36 students to one of the four groups, $n = 9$, see Figure 3.2 for the design of the experiment.

Soon after all the students' critiques were uploaded into WebCT, the researcher revealed the critiques for students viewing. The *content module* and *single page* tools of WebCT were used to upload the students' critiques for viewing. The students could view the list of uploaded critiques by selecting "ViewPeerAssignment" option in the left navigation bar of the course homepage. A view of the "ViewPeerAssignment" page identifying each uploaded assignment with an assigned student number is in Appendix D. On clicking the hyperlinked student number, students were able to view the complete assignment. Depending on the groups, students were able to view the critiques in the anonymous group and the named group. In the anonymous group, a number identified the students' critique. In the named group, the students' critiques indicated the student's name and a number. A sample of uploaded critiques as seen by the students in the anonymous group and the named group placed at Appendix E and Appendix F.

After all students' critiques were uploaded for other students to view, the instructor randomly assigned three student critiques to each peer for assessment. A computer software program, Microsoft Excel, was used to randomly assign student critiques to peer assessors. The formula used for random assignment ensured that (i) the students' critiques and their peer assessors were from the same group, (ii) each student's critique was assessed by three different peer assessors, (iii) each peer assessor assessed three different critiques, (iv) peer assessors do not assess their own critiques.

Using the "Course Mail' in WebCT, the instructor informed each peer assessor of the critiques they were to assess, e.g., please assess student (1), student (5), and student (7). Apart from the student number, peer assessors in each group received the following information:

> *Group 1,* the peer assessors in the anonymous, more-accountable group: Please submit your peer assessment by 08 June. Provide useful feedback on the students' critiques. Your timely submission and the quality of your feedback on the student assignments will be a part of your participation grade for this course. Your assessment will be anonymous.
>
> *Group 2,* the peer assessors in the named, more-accountable group: Please submit your peer assessment by 08 June. Provide useful feedback on the students' assignments. Your timely submission and the quality of your feedback on the student assignments will be a part of your participation grade for this course. Your assessment will be sent to the students, with your name, whose assignments you assessed. You will also receive the feedback on your assignments from your peers.
>
> *Group 3,* the peer assessors in the anonymous, less-accountable group: Please submit your peer assessment by 08 June. Timely submission of

your assessment will be a part of your participation grade for the course. Your assessment will be anonymous.

Group 4, the peer assessors in the named, less-accountable group: Please submit your peer assessment by 08 June. Timely submission of your assessment will be a part of your participation grade for the course. Your assessment will be sent to the students, with your name, whose assignments you assessed. You will also receive the feedback on your assignments from your peers.

Peer assessment included assigning numeric marks (on a scale of 10) and providing qualitative comments. Although the instructor did not provide specific criteria to the assessors for assessment, peer assessors were told to focus only on the two components, the problem and the method sections, since only these two sections were critiqued. The instructor's message on the discussion board about criteria for assessing the students' critiques is in Appendix G. Peer assessors were given one week to assess other students' critiques and submit their assessments (a numeric mark and qualitative comments on the assessed critiques).

Week 4. In the fourth week, the peers submitted their assessments in WebCT by selecting "SubmitPeerAssessment" option from the left navigation bar on the course homepage. WebCT's survey tool was used to submit peer

assessments. To submit the peer assessments, students were told to (1) click on the link "SubmitPeerAssessment" on the course menu and then on "Quantitative-Assignment 1, (2) provide comments and a grade (/10) under the space provided for each assignment assessed, (3) click "Save Answer" after each assessment to save your response. Your response will not be saved till you click "Save Answer", (4) click "Finish" to submit your assessment. Appendix H shows a view of the "SubmitPeerAssessment" screen to submit peer assessments.

Week 5: Once all peer assessments were submitted, using the compile feature in WebCT, the researcher compiled the peer assessments and e-mailed them to the students. Each student received three peer assessments. Appendix I show a sample of compiled peer assessment (peer assigned marks and qualitative comments) sent to each student. The instructor independently assessed the student critiques. The instructor's assessment involved assigning numeric marks (on a scale of 1 to 10) on the student's critiques. The instructor posted the instructor-assigned mark on each student's critique using the *manage student* option of the *manage course* feature of WebCT. The students were able to view the instructor-assigned mark, in *my grades* feature of the student tools in WebCT, labeled as "My Grades". This feature only allowed the students to view their own marks.

Week 6: Once the students received peer feedback on the first critique (critique 1), they were asked to critique a second published education research article (critique 2). This article also focused on quantitative research methods. However, this article was more difficult compared to the first article. For critique 2, the students had to critique four components, namely: the problem, the method, the results and the conclusions sections of the article. Appendix J shows the details on the criteria for critiquing the second research article. Only the instructor assessed the second critique. Instructor's assessment involved assigning a numeric mark (on a scale of 1 to 20) on the student's critiques. After submitting the second critique, the students were asked respond to a questionnaire in WebCT *survey* tool regarding their perceptions on the (1) learning benefits from assessing student assignments and peer feedback and (2) peer assessment procedure followed in this study. Copy of the questionnaire on students' perception is in Appendix K.

Chapter three was a report on the participants, materials, the instruments, the design of the experiment, and the procedure followed in the experiment to determine the effect of anonymity and peer accountability on online peer assessment. Chapter four will provide detailed results of the analysis of the data collected using the methods described in this chapter.

CHAPTER 4

Results

Chapter four is a report on the analysis on the data collected to determine the effects of anonymity and peer accountability on peer over-marking, critical peer comments, and the quality of peer comments during online peer assessment. The effects of the independent variables on the dependent measures are discussed in the context of the four research questions.

4.1 Peer Over-Marking

Does anonymous online peer assessment affect peer over-marking? The results from a 2 x 2 chi-square test indicated that the relationship between anonymity and the number of peer assessors over-marking was statistically significant, χ^2 (1, N = 36) = 4.050, p = .044, with a medium effect size, W = 0.335. As hypothesized, fewer peer assessors over-marked in the anonymous group (7 of 18, i.e., 39%) compared to the peer assessors in the named group (13 of 18, i.e., 72%). Table 4.1 shows the number of peers who over-marked, under-marked and assigned an identical-mark relative to the instructor. *Peer over-marking* was defined as the peer assessors assigning higher marks relative to the instructor-assigned mark, on a student's critique.

Three peers assessed each student's critique. The *peer-assigned marks* were the average of the numeric mark assigned by the three peer assessors.

Table 4.1

Number of Peers Who Over-Marked, Under-Marked, and Assigned an Identical-Mark Relative to the Instructor

Anonymity groups	Peer accountable groups	Number of peers		
		Over-Marked	Under-Marked	Identical-Marked
Anonymous	More-accountable	4	3	2
	Less-accountable	3	5	1
		7* (39%)	8 (44%)	3 (17%)
Named	More-accountable	5	2	2
	Less-accountable	8	1	0
		13 (72%)	3 (17%)	2 (11%)
Total		20 (56%)	11 (30%)	5 (14%)

Note. Number of peers in each of the four groups was 9. *$p < .05$.

The data on the peer-assigned marks were further analyzed to determine the effect of anonymity on peer under-marking (average of the three peer-assigned marks lower than the instructor-assigned mark) and peer identical-marking (average of the three peer-assigned marks equal to the instructor-assigned mark). The results indicated that the relationship between anonymity and the number of peers who under-marked was not statistically significant, χ^2 (1, N = 36) = 3.273, p = .070, with a medium effect size, W = 0.301. Further, the relationship between anonymity and the number of peers who assigned an identical-mark relative to the instructor, was also not statistically significant, χ^2 (1, N = 36) = 0.232, p = .630, with a small effect size, W = 0.080.

The results of the effect of peer accountability on the number of peers over-marking, under-marking, and identical-marking were also tested. The results indicated that the relationship between peer accountability and the number of peers who over-marked was not statistically significant, χ^2 (1, N = 36) = 0.450, p = .502, with a small effect size, W = 0.112. The relationship between peer accountability and the number of peers who under-marked was also not statistically significant, χ^2 (1, N = 36) = 0.131, p = .717, with a very small effect size, W = 0.060. Further, the relationship between peer accountability and the number of peers who assigned an identical-mark relative to the instructor's mark was not statistically significant, χ^2 (1, N = 36) =

2.090, p = .148, with a small effect size, W = 0.240. Details of peers and instructor-assigned mark on students' first critique are in Appendix L.

To summarize the results of the effect of anonymity (anonymous and named) on peer marking: (a) Significantly fewer peer assessors over-marked in the anonymous group compared to the named group. This finding supports the hypothesis. (b) There was no statistically significant difference in the number of peer assessors who under-marked or assigned an identical-mark.

To summarize the results of the effect of peer accountability (more-accountable and less-accountable) on peer marking: There was no statistically significant difference in the number of peer assessors who (a) over-marked, (b) under-marked, and (c) assigned an identical-mark relative to the instructor-assigned mark.

4.2 Critical Comments in Peer Feedback

Does anonymous online peer assessment facilitate critical comments in peer feedback? The results from a 2 x 2 chi-square test indicated that the relationship between anonymity and the number of critical comments made by the peer assessors was statistically significant, χ^2 (1, N = 767) = 32.368, p = .000, with a small effect size, W = 0.205. As hypothesized, peer assessors in the anonymous group made significantly more critical comments, n = 185, compared to the peer assessors in the named group, n = 157. The peer assessors in the named group made significantly more positive comments, n

= 282, compared to the peer assessors in the anonymous group, n = 143. Table 4.2 shows the number of critical and positive comments made by the peer assessors.

Critical comments were the negative statements, also called weaknesses, indicated by the peer assessor on another student's work (e.g., "the statement of purpose is vaguely stated"). *Positive comments* were the strengths indicated by the peer assessor on another student's work (e.g., "very good points on threats to external validity"). All comments made by the peer assessors were placed in a positive or critical category. The procedure of categorizing the comments as positive and critical (negative) was the same as followed by Falchikov (1995).

Total comments were the sum of positive and critical comments made by the peer assessors. The relationship between anonymity and the total comments made by the peer assessors was statistically significant, χ^2 (1, N = 767) = 16.060, p = .000, with a small effect size, W = 0.145. Peer assessors in the named group made a significantly higher number of comments, n = 439, than the peer assessors in the anonymous group, n = 328. The relationship between peer accountability and the number of critical comments made by the peer assessors was not statistically significant, χ^2 (1, N = 767) = 0.489, p = .485, with a negligible effect size, W = 0.025. The number of critical comments made by the peer assessors in the more-accountable group and less-accountable group were195 and 147, respectively.

