Chapter 20

Effects of Anonymity and Accountability During Online Peer Assessment

Gunita Wadhwa  
Memorial University, Canada

Henry Schulz  
Memorial University, Canada

Bruce L. Mann  
Memorial University, Canada

Learning Objectives

1. Describe the research question or problem in the introduction.

2. Summarize in your own words orally or in writing the gist of the literature review, including the theoretical framework to support the questions, problem, or need.

3. Describe the gist of the methodology, including:
   a. The participants [e.g., level, prerequisites, prior knowledge of the dependent variable under consideration, motivation to learn, intrinsic better job, etc.]
   b. The materials in depth [e.g., online, tools, software utilities, workbooks, writing materials, verbal instructions, etc.]
Abstract

A 2×2 experiment was conducted to determine the effects of anonymity (anonymous vs. named) and peer-accountability (more-accountable vs. less-accountable) on peer over-marking, and on the criticality and quality of peer comments during online peer assessment. Thirty-six graduate students in a Web-based education research methods course were required to critique two published research articles as a part of their course. Peer assessment was carried out on the first critique. Students were randomly assigned to one of the four groups. Peer assessors were randomly assigned three students’ critiques to assess. Peer assessors and the students being assessed were from the same group. Peer assessors assigned a numeric mark and commented on students’ critiques. The four main results were: First, significantly fewer peer assessors over-marked (i.e., assigned a higher mark relative to the instructor) in the anonymous group as compared to the named group (p < .04). Second, peer assessors in the anonymous group provided a significantly higher number of critical comments (i.e., weaknesses) as compared to the named group (p < .01). Third, peer assessors in the named group...
and the more-accountable group made a significantly higher number of quality comments (i.e., 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 (p < .01). Lastly, the students’ responses to the questionnaire indicated that they found the peer assessment process helpful. This study suggests that in online peer assessment, the anonymity and the degree of peer-accountability affect peer marking and comments.

Introduction

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). As an instructional method, peer assessment can help learners develop critical (Searby & Ewers, 1997), evaluative (Blumhof & Stallibrass, 1994), and analytical skills (Falchikov, 1995). Peer assessment has a history of 30 years of practice and has been widely used in higher education settings (Falchikov, 1995; Rada, 1998; Topping et al., 2000). With increasing interest in Web-based learning, online peer assessment is also gaining popularity (Topping, 1998).

Issues in Peer Assessment

Despite growing interest in peer assessment, concerns still remain regarding the reliability and validity of peer ratings. Some of the concerns 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; Slujisjams, 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; Slujisjams 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; Ormond, 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; Slujisjams 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, i.e., assign a higher mark 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 by Falchikov (1996), Searby and Ewers (1997), and Topping et al. (2000) showed 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 chapter 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.

### Variables

• **Anonymity:** The concept of anonymity has been studied in various settings and context, such as students’ response to teacher evaluation (e.g., McCollister 1985; Stone, Spool, & Robinowitz, 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; Connolly et al., 1990; Davis, 2000; Falchikov, 1995; Haaga, 1993; Jessup, Connolly, & Galegher, 1990; 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. Whereas, others (e.g., Hiltz et al., 1989; Valacich, Jessup, Dennis, & Nunamaker, 1992; Zhao, 1998) indicate that anonymity reduces responsibility resulting in more careless and less concerned responses. Yet, other studies (Ellis, 1984; George, Easton, Nunamaker Jr., & Northerchaft, 1990) found no difference in participants’ response due to anonymity.

Pinsonneault and Heppel (1998) found anonymity to interact with situational variables, such as, group unity. They further suggest that the effects of anonymity also depend on the context variables, such as, accountability cues, deindividuation, private self-
awareness, and attentional cues. This study examined the interaction of anonymity and peer-accountability on peer assessment.

• **Peer-accountability:** Some researchers (Topping et al., 2000; Zhao, 1998) suggest that incorporating peer-accountability in peer assessment may improve the quality of peer comments. Although there is no empirical evidence on the effect of peer-accountability on peer comments in online peer assessment, it has shown positive results in student response to teacher-evaluation questionnaires and small group interactions (Gordon & Stuecher, 1992; Price 1987). These studies further suggest that varying degrees of peer accountability may also affect the student response.