Table 4.2

Number of Critical and Positive Comments Made by the Peers Assessors

Anonymity groups	Peer accountable groups	Total number of comments (*N*)								
		Critical[a]			Positive[b]			Total comments (critical and positive)		
		M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*N*
Anonymous	More-accountable	10.00	5.10	90	8.33	6.40	75	9.17	5.68	165
	Less-Accountable	10.56	10.63	95	7.56	10.26	68	9.06	10.25	163
		10.28	8.09	185 (56%)	7.94	8.31	143 (44%)	9.11	8.17	328
Named	More-accountable	11.67	9.30	105	19.78	9.05	178	15.72	9.83	283
	Less-accountable	5.78	5.67	52	11.56	7.60	104	8.67	7.15	156
		8.72	8.06	157 36%	15.67	9.15	282 64%	12.19	9.20	439
Total		9.50	8.00	342 (45%)	11.81	9.46	425 (55%)	10.65	8.78	767

Note. [a] A critical comment was a weakness indicated by a peer assessor. [b]A positive comment was a strength indicated by a peer assessor. *n* indicates the number of comments in each category.

Similarly, the relationship between peer accountability and the number of positive comments made by the peer assessors was not significant. The number of positive comments made by the peer assessors in the more-accountable group and less-accountable group were 253 and 172, respectively. However, the relationship between peer accountability and the total number of comments (sum of critical and positive comments) made by the peer assessors was significant, χ^2 (1, N = 767) = 21.700, p = .000, with a small effect size, W = 0.168.The peer assessors in the more-accountable group made a significantly higher number of comments, n = 448, than the peer assessors in the less-accountable group, n = 319. The details of critical and positive comments made by the peer assessors are in Appendix M.

To summarize the results of the effect of anonymity (anonymous and named) on the number of critical and positive comments made by the peer assessors: (a) The peer assessors in the anonymous group made significantly more critical comments compared to the peer assessors in the named group. This finding supports the hypothesis. (b) The peer assessors in the named group made significantly more positive comments compared to the peer assessors in the anonymous group. (c) The peer assessors in the named group made significantly more comments (total number of critical and positive) comments compared to the peer assessors in the anonymous group.

To summarize the results of the effect of peer accountability (more-accountable and less-accountable) on the number of critical and positive

comments made by the peer assessors: (a) There was no statistically significant difference in the number of critical comments made by the peer assessors. (b) There was no statistically significant difference in the number of positive comments made by the peer assessors. (c) The peer assessors in the more-accountable group made significantly more comments (total number of critical and positive) compared to the peer assessors in the less-accountable group.

4.3 Quality Comments Made by the Peer Assessors

In online peer assessment, how does a more-accountable or less-accountable peer assessor affect the quality of peer comments? The results from a 2 x 2 chi-square test indicated that the relationship between peer accountability and the quality of peer comments was significant, χ^2 (1, N = 856) = 32.566, p = .000, with a small effect size, W = 0.195. As hypothesized, the peer assessors in the more-accountable group provided a significantly higher number of quality comments, n = 389, compared to the peer assessors in the less-accountable group, n = 236. The peer assessors in the less-accountable group provided significantly higher number social comments, n = 139, compared to the peer assessors in the more-accountable group, n = 95. Table 4.3 shows the number of quality comments and social comments made by the peer assessors.

Table 4.3

Number of Quality Comments and Social Comments Made by the Peer Assessors

Peer accountable groups	Anonymity groups	Total number of comments (*N*)								
		Quality comments[a]			Social comments[b]			Total comments (quality + social)		
		M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	N
More-accountable	Anonymous	14.67	4.44	132	5.33	1.94	48	10.00	5.84	180
	Named	28.56	18.27	257	5.22	2.39	47	16.89	17.43	304
		21.61	14.75	389 (80%)	5.28	2.10	95 (20%)	13.44	13.28	484
Less-accountable	Anonymous	11.44	14.95	103	7.33	3.00	66	9.39	10.67	169
	Named	14.78	9.43	133	7.78	4.41	70	11.28	8.00	203
		13.11	12.25	236 (63%)	7.56	3.67	139 (37%)	10.33	9.34	372
Total		17.36	14.04	625 (73%)	6.42	3.17	231 (27%)	23.78	6.85	856

Note. [a]Quality comments were cognitive statements made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement. [b]Social comments were general statements not related to a specific content of subject matter. *n* indicates the number of comments in each category.

Each comment made by the peer assessors was placed in one of the two categories, namely: quality or social, see Figure 2.1. *Quality comments* were cognitive statements made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement. Quality comments were identified as either surface level or in-depth level cognitive statements. *Social comments* were general statements made by the peer assessors that were not related to any specific content area. The method of identifying the peer comments as quality and social was same as done by Henri (1992) and Hara et al. (2000).

Total comments were the sum of quality comments and social comments made by the peer assessor. The relationship between peer accountability and total comments made by the peer assessors was statistically significant, χ^2 (1, N = 856) = 14.650, p = .000, with a small effect size, W = 0.130. The peer assessors in the more-accountable group made significantly higher number of comments, n = 484, compared to the peer assessors in the less-accountable group, n = 372.

The relationship between anonymity and quality of peer comments was statistically significant, χ^2 (1, N = 856) = 9.478, p = .002, with a small effect size, W = 0.105. Peer assessors in the named group made a significantly higher number of quality comments, n = 390, compared to the peer assessors in the anonymous group, n = 235. However, the number of social comments made by the peer assessors in the named group and the

anonymous group were 117 and 114, respectively (see Table 4.3). The relationship between anonymity and total number of comments (sum of quality comments and social comments) made by the peer assessors was statistically significant, χ^2 (1, N = 856) = 29.160, p = .000, with a small effect size, W = 0.184. The peer assessors in the named group made significantly higher number of comments, n = 507, compared to the peer assessors in the anonymous group, n = 349.

As mentioned earlier, quality comments were cognitive statements that were identified as either surface level or in-depth level cognitive comments. Table 4.4 shows the number of surface level and in-depth level cognitive comments made by the peer assessors. *Surface level cognitive comments* were the statements made by the peer assessors indicating the strengths and weaknesses in students' work without any suggestion, justification and elaboration. *In-depth level cognitive comments* were the statements made by the peer assessors indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions for improvement, and reasoned responses.

Table 4.4

Number of Surface Level and In-depth Level Cognitive Comments Made by the Peer Assessors

Peer accountable groups	Anonymity groups	Total number of comments (*N*)								
		Surface level[a]			In-depth level[b]			Total quality comments (surface + in-depth)		
		M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*N*
More-accountable	Anonymous	9.78	3.27	88	4.89	4.17	44	14.67	4.44	132
	Named	13.11	7.75	118	15.44	12.86	139	28.56	18.27	257
		11.45	6.02	206 (53%)	10.17	10.74	183 (47%)	10.81	8.61	389
Less-accountable	Anonymous	5.89	5.11	53	5.56	11.40	50	11.44	14.95	103
	Named	8.67	5.29	78	6.11	6.72	55	14.78	9.43	133
		7.28	5.24	131 (56%)	5.84	9.08	105 (44%)	6.56	7.35	236
Total		9.37	5.95	337 (54%)	8.01	10.05	288 (46%)	17.36	14.04	625

Note. [a]Surface level comments were statements indicating the strengths and weaknesses in students' work without any suggestion, justification and elaboration. [b]In-depth level comments were statements indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions, and reasoned responses. *n* indicates comments in each category.

Between the peer accountable groups (more-accountable and the less-accountable), the difference in the number of surface level and in-depth level cognitive comments made by the peer assessors was not statistically significant, χ^2 (1, N = 625) = 0.572, p = .450, with a negligible effect size, W = 0.030. However, between the anonymity groups (anonymous and named), the difference in the number of surface level and in-depth level cognitive comments made by the peer assessors was statistically significant, χ^2 (1, N = 625) = 7.967, p = .005, with a small effect size, W = 0.112. Peer assessors in the named group made a significantly higher number of in-depth level cognitive comments, n = 194, compared to the peer assessors in the anonymous group, n = 94. Peer assessors in the named group also made more surface level cognitive comments, n = 196, than the peer assessors in the anonymous group, n = 141, see Table 4.4. Details of social and quality comments made by the peer assessors are in Appendix N.

To summarize the results of the effect of peer accountability (more-accountable and less-accountable) on social and quality comments made by the peer assessors: (a) The peer assessors in the more-accountable group provided significantly more quality comments compared to the peer assessors in the less-accountable group. This finding supports the hypothesis. (b) The peer assessors in the less-accountable group provided significantly more social comments compared to the peer assessors in the more-accountable group. (c) The peer assessors in the more-accountable group provided

significantly more comments (total number of quality comments and social comments) compared to the peer assessors in the less-accountable group. (d) There was no statistically significant difference in the number of surface level and in-depth level cognitive comments made by the peer assessors.

To summarize the results of the effect of anonymity (anonymous and named) on social and quality comments made by the peer assessors: (a) The peer assessors in the named group provided significantly more quality comments compared to the peer assessors in the anonymous group. (b) The peer assessors in the named group provided significantly more social comments compared to the peer assessors in the anonymous group. (c) The peer assessors in the named group provided significantly more comments (total number of quality comments and social comments) compared to the peer assessors in the anonymous group. (d) The peer assessors in the named group provided significantly more in-depth level cognitive comments compared to the peer assessors in the anonymous group. (e) The peer assessors in the named group provided significantly more surface level cognitive comments compared to the peer assessors in the anonymous group.

4.4 Student Performance in Critiquing Research Articles

How does peer assessment in a graduate web-based education research methods course affect student performance in critiquing research articles? A repeated measures ANOVA was conducted to determine the effect of the interaction of anonymity and peer accountability on the difference in students' performance from critique 1 to critique 2. The results indicated that there was no statistically significant interaction of anonymity and peer accountability on the difference in students' performance: F (1) = 0.000, p = 1.00, η^2 = 0.00. This finding does not support the hypothesis. Table 4.5 shows the summary of repeated measures ANOVA with within-subjects effects of anonymity and peer accountability.

Student performance was defined as the student's ability to critique published education research articles. To determine the students' performance, the instructor-assigned marks on the students' critique 1 and critique 2 were compared.

Further, there was no statistically significant, F (1) = 0.336, p = .566, η^2 = .010, effect of peer accountability on the difference in students' performance from critique 1 to critique 2. However, there was a statistically significant, F (1) = 4.360, p =. 045, η^2 = .120, effect of anonymity on the difference in students' performance from critique 1 to critique 2.

Table 4.5

Summary of Repeated Measures ANOVA with Within-Subjects Effects of Anonymity and Peer Accountability

Variables	*SS*	*df*	*MS*	*F*	*p*	η^2
Critique (Critique 1 - Critique 2)	3.125	1	3.125	0.121	0.730	0.004
Critique x Anonymity	1.125	1	1.125	4.360	0.045*	0.120
Critique x Peer accountability	8.861	1	8.861	0.336	0.566	0.010
Critique x Anonymity x Peer accountability	0.000	1	0.000	0.000	1.00	0.000
Error (Critique)	8.257	32	0.258			

*$p < .05$

In the named group, the instructor-assigned marks improved from critique 1 (M = 7.84, SD = 0.61) to critique 2 (M = 8.13, SD = 0.65). However, in the anonymous group, the instructor-assigned marks decreased from critique 1 (M = 8.17, SD = 0.66) to critique 2 (M = 7.96, SD = 0.47). Table 4.6 shows the means of the instructor-assigned marks on the students' critique 1 and critique 2.

Table 4.6

Means' of the Instructor-Assigned Marks on the Students' Critique 1 and Critique 2

Anonymity groups	Peer accountable groups	Means of the instructor-assigned marks on the students' critiques			
		Critiques 1		Critique 2	
		M	*SD*	*M*	*SD*
Anonymous	More-accountable	8.17	0.66	8.03	0.51
	Less-accountable	8.17	0.71	7.89	0.47
		8.17	0.66	7.96	0.47
Named	More-accountable	8.06	0.68	8.42	0.74
	Less-accountable	7.61	0.49	7.83	0.41
		7.84	0.61	8.13	0.65
Overall class performance		8.00	0.65	8.04	0.57

Note. Number of students in each of the four groups was 9.

Figure 4.1 show means' of the students' performance from critique 1 to critique 2 in the anonymity groups.

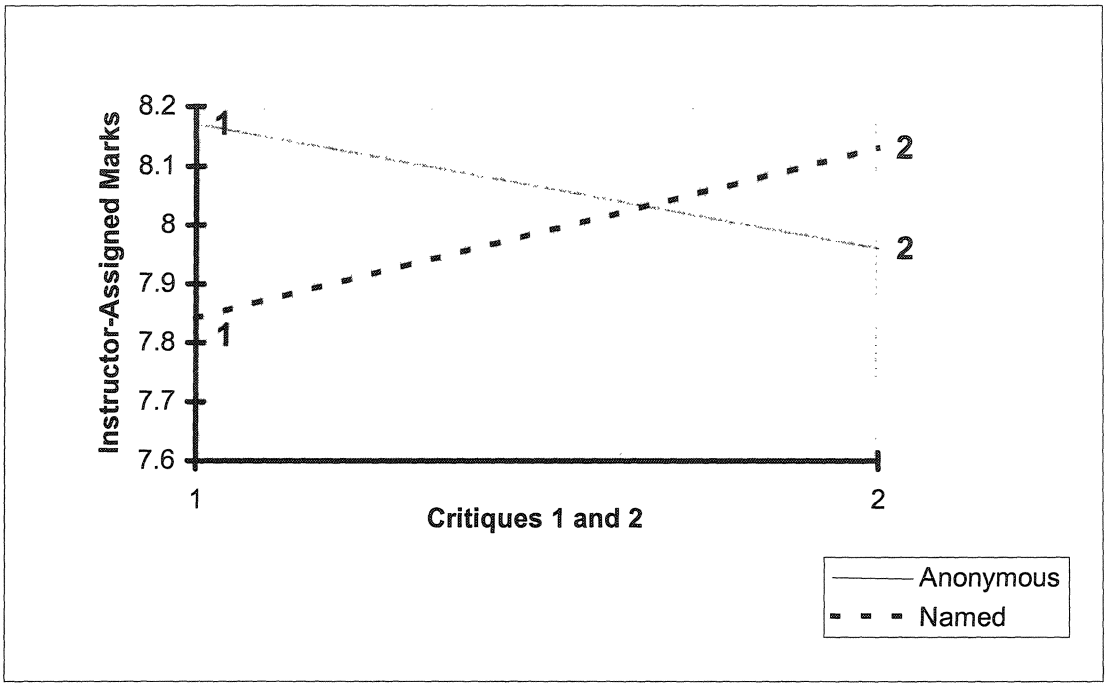


Figure 4.1. Means' of the students' performance from critique 1 to critique 2 in the anonymity groups.

Finally, there was no statistically significant difference, F (1) = 0.121, p = .730, η^2 = .004, in the instructor-assigned students' marks from critique 1 (M = 8.00, SD = 0.65) to critique 2 (M = 8.04, SD = 0.57), see Table 4.6. Details of instructor-assigned marks on students' critique1 and 2 are in Appendix O.

The results of the effect of anonymity and peer accountability on the students' performance can be summarized as: (a) There was significant effect of the interaction of anonymity and peer accountability on the students' performance. (b) There was no significant effect of the peer accountability variable on the difference in students' performance from critique 1 to critique 2. (c) There was a significant effect of the anonymity (anonymous and

named) variable on the difference in students' performance from critique 1 to critique 2. The students' performance improved in the named group. However, it decreased in the anonymous group. (d) Overall there was no significant improvement in the students' performance from critique 1 to critique 2. This finding was contrary to the hypothesis.

Questionnaire analysis. In addition to determining the improvement in the student performance, based on the instructor-assigned marks on the students' critiques, the students were also asked to respond to a questionnaire. The questionnaire was constructed to determine whether the students' (a) learned from assessing other students' work (b) learned from receiving peer feedback, and (c) found the peer assessment procedure easy to follow. The data indicated that the students learned more from assessing and viewing other students' critiques than from the peer feedback. The students' response to the questionnaire also showed that they found the peer assessment process in this study, easy to follow and that they would recommend this process in other courses. Of the 36 students in the study, 35 responded to the questionnaire. Table 4.7 shows the students' response on learning benefits derived from peer assessment process followed in this study.

Table 4.7

The Students' Response on Learning Benefits Derived From the Peer Assessment Process Followed in This Study

Statements	Agree (%)	Don't Know (%)	Disagree (%)
I learned about the critiquing research articles by assessing other students' critiques (assignment).	83	3	14
I learned about critiquing research articles from viewing other students' critiques (assignment).	91	0	9
I learned about critiquing research articles from completing my own critique (assignment).	94	3	3
I learned about critiquing research articles from peer feedback.	66	3	31
I found the peer comments helpful.	83	6	11

Note. Total number of students who responded to the questionnaire was 35.

A detailed analysis of students' response to the questionnaire, on the learning benefits derived from the peer assessment process, is as follows.

Assessing other students' critiques (assignments): Of the 35 students, 83% (29 of 35), reported that they learned about critiquing articles by assessing other students' critiques. In the anonymity groups, approximately the same number of students in the named group, 52% (15 of 29), and the anonymous group, 48% (14 of 29), reported that they learned about critiquing articles by assessing other students' critiques. Similarly, in the peer accountable groups, approximately the same number of students in the more-accountable group, 48% (14 of 29), and the less-accountable group, 52% (15 of 29), reported that they learned about critiquing articles by assessing other students' critiques.

Viewing other students' critiques: Majority of students, 91% (32 of 35), reported that they learned about critiquing articles by viewing other student critiques. In the anonymity groups, more students in the named group, 56% (18 of 32), indicated that they learned about critiquing articles by viewing other students' critiques compared to the students in the anonymous group, 44% (15 of 32). In the peer accountable groups, equal number of students, 50% (16 of 32), in the more-accountable and less-accountable groups indicated that they learned about critiquing articles from viewing other students' critiques.

Own critiques: The analysis indicated that, 94% (33 of 35) students learned about critiquing articles by completing their own critiques. In the anonymity groups, more students in the named group, 55% (18 of 33), indicated that they learned from completing their own critiques compared to the anonymous group, 45% (15 of 33). In the peer accountable groups, approximately the same number of students in the more-accountable group, 52% (17 of 33), and less-accountable group, 48% (16 of 33), reported that they learned from completing their own critiques.

Peer feedback: Only 66% (23 of 35) students reported that they learned about critiquing articles from peer feedback. In the anonymity groups, 52% (12 of 23) students in the named group and 48% (11 of 23) students in the anonymous group indicated that they learned from peer feedback. Similarly, in the peer accountable groups, 52% (12 of 23) students in the less-accountable group and 48% (11 of 23) students in the more-accountable group reported that they learned about critiquing articles from peer feedback.

Helpful comments: Student reported, 83% (29 of 35), that they found peer comments as being helpful. In the anonymity groups, more students in the named group, 55% (16 of 29), found the peer comments helpful compared to the anonymous group, 45% (13 of 29). In the peer accountable groups, more students in the less-accountable group, 55% (16 of 29), found the peer comments helpful compared to the students in the more-accountable group, 45% (13 of 29).

Table 4.8 shows the students' views on the peer assessment process followed in this study. A detailed analysis of students' response to the questionnaire, on the peer assessment process and the procedures followed in the study, is follows.

Peer assessment process: Of the 35 students', 69% (24 of 35) reported that they found the peer assessment process in this study, easy to follow. In the anonymity groups, more students in the named group, 54% (13 of 24), found the peer assessment process easy to follow compared to the anonymous group, 46% (11 of 24). In the peer accountable groups, more students in the more-accountable group, 54% (13 of 24), found the peer assessment process easy to follow compared to the students in the less-accountable group, 46% (11 of 24).

Recommend peer assessment process: Of the 35 students, 71% (25 of 35), reported that they would recommend the peer assessment process followed in this study in another course. In the anonymity groups, more students in named group, 52% (13 of 25), indicated that they would recommend the peer assessment process followed in this study in another course, compared to the students in the anonymous group, 42% (12 of 25). In the peer accountable groups, more students in the less-accountable group, 52% (13 of 25), indicated that they would recommend the peer assessment process followed in this study in another course, compared to students in the more-accountable group, 48% (12 of 25).

Table 4.8

Students' Views on the Peer Assessment Process Followed in This Study

Statements	Agree (%)	Don't Know (%)	Disagree (%)
I found the peer assessment process in this course easy to follow.	69	0	31
I would recommend this peer assessment process in another course.	71	6	23
I found the grading scheme fair.	51	9	40
The number of assignments I assessed was reasonable.	94	0	6
The instructor's general feedback was sufficient.	31	23	46
I prefer to know the person whose assignment I am assessing.	37	3	60
As an assessor, I prefer to indicate my name on the assessments I submit.	63	0	37
My comments would have been different if I had known the person whose assignment I was assessing.	14	6	80

Grading scheme: The results of the questionnaire analysis indicated that 51% (18 of 35) students found the grading scheme fair. In the anonymity groups, more students in the anonymous group, 61% (11 of 18) found the grading scheme fair compared to the students in the named group, 39% (7 of 18). In the peer accountable groups, more students in the more-accountable group, 56% (10 of 18), found the grading scheme fair compared to the students in the less-accountable group, 44% (8 of 18).