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 counter-arguments” (p. 75). Gordon and Stuecher (1992) examined the differences in students’ responses on teacher-evaluation questionnaires, 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 and identifiability had no impact when the group members were accountable for their decisions. In their study, with a 2×2 (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.

One way of incorporating peer accountability in peer assessment could be by 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. Tsai et al. (2001), in a study on networked peer assessment, reviewed and graded the quality of peer assessor’s comments to encourage assessors to provide helpful comments.

### Aims of the Study

The aim of this study was to determine the effects of anonymity and the degree of peer-accountability on peer assessment. The four research questions 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 more- or less-accountability 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?

**Method**

**Participants**

Thirty-six graduate students (22 females and 14 males, 26-55 years of age) enrolled in a Web-based education research methods course for spring semester 2003 agreed to participate in the experiment. 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 Present Study**

The research methods course was a compulsory course for all education graduate students at the university. The duration of the course was 16 weeks. The course was offered online through the 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. However, 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.

**Experimental Design**

A 2 (anonymity)× 2 (peer-accountability) factorial design was used. Nine students were randomly assigned to each of the four groups (i.e., anonymous, more-accountable group; named, more-accountable group; anonymous, less-accountable group; and named, less-accountable group). All groups received the same course assignments, which were assessed using the same procedure (see Appendix B).
**Anonymity** was defined as anonymous versus named. 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** was defined as more-accountable versus less-accountable. 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.

### Peer Assessment Process Followed in the Study

The peer assessment process followed in this experiment was adopted from previous studies (Falchikov, 1995, 1996; Mann, 2000; Sluisman 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. Figure 1 shows the online peer assessment process followed in this experiment.

*Figure 1. Peer assessment process in a WebCT course*
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. Appendix H shows the summary of WebCT tools used for peer assessment activities in the experiment. 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. 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 (see Table 1).

For the experiment, students were asked to critique two published education research articles that used quantitative methods (critiques 1 and 2). This constituted the requirement for the course, which contributed one-third (1/3) to the entire course. For critique 1, students were asked to critique the problem and methods section of the article. The instructor provided the criteria to critique (see Appendix B). After all critiques were submitted in WebCT, students were able to view other students' submissions by clicking on the “View Peer Assignment” option on the course homepage. Each student submission was identified with a unique student number assigned by the instructor. Appendix C shows a view of the “View Peer Assignment” page indicating each uploaded assignment with an instructor-assigned student number. The instructor randomly assigned three student numbers (student critiques) to each peer for assessment. The students and their peer assessors were from the same group. On clicking the instructor-assigned student number, students in the anonymous group were able to view critiques indicating only the student number. However, the students in the named group were able to view student critiques indicating the student’s name and number (see sample in Appendix D). Peer assessment involved assigning a numeric mark on a scale of 1 to 10, and providing comments on the assessed critiques. Peer assessors were asked to use their own criteria for assessing student critiques. The instructor assessed each student critique independent of the peers, also on a scale of 1 to 10.

Each student received three peer assessments (numeric mark and qualitative comments) as well as the instructor’s mark on their critique. The peers’ assessments on each student critique were compiled and e-mailed to each individual student. The students in the anonymous group received anonymous peer assessments indicating the peer assessor’s

<table>
<thead>
<tr>
<th>Week</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Students are given the first article to critique (critique 1).</td>
</tr>
<tr>
<td>3</td>
<td>Students submitted their first critique online using the assignment tool in WebCT. Student submissions were uploaded for other students to view and peer assessment.</td>
</tr>
<tr>
<td>4</td>
<td>The peers submitted their assessments in WebCT using the Survey tools.</td>
</tr>
<tr>
<td>5</td>
<td>The researcher compiled the peer assessments and e-mailed them to each individual student.</td>
</tr>
<tr>
<td>6</td>
<td>Students are given the second article to critique.</td>
</tr>
</tbody>
</table>

Table 1. The schedule of activities followed in the experiment
number. The students in the named group received peer assessments indicating the peer assessor’s name and number. Appendix E show a sample of compiled peer assessment (peer assigned marks and qualitative comments) sent to a student in the named group.