Number of critiques assessed: Majority, 94% (33 of 35), of the students reported that they assessed reasonable number of student critiques. In the anonymity groups, more students in the named group, 56% (18 of 33), found the number of critiques assessed as reasonable compared to the students in the anonymous group, 48% (15 of 33). In the peer accountable groups, more students in the less-accountable group, 52% (17 of 33) found the number of critiques assessed as reasonable compared to the students in the more-accountable group, 48% (16 of 33).

Instructor's feedback: Only 31% (11 of 35) students reported that they found the instructor's feedback sufficient. In the anonymity group, more students in the anonymous group, 73% (8 of 11), found the instructor's feedback sufficient compared to the students in the named group, 27% (3 of 11). In the peer accountable groups, more students in the less-accountable group, 64% (7 of 11), found the instructor's feedback sufficient compared to the students in the more-accountable group, 36% (4 of 11).

Indicate name as peer assessors: Students' response to the questionnaire analysis indicated that 63% (22 of 35) students reported that as a peer assessor they would prefer to indicate their name on the assessments they submit. In the anonymity groups, more students from the named group, 59% (13 of 22), preferred to indicate their names as peer assessors compared to the students in the anonymous group, 41% (9 of 22). In the peer accountability groups, more students in the less-accountable group, 64% (14 of 22), preferred to indicate their names as peer assessors compared to the students in the more-accountable group, 36% (8 of 22).

Prefer to know the name of the student being assessed: Of the 35 students', 37% (13 of 35) students reported that they would like to know the name of the student they were assessing. In the anonymity groups, more students in the named group, 69% (9 of 13), preferred to know the name of the student they were assessing compared to the students in the anonymous group, 32% (4 of 13). In the peer accountable groups, more students in the more-accountable group, 54% (7 of 13) preferred to know the name of the student they were assessing compared to the students in the less-accountable group, 46% (6 of 13).

Peer comments vary with the anonymity condition: The results of the questionnaire analysis indicated that only 14% (5 of 35) students reported that they would provide different comments if they knew the name of the student whose critique (assignment) they were assessing. In the anonymity

groups, more students in the anonymous group, 60% (3 of 5), indicated that they would provide different comments if they knew the name of the student whose assignment they were assessing compared to the students in the named group, 40% (2 of 5). In the peer accountable groups, more students in the less-accountable group, 60% (3 of 5), indicated that they would provide different comments if they knew the name of the student whose assignment they were assessing compared to the students in the named group, 40% (2 of 5).

4.5 Summary of the Results

Results of this thesis study, organized by four research questions, can be summarized as follows:

1. As predicted, fewer peer assessors *over-marked* (i.e., assigned a mark higher relative to the instructor) in the anonymous group compared to the number of peer assessors in the named group.
2. As predicted, peer assessors in the anonymous group provided more *critical comments* (i.e., number of negative comments) compared to peer assessors in the named group.
3. As predicted, peer assessors in the more-accountable group provided more *quality comments* (i.e., the number of cognitive comments made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for

improvement) compared to the peer assessors in the less-accountable group.

4. Contrary to the hypothesis, there was no significant improvement in the students' performance after the peer assessment exercise. However, the students' response to the questionnaire indicated that they learned more from viewing and assessing other students' critique than from peer feedback. The students also reported that they found the peer assessment process in the study easy to follow. However, there was a mixed response on the grading scheme followed in the study.

Chapter four was a report on the data analysis and results on the experiment to determine the effect of anonymity and peer accountability during peer assessment in an online research methods course. Chapter five discusses the rationale for the results obtained, the conclusions drawn from the experiment, limitations of the study, and recommendations for future research.

CHAPTER 5

Discussion

Chapter five is a discussion of the results of the experiment, its conclusions, the limitations, and recommendations for future research.

5.1 Conclusions and Contributing Factors

This master's thesis research examined the effects of anonymity and peer accountability on peer over-marking, and the criticality and quality of peer comments in online peer assessment. Based on the four research questions and the data reported in chapter 4, following conclusions were drawn:

Peer over-marking. The first conclusion drawn from the results was that anonymous online peer assessment reduced the number of peer assessors over-marking. *Peer over-marking* was defined as the peer assessors assigning higher marks relative to the instructor, on a students' critique.

Previous research (Boud & Homes, 1995; Falchikov, 1986, 1995; Kelmar, 1993; Pond et al., 1995; Mowl & Pain, 1995; Rushton et al., 1993; Sluijsmans et al., 2001) indicates that peer assessors' have a tendency to

over-mark. This may be due to loyalty towards friends and social pressure. Consistent with the literature, this study found that overall peer assessors had a tendency to over-mark (see Table 4.1). However, this study showed that anonymous peer assessment reduced the numbers of peer assessors over-marking. This finding supports the hypothesis. Notably, there was no significant effect of anonymity (anonymous and named) on the number of peer assessors under-marking and assigning identical marks. The effect of the peer accountability (more-accountable and less-accountable) variable on peer marking was also examined. The results indicated that there was no significant relationship between peer accountability and the number of peer assessors over-marking, under-marking, and assigning identical marks.

Critical peer comments in peer feedback. The second conclusion of this thesis is that anonymous online peer assessment enhanced critical comments in peer feedback. *Critical comments* were the negative comments (or weaknesses) made by the peer assessors on the students' critiques.

Consistent with the literature, in this study, the peer assessors in the named group provided more complimentary comments. However, the peer assessors in the anonymous group provided more critical comments. This finding supports the hypothesis. Notably, the peer assessors in the named group provided significantly more comments (sum of positive and critical comments). Varying the degree of peer accountability (more-accountable and

less-accountable) did not affect the number of critical or positive comments made by the peer assessors. However, peer assessors in the more-accountable group provided significantly more comments.

Quality comments made by the peer assessors. The third conclusion drawn from the results was that in online peer assessment two variables, namely: peer accountability and anonymity significantly affect the quality of peer comments. *Quality comments* were statements made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement. The results in this study showed that between the peer accountability groups (more-accountable and less-accountable), peer assessors in the more-accountable group provided significantly higher number of quality comments compared to the less-accountable group (see Table 4.3). This finding supports the hypothesis.

It was interesting to note that between the anonymity groups (named group and anonymous), the peer assessors in the named groups made significantly more quality comments compared to peer assessors in the anonymous groups. One explanation for this could be that since in the named group the peer assessors were identifiable by their names, it forced them to give meaningful comments even if they are more complimentary than critical.

Therefore, it can be concluded that in online peer assessment the interaction of peer accountability and anonymity improves the quality of peer

comments. As mentioned in chapter 2, a comprehensive review of the extant literature published in and before 2003, found nothing to indicate the effect of peer accountability on peer comments in online peer assessment. Therefore, the observations made in this study on the effects of peer accountability are precedence setting.

Students' performance in critiquing research articles. Finally, the effect of peer assessment on students' performance was examined. *Students' performance* was defined as the student's ability to critique research articles. The data showed mixed results. Overall, there was no significant improvement in students' performance from critique 1 to critique 2 (see Table 4.5). This finding was contrary to what was expected. Also, there was no effect of the interaction of anonymity and peer accountability on the students' performance. Further, there was no effect of peer accountability on the students' performance. However, there was a significant effect of anonymity on the students' performance. The instructor-assigned marks from critique 1 to critique 2 improved significantly for the students in the named group. On the other hand, the difference in the instructor-assigned marks from critique 1 to critique 2 showed a decrease in the anonymous group (see Figure 4.1).

The improvement in the students' performance in the named group can be partly attributed to the number of quality comments provided by the peer assessors and received by the students in that group. In this study, peer

assessors in the named group, provided more quality comments compared to the peer assessors in the anonymous group (see Table 4.3). Similarly, students in the named group received more quality comments compared to the peer assessors in the anonymous group. Since the participants in the group that provided and received more quality comments (named group) also showed improvement their in performance, it could be concluded that there is a relationship between the quality comments and the students performance. Furthermore, between the named groups, the students in the more-accountable group provided more quality comments compared to students in the less-accountable group. The students in the same group (named, more-accountable group) also showed significant improvement in their performance compared to their counterparts (named, less-accountable group). This finding further strengthens the view that there is a relationship between the quality of peer comments and the students' performance.

There is some evidence in the literature (Webb, 1995) to indicate that the level of elaboration of comments is related to achievement. Webb's extensive review of the empirical studies on peer interactions in small groups (Webb, 1989) suggests that there is a positive correlation between giving an elaborated explanation and learner's achievement. Therefore, consistent with Webb's view, the data in this study showed that the students who provided more quality comments also showed significant improvement in their performance. However, further analysis to examine the pattern of the quality

of peer comments, was deemed beyond the scope of the experimental work on this master's thesis.

There seem to be various reasons for the lack of significant improvement in overall class performance. First, the instructor did not verify the correctness of the peer assessors' comments. Therefore, even though students may have received substantial peer comments, the content may not have been correct. As a result, the peer comments may not have contributed to learning and improvement in students' performance. The second reason for the lack of significant improvement in the students' performance could be that this experiment did not include a control group to compare the impact of peer assessment on the students' performance. The results on the students' performance may have varied if the performance of the students in the experimental group who participated in the peer assessment process (i.e., viewing and assessing other students' critiques, and providing peer comments and receiving peer feedback) was compared with the students in the control group who did not participate in the peer assessment process. The third reason for no significant improvement in the students' performance could be that the students' performance was measured after one peer assessment exercise. Studies (Anderson et al., 2001; Kuhn, 1991; Sluijsmans et al., 2001) indicate that the students ability to critique, assess and evaluate, improves with practice. Therefore, monitoring student's progress over a period of time may show different results. Finally, another reason for no significant

improvement in the students' performance could be the type and the level of difficulty of the content. In this study, the first article (critique 1), on which the peer assessment exercise was conducted, was simple and straightforward. However, the second article (critique 2), on which the improvement in the students' performance was measured, was more difficulty compared to the first one. This may have affected the results on the students' performance.

Questionnaire analysis. In addition to determining the improvement in the students critiquing ability based on the instructor-assigned marks, the students were also asked to respond to a questionnaire. The questionnaire was constructed to determine whether the students' perceived the peer assessment exercise beneficial and easy to follow. Information was sought in three areas: (1) learning benefits from viewing and assessing other students' work, (2) learning benefits from receiving peer feedback, and (3) the peer assessment procedure followed in this study. The data on the students response to the questionnaire showed that the students found the peer assessment exercise beneficial (see Table 4.6). However, they perceived more learning benefits from viewing and assessing other students' work compared to receiving peer feedback. The students also found the peer assessment process in this study easy to follow. In fact, majority indicated that they would recommend the peer assessment process followed in this study, in other courses (see Table 4.7).

To summarize the conclusions drawn from the results on the effects of anonymity and peer accountability on peer over-marking, and the criticality and quality of peer comments in a graduate web-based education research methods course: (a) anonymous online peer assessment reduced peer over-marking; (b) anonymous online peer assessment enhanced critical comments in peer feedback; (c) Increased degree of peer accountability improved the quality of peer comments. Under certain conditions anonymity also improved the quality of peer comments; (d) Even though there is no clear evidence to indicate improvement in students performance, the students' response to the questionnaire indicated that they found the peer assessment process useful. The results of the questionnaire suggest that the students' learned more from viewing and assessing other students work than from peer feedback.

5.2 Implications from these Conclusions

The first implication from the discussion of the results of the study is that anonymous online peer assessment can affect peer over-marking. In this study, the peer assessors in the named group had a tendency to assign a higher mark relative to the instructor's mark. However, there was no significant difference in under-marking or identical-marking by the peer assessors in the anonymous group and the named group.

Closely associated with this prediction, the second implication is the likelihood that online peer assessment enhances critical comments in peer

feedback. In this study, the peer assessors in the anonymous group provided more critical comments and the peer assessors in the named group provided more complimentary comments.

The third implication from the results of the study is that in online peer assessment, degree of peer accountability does affect the quality of comments made by the peer assessors. In this study, the peer assessors in the more-accountable group provided more quality comments compared to the peer assessors in the less-accountable group. Further, the interaction of anonymity with peer accountability also affected the quality of peer comments.

The fourth implication is that the peer assessment process as described in chapter 3, may not improve the students' critiquing ability as predicted. However, there seems to be a relationship between quality of peer comments and students' performance. Studies in past have attempted to measure the impact of peer assessment on higher order learning skills such as critical thinking (Anderson et al., 2001; MacPherson, 1999), assessment skills (Sluijsmans et al., 2002), evaluative skills (Falchikov, 1986). However, no clear results on learning benefits have been reported in these studies. Falchikov (2001) notes that, "lack of evidence of cognitive development may simply reflect the difficulty of measuring it" (p.74). Despite unclear empirical evidence of the effects of peer assessment on student learning, all studies reported that students found the peer assessment process helpful (Anderson

et al., 2001; Falchikov, 1996; Sluijsmans et al., 2002; Topping et al., 2001). Consistent with these studies, the students response to the questionnaire in this study also showed that they found the peer assessment process a useful learning tool.

An overall conclusion is that online peer assessment, as an instructional method, can be incorporated in the course curriculum to provide practice to students in developing critical, evaluative and analytical skills. However, the emphasis should be on ensuring that the peer assessors provide quality feedback. Searby & Ewers (1997) stated, "It is difficult to prove conclusively that the use of peer assessment will improve the quality of students work. Even if it does not, it is still a valuable exercise as it clarifies the learning goals for students through the development of criteria for assessment" (p.381-82).

5.3 Limitations Of The Study

The first limitation of this study was that no attempt was made to understand the degree to which the participants knew each other. It remains to be determined, whether in online peer assessment the degree to which the participants know each other, affects peer marking and peer comments. As Bump (1990) suggests that in an online setting removing or stating the names may not make the condition clearly anonymous or identifiable, as the students may not necessarily know their peers despite knowing their names.

The second limitation of this study was that it did not determine the effect of anonymity and peer accountability on the degree of agreement between the peer-assigned marks and the instructor-assigned marks. In this study, the peer-assigned marks and the instructor-assigned marks were compared to determine peer over-marking. Empirical studies (e.g., Falchikov & Goldfinch, 2000; Falchikov, 1986,1995; Magin, 2001) on the degree of agreement between the peer-assigned marks and the instructor-assigned marks have been determined either through correlation analysis or through analysis of marks. Many of these studies found a high degree of agreement in peer-instructor mark (e.g., Falchikov & Goldfinch, 2000; Falchikov, 1995). However, some other studies (e.g., Mowl & Pain, 1995) found poor agreement between the peer and the instructor's mark. In this study, it remains to be examined how anonymity and peer accountability in online peer assessment affect the validity of peer-assigned mark relative to the instructor's mark.

The third limitation of this study was that the instructor did not verify correctness of peer comments. Therefore, even though the assessor may have provided substantial feedback, peer comments may not be correct. This may affect learning. Further, the instructor did not provide any qualitative feedback on the students' critiques. In this study design, the instructor's assessment included only assigning a numeric mark on a student's critique. As mentioned in chapter 3, prior to this study, the design of the experiment

was tried with another group of students. Based on the observations made during the trials, there were two reasons for the absence of the instructor's qualitative comments on the students' critiques. First, during the trials it was observed that since the instructor provided qualitative feedback, the students did not take peer comments seriously. This attitude is consistent with the literature that indicates that students trust the instructor's feedback more than peer feedback (e.g. Davis, 2000; Falchikov 2001; Pond et al., 1995, Sluijsmans et al., 2001, Topping et al., 2000; Zhao, 1998). Second, during the trials it was noted that when both the instructor and the peers provided qualitative feedback, it was difficult to determine whether the student benefited from the instructor's feedback or peer feedback. However, in the absence of the instructor's comments, verifying the correctness of peer comments seems important as this may affect student learning. The effect of verifying correctness of peer comments, on student learning needs to be examined.

The fourth limitation of this study was that student' attitude towards peer assessment was not taken into account. O'Donnell & Topping (1998) suggest that the efficacy of feedback depends on both the giver and the receiver. Some studies (Falchikov, 2001; O'Donnell & Topping, 1998) found that male students may not act upon peer feedback as female students. Therefore, in this study, even though the peer assessor may have provided substantial feedback but the student assessed may not have acted upon the

peer comments due to a personality type. Also, the students learning styles were not taken into consideration. For instance, Lin et al. (2001) found that students' with high executive thinking styles provided better feedback than their low executive counterparts. Similarly, Webb (1995) suggested that it is important to know whether the student assessed understood peer comments. Therefore, it may be important to examine how the peer assessors provide the feedback and how do the students' assessed incorporate peer feedback. This may affect student performance.

The fifth limitation of this study was that the difference in students' ability to critique research articles were judged based on only one peer assessment exercise. Studies indicate that critiquing skills and assessment skills improve with practice (Anderson et al., 2001, Slujismans et al., 2002). Therefore, further opportunities in assessing other students' work may improve quality of peer comments and students ability to critique. Also, the improvement in students' performance was measured on the research articles that were different in terms of complexity and the level of difficulty. Thus the difference in performance should be measured using a more reliable and accurate method.

In sum, the process of describing the limitations of this thesis study provides the necessary narrow context for the conclusions of the study. The following section summarizes the conclusions drawn from this study.

5.4 Summary

This thesis study attempted to examine the effects of anonymity and peer accountability on peer marking and peer comments during peer assessment in a graduate web-based education research methods course. The data indicated that the interaction of anonymity and peer accountability helped in minimizing problems in peer marking and comments. This finding may help in enhancing the benefits expected from the peer assessment process. O'Donnell & Topping (1998) suggest "peer feedback might be of poorer quality than that provided by teacher. However, peer feedback is usually available in greater volume and with greater immediacy than teacher feedback, which might compensate for any quality disadvantage" (p. 262).

Despite the encouraging results, extended interventions of anonymity and peer accountability during online peer assessment may be required to produce a more comprehensive understanding of solutions to these research questions. In doing so, it may be prudent to modify the model of peer assessment or replace some of the procedures and instruments with others of more complex or sensitive design. In any case, that task is beyond the scope of this thesis. It is hoped that these findings provide some direction for researchers and educators about peer assessment in an online learning environment.

REFERENCES

Ahern, T. C., Peck, K., & Laycock, M. (1992). The effects of teacher discourse in computer-mediated discussion. *Journal of Computing Research, 8*(3), 291-309.

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*, 1-32.

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Boston: Allyn and Bacon.

Allen, C. L. (1992). Multimedia learning environments designed with organizing principles from non-school settings. In E.DeCorte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 464-484). London:Springer-Verlag.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*, 349-356.

Berge, Z. L., & Collins, M. (1993). Computer conferencing and on-line education. *The Arachnet Electronic Journal on Virtual Culture, 1*(3). Retrieved August 19, 2003, from http://www.emoderators.com/papers/bergev1n3.html

Blumhof, J., & Stallibrass, C. (1994). *Peer assessment*. Hatfield, UK: University of Hertfordshire Press.

Bordia, P. (1997). Face-to-Face versus computer-mediated communication: A synthesis of the experimental literature. *Journal of Business Communications, 34*(1), 99-120.

Borman, W., White, L., & Dorsey, D. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80*, 168-177.

Bostock, S.J. (2000). *Computer assisted assessment - experiments in three courses*. Retrieved August 19, 2003, from http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/caa-ktn.htm

Boud, D., & Holmes, H. (1995). Self and peer marking in a large technical subject. In D. Boud, *Enhancing learning through self-assessment* (pp. 63-78). London: Kogan Page.

Bump, J. (1990). Radical changes in class discussion using networked computers. *Computers and Humanities*, *24*, 49-65.

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups. *Management Science*, *36*(6), 698-703.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10,* S33-S49.

Dancer, W., & Dancer, J. (1992). Peer rating in higher education. *Journal of Education for Business, 67*, 306-309.

Davis, P. (2000). Computerized peer assessment. *Innovations in Education and Training International, 37*(4), 346-355.

Davis, R., & Berrow, T. (1998). An evaluation of the use of computer supported peer review for developing higher-level skills. *Computers and Education, 30* (1/2), 111-115.

Downing, T., & Brown, I (1997). Learning by cooperative publishing on world wide web. *Active Learning, 7,* 14-16.

Ellis, L (1984). The effects of anonymity on student ratings of college teaching and course quality. *Journal of Instructional Psychology, 11*, 182-186.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education, 11*(2), 146-166.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*, 175-187.

Falchikov, N. (1996). Improving learning through critical peer feedback and reflection. *Higher Education Research and Development, 19,* 214-218.

Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education, 22,* 385-397.

Falchikov, N., & 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.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London: Routledge Falmer.

Feenberg, A. (1987). Computer conferencing and the humanities. *Instructional Science, 16,* 169-186.

Fenwick, T., & Parsons, J. (2000). *The art of evaluation: a handbook for educators and trainers.* Toronto: Thompson Educational Publication.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education, 20*, 289-300.

Goldberg, M. (1997). Communication and collaboration tools in WebCT. *Proceedings of the Conference Enabling Network-Based Learning*, Espoo, Finland, May 28 - 30.