Next, students were given a second article to critique (critique 2) and were asked to critique four components: problem, method, results, and conclusions. The instructor provided the criteria to critique (see Appendix B). Only the instructor assessed critique 2. The instructor’s assessment involved only assigning a numeric mark on a scale of 1 to 20. Table 1 shows the schedule of activities followed in the experiment.

**Procedure for Data Analysis**

- **Peer over-marking**: Peer over-marking was operationalized as the peer assessors assigning a higher mark relative to the instructor. To determine peer over-marking, the peer-assigned mark (average of the three peers’ marks) on student’s critique was compared with the instructor-assigned mark on the same critique. It was expected that the number of peer over-markings would be higher in the named group compared to the anonymous group.

- **Critical peer comments**: Each written statement (comment) provided by the peer assessors on other students’ critiques was analyzed as being positive or critical. 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. Appendix F shows examples of the “positive” and “critical 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). 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.

- **Quality of peer comments**: Each written statement (comment) provided by the peer assessors on other students’ critiques was categorized as either a social or quality comment. 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 weaknesses 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. Appendix G 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, Bonk, and Angeli (2000). Quality of comments was the sum of surface level and in-depth level cognitive comments made by the peer assessors. Since the students were asked to critique only those two components of the research article, comments only related to the problem and the method sections of the critique were considered. Each comment (statement) was analyzed for the level of processing, as done by Hara et al. (2000).

The number of quality comments made by the peer assessors in the more-accountable group and the less-accountable group were counted and compared. 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.

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

### Results

#### Peer Over-Marking

The results from a 2×2 chi-square test indicated that the relationship between anonymity and peer over-marking (peer-assigned marks higher than the instructor’s mark) was statistically significant, $\chi^2(1, N=36) = 4.050, p = .044$, with a medium effect size, $W = 0.335$. As hypothesized, the number of peer over-marking was more in the named group (13 of 18, i.e., 72%) compared to the anonymous group (7 of 18, i.e., 39%). Table 2 shows the number of peers who over-marked, under-marked and assigned an identical-mark relative to the instructor.

There was no statistically significant difference in *under-marking* ($p = .070$), the number of peer-assigned marks lower than the instructor’s mark, or *identical-marking* ($p = .630$), the number of peer-assigned marks identical to the instructor’s mark. 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 no statistically significant effect of peer accountability on the number of peers who over-marked ($p = .502$), under-marked ($p = .717$) and assigned an identical-mark ($p = .148$).
Critical Peer Comments

The results from a $2 \times 2$ chi-square test indicated that the relationship between anonymity and the number of critical comments made by the peer assessors was statistically significant, $\chi^2 (1, N = 767) = 32.368$, $p = .000$, 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 3 shows the number of critical and positive comments made by the peer assessors. The relationship between anonymity and the number of total comments (sum of critical and positive comments) made by the peer assessors was also statistically significant, $p = .000$. 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 and positive comments made by the peer assessors was not statistically significant, $p = .485$. The number of critical comments made by the peer assessors in the more-accountable group and less-accountable group were 195 and 147, respectively. The number of positive comments made by the peer assessors in the more-accountable group and less-accountable group were 253 and 172, respectively.
Table 3. Number of critical and positive comments made by the peers assessors (Note. A critical comment was a weakness indicated by a peer assessor. A positive comment was a strength indicated by a peer assessor.)