Gordon, R. A., & Stuecher, U. (1992). The effects of anonymity and increased accountability on the linguistic complexity of teaching evaluations. *Journal of Psychology, 126*(6), 639-649.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17(*4), 397-431.

Haaga, D. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology, 20*(1), 28-32.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28,* 115-152.

Harkins, S., & Petty, R. (1992). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology, 43*, 1214-1229.

Helmore, P., & Magin, D. (1998). Peer and teacher assessment of conference presentations by final year students. In P. Lep Darvall and Z. Pudlowski (Ed.), Globalization of engineering education. *Proceedings of the 1st UICEE conference on engineering education, Melbourne,* 284-288.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Eds.), *Collaborative learning through computer conferencing* (pp. 117-136). Berlin: Springer-Verlag.

Hiltz, S. R., Turoff, M., & Johnson, K. (1989). Disinhibition, deindividuation, and group process in pen name and real name computer conferences. *Decision Support Systems, 5*(2), 217-232.

Hittleman, D. R., & Simon, A. J. (2002). *Interpreting educational research: An introduction for consumers of research* (3rd ed.). Columbus, OH: Merrill.

Howell-Richardson, C., & Mellar, H. (1996). A methodology for the analysis of patterns of participation within computer-mediated communication courses. *Instructional Science, 24,* 47-69.

Jacobs, R. M., Briggs, D. H., & Whitney, D. R. (1975). Continuous-progress education: Student self-evaluation and peer evaluation. *Journal of Dental Education, 39*(8), 535-541.

Jonassen, D.H., & Kwon, H. (2001). Communication patterns in computer-mediated versus face-to-face group problem solving. *Educational Technology Research and Development, 49*(1), 35-51.

Kahai, S., Avolio, B., & Sosik, J. J. (1998). Effects of source and participant anonymity and difference in initial opinions in an EMS context. *Decision Sciences Journal*, *29*(2), 427-460.

Kelmar, J. H. (1993). Peer assessment in graduate management education. *International Journal of Educational Management, 7*(2), 4-7.

Kuhn, D. (1991). *The skills of argument.* Cambridge: Cambridge University Press.

Kwok, R. C. W., & Ma, J. (1999). Use of a group support system for collaborative assessment. *Computers and Education, 32*(2), 109-125.

Leedy, P. D., & Ormrod, J. E. (2001). *Practical research: Planning and design* (7th ed.). New Jersey: Merrill Prentice Hall.

Lin, S. S. J., Liu, E. S. F., & Yuan, S. M. (2001). Web-based peer assessments: Feedback for students with various thinking styles. *Journal of Computer Assisted Learning, 17,* 420-432.

MacPherson, K. (1999). The development of critical thinking skills in undergraduate supervisory management units: Efficacy of student peer assessment. *Assessment and Evaluation in Higher Education, 24*(3), 273-284.

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26*(1), 53-63.

Makkai, T., & McAllister, I. (1992). Measuring social indicators in opinion surveys: A method to improve accuracy on sensitive questions. *Social Indicators Research, 27*(2), 169-186.

Mann, B. (2000). WebCT: Serving educators in Newfoundland & Labrador. *The New Morning Watch: Educational and Social Analysis, 28*(1/2). Retrieved August 19, 2003, from http://www.mun.ca/educ/faculty/mwatch/fall00/mann.htm.

McBearty, P. (1982). We should abolish anonymous evaluations of teachers by their students. *Chronicle of Higher Education, 25,* 48.

McCollister, R. J. (1985). Students as evaluators: Is anonymity worth protecting? *Evaluation and the Health Professions, 8,* 349-355.

Miyake, N. (1987). Constructive interaction and iterative process of understanding. *Cognitive Science, 10*, 151-177.

Mowrer, D. E. (1996). A content analysis of student/instructor communication via computer conferencing. *Higher Education, 32,* 217-241.

Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay-writing: A case study from geography. *Innovations in Education and Training International*, *32*(4), 324-336.

O'Donnell, A. M., & Topping, K. (1998). Peers assessing peers: Possibilities and problems. In K. Topping, & S. Ehly (Eds.), *Peer assisted learning* (pp. 255-278). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Donnell, A. M., & King, A. (1999). *Cognitive perspectives on peer learning.* Mahwah: NJ: Lawrence Erlbaum Associates.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 25*(1), 23-38.

Pinsonneault, A., & Heppel, N. (1998). Anonymity in group support systems research: A new conceptualization, measure, and contingency. *Journal of Management Information Systems, 14*(3), 89-108

Pond, K., Rehan, U., & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education and Training International*, *32*(4), 314-323.

Price, K. H. (1987). Decision responsibility, task responsibility, identifiabilty and social loafing. *Organizational Behavior and Human Decision Processes, 40*, 330-345.

Rada, R. (1998). Efficiency and effectiveness in computer-supported peer-peer learning. *Computers and Education, 30, (3/4),* 137-146.

Romiszowski, A., & Mason, R. (1996). Computer-mediated communication. In D. Jonassen (Eds.), *Handbook of research for educational communications and technology* (pp. 438-456). New York: Macmillan.

Rourke, L., & Anderson, T. (2002). Using peer teams to lead online discussion. *Journal of Interactive Media in Education, (1).*

Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: A case study. *Journal of Computer-Based Instruction, 20*(3), 75-80.

Searby, M. & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the school of music. Kingston University. *Assessment and Evaluation in higher Education, 22*(4), 371-383.

Sluijsmans, D. M. A., Moerkerke, G., Dochy, F., & Van Merriënboer, J. J. G. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation, 27*(2), 153-173.

Sluijsmans, D. M. A., Saskia, B., & Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education*, *27*(5), 443-454.

Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75.

Stone, E. F., Spool, M. D., & Robinowitz, S. (1977). Effects of anonymity and retaliatory potential on student evaluations of faculty performance. *Research in Higher Education, 6,* 313-325.

Swanson, D., Case, S., & Van Der Vleuten, C. (1991). Strategies for student assessment. In D. Boud, & G. Feletti (Eds.), *The challenge of problem based learning* (pp. 260-273). London: Kogan Page

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*, 74-83.

Topping, K., & Ehly, S. (Eds.). (1998). *Peer assisted learning*. Lawrence Erlbaum Associates, Mahwah: NJ.

Topping, K. J., Smith E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, *25*(2), 149-169.

Towler, L., & Broadfoot, P. (1992). Self-assessment in the primary school. *Educational Review, 44*(2), 137-151.

Tsai, C. C., Liu, E. Z., Lin, S. S. J., & Yuan, S. (2001). A networked peer assessment system based on a vee heuristic. *Innovations in Education and Teaching International, 38*(3), 220-230.

Walther, J. B., & Tidewell, L. C. (1995). Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing, 5*(4), 355-378.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21-40.

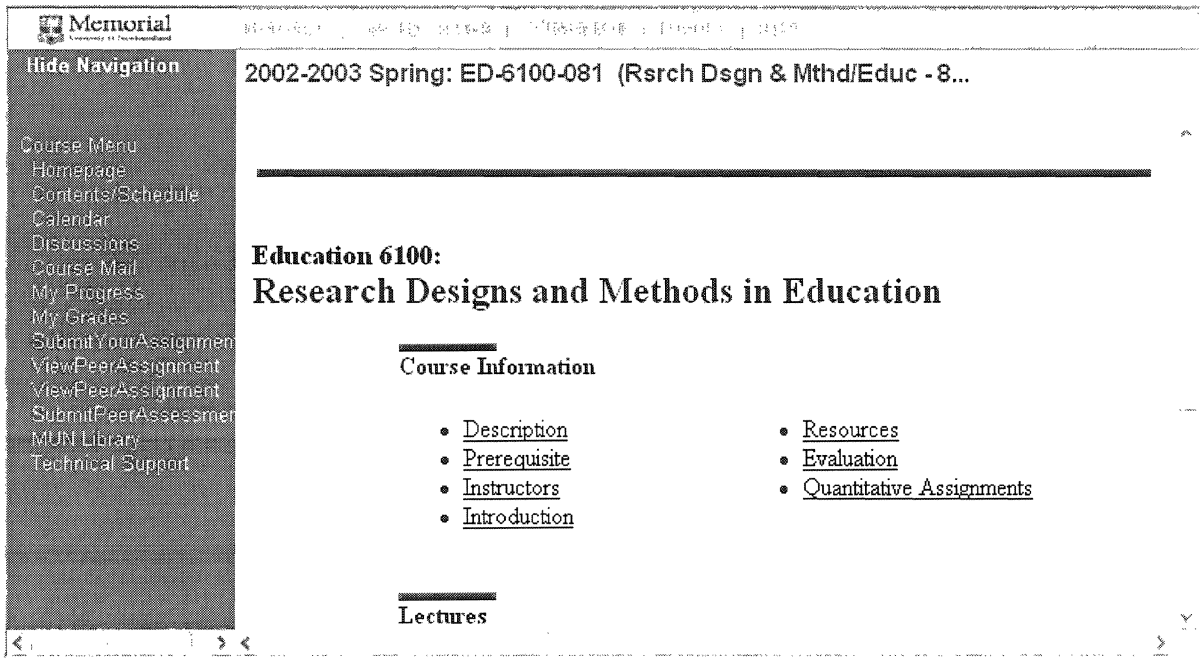
Webb, N. M. (1995). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz- Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 102-119). Cambridge: Cambridge University Press.

Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunications, 4*(4), 311-345.

Zhu, P. (1998). Learning and mentoring: Electronic discussion in a distance learning course. In C. J. Bonk & K. S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 233-259). Mahwah, NJ: Erlbaum.

Appendix A

A View Of The Course Homepage In WebCT For The Education Research Methods Course



Appendix B

Details On The First Research Article And The Criteria For Critiquing The Article

Critique 1

Parush, A., Hamm, H., & Shtub, A. (2002). Learning Histories in Simulation-based Teaching: The effects on Self-learning and Transfer, *Computers & Education*, 39(4), 319-332. (This research uses statistical procedures very common in experimental studies, analysis of variance).

You can access the article through the MUN library by clicking on the "MUN Library" link under the "Course Menu". Select the "Off Campus Login" option and indicates your UserID and PIN to access electronic journals. Select EJournals, under Quick Search, to access the article.

NOTE: In case you do not have a UserID and PIN to access the electronic journals or you are unable to access the article, you can contact the Library Information Services at 737-7427.

Procedures for Critiquing/Reviewing a Research Article

A critique consists of a clear identification of the study using ideas and terms associated with quantitative research methods (e.g., design, variables, experimental, survey). Second is an appraisal of the quality of the study using the notions of validity (internal, external, instrumentation), accuracy of reporting and correctness of interpreting results, generality of findings, clear implications for questions and theory (and practice). The general structure of the research report guides the critique:

Problem: clarity, development, etc. (basis in the literature, theory; hypotheses).

Method: participants (who, sampling), instruments (validity and reliability), design.

Results: summary, descriptive statistics; analysis (including inferential statistics).

Conclusions: answer to problems, discussion and implications (including literature).

For this article the critique will include only the Problem and the Method portion of the study. Your critique should be brief and focus on critique, and not description.

Appendix C

The Instructor's Message Posted On The Discussion Board Regarding The Purpose Of Peer Assessment In The Course

Subject: Peer assessment in the course

Reply Reply privately Quote Download

Message no. 12

Author: Henry Schulz (81275200203)

Date: Monday, May 12, 2003 2:44pm

Hello students in E6100.

As part of the course this spring we are including Peer Assessment of the first assignment for the Quantitative Methods portion of the course. This is intended to expedite your learning, and increase the feedback that you receive on your work in the course. Below is a brief summary of peer assessment and how it will be used in the course.

Peer Assessment

Peer assessment is a process whereby students in a class critique and assess each other's work. Peer assessment is a well-recognized strategy that enhances learning. It encourages student autonomy as well as critical thinking. It provides a mechanism for students to receive additional feedback (beyond what instructors can reasonably give), and it ensures that all students receive substantial and timely feedback. It is well suited to the course since the intent of the course is for you to learn to analyze and critique research reports in education.

What this means for the course.

For the **Quantitative Methods** there are two assignments. Both involve the review and critique of a research paper that uses a quantitative approach.

The first assignment requires that you critique selected aspects of the research paper identified in the "Quantitative Assignment" on the "Homepage" (the paper can be accessed on-line through the MUN library). You then submit this critique for others in the course to view, and provide an assessment of the critiques of three or

four other students in the course. We will identify the students whose assignments you are to assess (this is to ensure that all submissions are peer assessed).

The second assignment requires that you provide a complete review of a second research paper (again, we will identify it, and it too can be accessed using the MUN library on-line facilities).

The peer assessments of your first assignment should assist you in developing your understanding of quantitative methods, and thereby produce a more complete critique for the second assignment.

Quantitative Assignment 1

- You are to review and critique the research article. Click on the "Quantitative Assignment" link on the course "Homepage" to view the assignment details. The due date for submitting the assignment is **02 Jun 2003.**
- Once all the assignments have been submitted, we will upload student assignments for others in the class to view. You will be told on the "Discussions" forum when the assignments are ready for review. Since the class is quite large, you will be able to view approximately half of the class' assignments. However, you will assess only three or four of the assignments for the peer assessment. To ensure fairness, we will randomly select the ones which you are to peer assess (we will inform you of these through the "Course Mail").
- The peer assessment process will involve you assigning a numeric mark to the student paper and giving comments on it. I will mark each assignment independently (without viewing your marks and comments).
- Your peer assessment of the assignments should be submitted by **08 Jun 2003.**
- The grade given to the assignment will be the average of the three or four peer assessments and this average combined with the mark that I assign as instructor.
- Once you have submitted feedback on other students' assignments, your feedback will be sent to the students, and you will receive the peer feedback on your assignment within three days of submission.

Quantitative Assignment 2

- You are asked to review and critique a second research article. Click on the "Quantitative Assignment" link on the course "Homepage" to view the assignment details. The due date for submitting the second assignment is **16 Jun 2003.**
- There will be no peer assessment for Assignment 2. The mark assigned by me on the assignment will be the final mark for this assignment.

Your Final Mark On The Quantitative Methods

The average of your grade on the two assignments, and participation in this part of your course, will comprise your FINAL MARK for the Quantitative Methods component. There will be no final exam for the quantitative methods.

Henry Schulz MUN

Appendix D

A View Of The "ViewPeerAssignment" Page Identifying Each Uploaded Assignment With An Assigned Student Number.

2002-2003 Spring: ED-6100-081 (Rsrch Dsgn & Mthd/Educ - 81275)

Homepage > **ViewPeerAssignment**

Each of you will be able to view some student assignments. Click on the hyperlinked number to view the assignment.

Please check your "Course Mail" for the assignments to be assessed by you.

Please submit peer assessment by **08 Jun 2003**.

Table of Contents

1. Student (1)
2. Student (2)
3. Student (3)
4. Student (4)
5. Student (5)
6. Student (6)
7. Student (7)
8. Student (8)
9. Student (9)

Appendix E

A Sample Of An Uploaded Critique Of A Student In The Anonymous Group

Assignment 1

By

Student (1)

This critique is of the research article " Learning Histories in Simulation- Based Teaching: the effects on self-learning and transfer" by A. Parush, H. Hamm and A.Shtub and was published in Computers and Education 39 (2002) 319-332.

The research problem is clearly identified as to how the self-learning process interacts with simulation-based teaching and learning. This topic is of significance to study due to the fact that simulators have become an integral part of management and engineering students' education. The efficiency and effectiveness of the use of simulators would be valuable information for the engineering faculty in terms of the future use of such technology.

The literature review and subsequent discussion uses a historical context where teaching methods and individual learning processes have been studied over time and continue to be studied particularly with the onset of the use of technology.

Appendix F

A Sample Of An Uploaded Critique Of A Student In The Named Group

Quantitative-Assignment 1
By
Juan Gelded (Student 13)

Parush et al observed a growing number of opportunities where the built-in learning history were incorporated to several simulation based teaching tools, introducing a new concept in simulation-based teaching techniques. The use of simulated environments as a tools for learning is becoming quite popular due to the recognition of their teaching capabilities, reducing the learning process time and allowing students to perform better in task involved. Unfortunately, not too much research has been done to prove the theory that the use of previously recorded learning histories in simulation based learning environments will increase the efficiency and effectiveness of the learning process. The authors conducted a controlled experiment to demonstrate that the use of previous recorded experiences in simulators will improve trainees performance during the learning process itself.The authors explain the problem with clarity, although there is a great quantity of research and literature regarding the advantages of the use of simulated-based environments to enhance the learning process in complex scenarios, they recognize the lack of sufficient research to specifically prove the effectiveness and efficiency of learning histories in simulation-based teaching as literature suggests, therefore, they set an experiment to explore the effects of history.

Appendix G

The Instructor's Message On The Discussion Board About Criteria For Assessing Student Critiques

Message no. 367
Posted by **Henry Schulz (81275200203)** on Thursday, June 5, 2003 6:43am
Subject Assignments, peer assessment...

Thanks very much for submitting assignments as promptly as you did. Thanks also to Gunita for getting them uploaded. Although there were a couple of glitches (in getting the uploaded assignments to look exactly as you had written them) this has made it possible for you to see the kinds of work you have done and how much you have learned. I have tried to stay somewhat in the background as I think what many of you have said and done has contributed to your own understanding and to that of others. Just getting things out in front of others is a very important aspect of learning, and of the way we develop in education.

Many of you have asked for further guidance in the grading component. I asked that you use a scale of 1 to 10 in your numerical grade. I did not give a detailed rubric as part of what is important is to develop your own sense of the criteria that you should consider in looking at a piece of research and to critique it. Also, there are many aspects that can be addressed, and few right/wrong responses (except perhaps in deciding whether the research was experimental or not). Issues of internal validity and, particularly, external validity are complex and many aspects can be brought into consideration.

You were to critique the first two main parts of a research report: the introduction leading to a statement of the problem and hypotheses, and the design and methodology of the study. This provides some indication of the focus of the critique. A research paper should report the basis for the problem and hypotheses, which is a summary of relevant research and theory, and also of personal and professional experience. This should include the justification for and relevance of the research. The research paper should then provide a clear indication of the methods of the study: the individuals involved including the sampling procedures, the design of the study with particular note re how extraneous factors were controlled, the variables and how the dep var was measured and if it/they were valid and reliable, and the steps in the procedures (time, etc.). Internal and external validity are key concepts and should feature in the critique.

I have tried to make this so that all of you would receive important feedback--and more and more timely than I would ever be able to give. The grade component is important too. Note, to keep it in perspective, the grades you give in the peer assessment for this assignment will be averaged (over the three grades) and this will

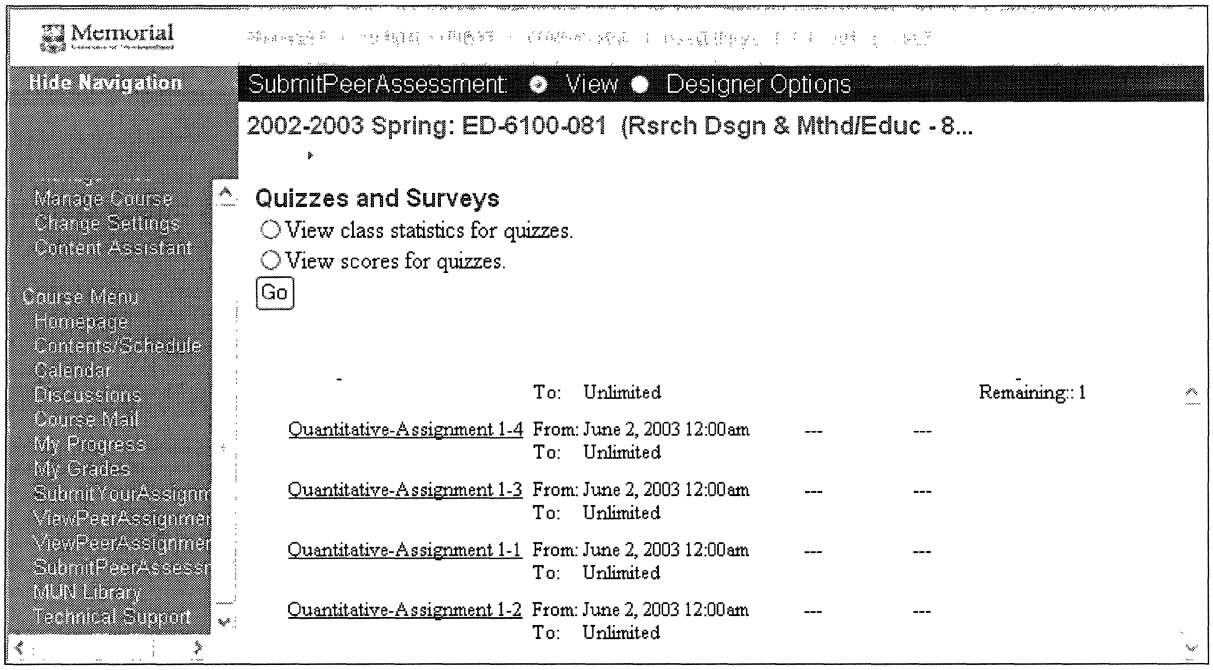
then be averaged with the grade that I give (which will be done independently of those given by you).

If you have questions, do not hesitate to ask. I will say more about the second quantitative assignment very shortly.

Henry Schulz

Appendix H

A View Of The "SubmitPeerAssessment" Screen To Submit Peer Assessments



Appendix I

A Sample Of Compiled Peer Assessment (Peer Assigned Marks And Qualitative Comments) Sent To Each Student

Student (10) - Nerissa Belben

How would you grade (/10) Nerissa's assignment? Please provide comments

User ID	Response
1 Leigh-Ann Ryder	Mark= 7/10 Comments: I thought it was very interesting that you and I started the assignment in a very similar way. I also felt that the main objectives were stated in the 2 questions you cited. As well, the "transfer of what was learnt" seemed unclear to me and I wondered if it should not have been operationally defined. The hypotheses you selected were given early in the article, but it could be mentioned that other, more pointed hypotheses, are also listed a little further along in the article. Mentioning these as well could add further strength to the cause and effect relationship. I was interested in what you wrote about classifying the experiment as quasi... as I had not considered that. With the threats to internal validity I wasn't sure that fatigue would enter in as it was a factor for both groups.
6 Cynthia Gardiner	Mark= 8.5/10 Comments: -good paraphrasing and identification of the problem statement; -identified research hypotheses and type of hypotheses; -causal comparative relationship identified; -identified the independent and dependent variables, but not any control or intervening variables; -method section of critique provides too much text as description of the study that is already known by the reader, -identified random assignments the determining factor for a true experimental design; -diagram of experiment design is not accurate.
9 David Conway	Mark=8.5 /10 Comments: Overall, good analysis of this research study. Your statement of the problem, hypothesis and research methods were clear and concise. Comparing the problem statement to the definition in Bieger and Gerlach was an effective way to evaluate if the problem statement in this study met the criteria for problem statements. You also raised an interesting point when stating what is meant by "the transfer of learning" in that it lacks clear definition in this study. When reviewing this study, I also did not consider the fact that the subjects were .....

Appendix J

Details On The Criteria for Critiquing The Second Research Article

Procedure for Critiquing / Reviewing a Research Article

A critique consists of a clear identification of the study using ideas and terms associated with quantitative research methods (e.g., design, variables, experimental, survey). Second is an appraisal of the quality of the study using the notions of validity (internal, external, instrumentation), accuracy of reporting and correctness of interpreting results, generality of findings, clear implications for questions and theory (and practice).

The general structure of a research report guides the critique:

- Problem: clarity, development, etc. (basis in the literature, theory; hypotheses).
- Method: participants (who, sampling), instruments (validity and reliability), design.
- Results: summary, descriptive statistics; analysis (including inferential statistics).
- Conclusions: answer to problems, discussion, implications (including literature).

For Assignment 1 the critique will include only the Problem and the Method portion of the study. Your critique should be brief and focus on critique, and not description.

Appendix K

Questionnaire On Students' Perception

Scale

(5) Strongly Agree	(4) Agree	(3) Undecided	(2) Disagree	(1) Strongly Disagree

Statements

About Learning

1. I learned about the assignment by assessing other assignments
2. I learned about the topic from completing my own assignment.
3. I learned about the topic from viewing other assignments
4. I learned about the topic from peer feedback
5. I found the peer comments helpful

About Process

1. I found the grading scheme fair
2. I found the peer assessment process in this course easy to follow
3. I would recommend this peer assessment process in another course
4. As an assessor, I prefer to indicate my name on the assessments I submit
5. My comments would have been different if I had known the person whose assignment I was assessing
6. I prefer to know the person whose assignment I am assessing
7. The number of assignments I assessed was reasonable.

Appendix L

Details Of Peer And Instructor-Assigned Mark On Students' First Critique

Groups	Student Number	Peer Marks			Peer-Assigned Mark (P1+P2+P3/3)	Instructor-Assigned mark	Difference
		Peer 1	Peer 2	Peer 3	(P)	(I)	(P – I)
Group 1 Anonymous More-accountable	1	8.00	9.50	8.50	8.67	8.50	+0.17
	2	7.00	8.50	8.50	8.00	8.00	0.00
	3	8.00	8.00	7.00	7.67	7.50	+0.17
	4	8.00	7.00	6.50	7.17	8.00	-0.83
	5	6.50	8.70	7.50	7.57	8.50	-0.93
	6	8.50	9.00	9.00	8.83	8.50	+0.33
	7	8.00	6.00	8.00	7.33	9.50	-2.17
	8	7.00	8.00	7.50	7.50	7.50	0.00
	9	9.00	9.00	7.50	8.50	7.50	+1.00
Group 2 Named More-accountable	10	7.00	8.50	8.50	8.00	8.00	0.00
	11	7.00	8.50	8.50	8.00	7.50	+0.50
	12	8.00	7.50	8.50	8.00	8.50	-0.50
	13	9.00	8.50	8.00	8.50	8.50	0.00
	14	9.00	8.50	8.50	8.67	9.50	-0.83
	15	8.00	9.00	7.50	8.17	7.50	+0.67
	16	9.00	8.75	8.00	8.58	7.50	+1.08
	17	9.50	8.00	9.00	8.83	8.00	+0.83
	18	8.00	7.50	8.00	7.83	7.50	+0.33
Group 3 Anonymous Less-accountable	19	9.00	8.50	6.00	7.83	8.00	-0.17
	20	9.00	9.00	7.50	8.50	9.00	-0.50
	21	7.40	8.00	8.50	7.97	7.50	+0.47
	22	8.00	8.00	8.50	8.17	8.50	-0.33
	23	7.50	9.00	6.50	7.67	8.50	-0.83
	24	9.00	8.50	8.00	8.50	8.50	0.00
	25	7.50	8.00	8.00	7.83	7.00	+0.83
	26	7.75	8.50	8.00	8.08	7.50	+0.58
	27	8.35	8.00	9.50	8.62	9.00	-0.38
Group 4 Named Less-accountable	28	8.50	8.00	9.00	8.50	7.50	+1.00
	29	9.00	8.50	9.00	8.83	8.00	+0.83
	30	8.00	8.80	8.00	8.27	8.50	-0.23
	31	7.50	8.70	8.00	8.07	7.50	+0.57
	32	8.90	8.50	7.50	8.30	8.00	+0.30
	33	9.50	8.50	7.00	8.33	7.00	+1.33
	34	9.00	6.00	8.00	7.67	7.00	+0.67
	35	8.50	7.50	8.00	8.00	7.50	+0.50
	36	8.00	9.00	9.00	8.67	7.50	+1.17

Appendix M

Details Of Critical And Positive Comments Made By The Peer Assessors

Groups	Student Number	Comments		Total Comments (C + P)
		Critical (C)	Positive (P)	
Group 1 Anonymous More-accountable	1	13	0	13
	2	16	18	34
	3	4	19	23
	4	11	8	19
	5	9	9	18
	6	1	8	9
	7	16	5	21
	8	8	5	13
	9	12	3	15
		90	**75**	**165**
Group 2 Named More-accountable	10	2	10	12
	11	3	25	28
	12	20	22	42
	13	13	15	28
	14	17	36	53
	15	30	27	57
	16	7	8	15
	17	9	13	22
	18	4	22	26
		105	**178**	**283**
Group 3 Anonymous Less-accountable	19	2	3	5
	20	15	6	21
	21	29	32	61
	22	6	2	8
	23	4	3	7
	24	27	15	42
	25	6	0	6
	26	2	7	9
	27	4	0	4
		95	**68**	**163**
Group 4 Named Less-accountable	28	1	8	8
	29	6	16	22
	30	2	28	30
	31	2	6	8
	32	8	14	22
	33	0	4	4
	34	17	14	31
	35	4	4	8
	36	12	10	22
		52	**104**	**156**

Appendix N

Details Of Social And Quality Comments Made By The Peer Assessors

Groups	Student Number	Social comments (S)	Quality Comments (Q)		Quality comments (Q)	Total Comments (S + Q)
			Surface level	In-depth level		
Group 1 Anonymous More-accountable	1	6	11	0	11	17
	2	6	11	12	23	29
	3	8	13	6	19	27
	4	6	10	8	18	24
	5	4	15	0	15	19
	6	6	9	3	12	18
	7	1	4	9	13	14
	8	6	8	2	10	16
	9	5	7	4	11	16
		48	**88**	**44**	**132**	**180**
Group 2 Named More-accountable	10	4	8	6	14	18
	11	6	9	13	22	28
	12	0	21	8	29	29
	13	5	6	11	17	22
	14	6	10	28	38	44
	15	6	30	43	73	79
	16	9	9	18	27	36
	17	6	10	11	21	27
	18	5	15	1	16	21
		47	**118**	**139**	**257**	**304**
Group 3 Anonymous Less-accountable	19	8	2	0	2	10
	20	3	9	5	14	17
	21	9	11	35	46	55
	22	9	5	0	5	14
	23	12	7	2	9	21
	24	10	15	8	23	33
	25	6	0	0	0	6
	26	5	4	0	4	9
	27	4	0	0	0	4
		66	**53**	**50**	**103**	**169**
Group 4 Named Less-accountable	28	3	7	0	7	10
	29	15	8	4	12	27
	30	9	10	15	25	34
	31	13	2	0	2	15
	32	5	11	4	15	20
	33	11	4	0	4	15
	34	5	4	18	22	27
	35	6	13	4	17	23
	36	3	19	10	29	32
		70	**78**	**55**	**133**	**203**

Appendix O

Details Of Instructor-Assigned Marks On Students' Critique1 And 2

Groups	Student Number	Instructor-assigned mark	
		Critique (1)	Critique (2)
Group 1 Anonymous More-accountable	1	8.50	8.75
	2	8.00	8.00
	3	7.50	8.25
	4	8.00	8.00
	5	8.50	7.25
	6	8.50	8.25
	7	9.50	8.50
	8	7.50	8.00
	9	7.50	7.25
Group 2 Named More-accountable	10	8.00	7.00
	11	7.50	8.25
	12	8.50	8.75
	13	8.50	8.00
	14	9.50	9.75
	15	7.50	8.75
	16	7.50	8.25
	17	8.00	8.75
Group 3 Anonymous Less-accountable	18	7.50	8.25
	19	8.00	7.50
	20	9.00	8.75
	21	7.50	8.25
	22	8.50	7.75
	23	8.50	7.75
	24	8.50	7.50
	25	7.00	8.25
	26	7.50	7.25
	27	9.00	8.00
Group 4 Named Less-accountable	28	7.50	7.50
	29	8.00	7.50
	30	8.50	8.25
	31	7.50	7.50
	32	8.00	7.25
	33	7.00	8.25
	34	7.00	8.25
	35	7.50	8.25
	36	7.50	7.75