<table>
<thead>
<tr>
<th>Anonymity groups</th>
<th>Peer accountable Groups</th>
<th>Critical</th>
<th>Positive</th>
<th>Total comments (critical and positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>More-Accountable</td>
<td>10.00</td>
<td>5.10</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>10.28</td>
<td>8.09</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(56%)</td>
<td>(44%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>More-Accountable</td>
<td>11.67</td>
<td>9.30</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Less-Anonymous</td>
<td>5.78</td>
<td>5.67</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>8.72</td>
<td>8.06</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36%</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>9.50</td>
<td>8.00</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45%)</td>
<td>(55%)</td>
<td></td>
</tr>
</tbody>
</table>
Quality of Peer Comments

The results from a $2 \times 2$ chi-square test indicated that the relationship between peer accountability and the quality of peer comments was significant, $\chi^2(1, N = 856) = 32.566$, $p = .000$, 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 shows the number of quality comments and social comments made by the peer assessors. The relationship between peer accountability and total comments made by the peer assessors was statistically significant, $\chi^2(1, N = 856) = 14.650$, $p = .000$, 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, $\chi^2(1, N = 856) = 9.478$, $p = .002$, 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). The relationship between anonymity and total number of comments (sum of quality comments and social comments) made by the peer assessors was statistically significant, $p = .000$. The peer assessors in the named group made significantly higher number of comments, $n = 507$, compared to the peer assessors in the anonymous group, $n = 349$.

The results of the effect of peer-accountability and anonymity on the number of surface-level and in-depth level cognitive comments made by the peer assessors was also tested. 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, $\chi^2(1, N = 625) = 0.572$, $p = .450$, 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, $\chi^2(1, N = 625) = 7.967$, $p = .005$, 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, $n = 141$. Table 5 shows the number of surface level and in-depth level cognitive comments made by the peer assessors.

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
Table 4. Number of quality comments and social comments made by the peer assessors (Note. *Quality comments were cognitive statements made by the peer assessors indicating strengths and weakness along with reasoned responses and suggestions for improvement. *Social comments were general statements not related to a specific content of subject matter. n indicates the number of comments in each category.)

<table>
<thead>
<tr>
<th>Peer accountable groups</th>
<th>Anonymity groups</th>
<th>Quality comments</th>
<th>Social comments</th>
<th>Total comments (quality + social)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anonymous</td>
<td>M SD n</td>
<td>M SD n</td>
<td>M SD n</td>
</tr>
<tr>
<td>More-accountable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anonymous</td>
<td>14.67 4.44 132</td>
<td>5.33 1.94 48</td>
<td>10.00 5.84 180</td>
<td></td>
</tr>
<tr>
<td>Named</td>
<td>28.56 18.27 257</td>
<td>5.22 2.39 47</td>
<td>16.89 17.43 394</td>
<td></td>
</tr>
<tr>
<td>Less-accountable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anonymous</td>
<td>11.44 14.95 103</td>
<td>7.33 3.00 66</td>
<td>9.39 10.67 169</td>
<td></td>
</tr>
<tr>
<td>Named</td>
<td>14.78 9.43 133</td>
<td>7.78 4.41 70</td>
<td>11.28 8.00 203</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17.56 14.04 625</td>
<td>6.42 3.17 231</td>
<td>9.34 6.85 856</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
Table 5. Number of surface level and in-depth level cognitive comments made by the peer assessors (Note. Surface level comments were statements indicating the strengths and weaknesses in students' work without any suggestion, justification and elaboration. In-depth level comments were statements indicating the strengths and weaknesses in the student's work that contained supporting arguments, suggestions, and reasoned responses.)

<table>
<thead>
<tr>
<th>Peer accountable groups</th>
<th>Anonymity groups</th>
<th>Surface level&lt;sup&gt;a&lt;/sup&gt;</th>
<th>In-depth level&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Total quality comments (surface + in-depth)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>Named</td>
<td>13.11</td>
<td>7.75</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.45</td>
<td>6.02</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(53%)</td>
</tr>
<tr>
<td>Less-accountable</td>
<td>Anonymous</td>
<td>5.89</td>
<td>5.11</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Named</td>
<td>8.67</td>
<td>5.29</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.28</td>
<td>5.24</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(56%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>9.37</td>
<td>5.95</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(54%)</td>
</tr>
</tbody>
</table>
The results indicated that there was no statistically significant interaction of anonymity and peer accountability on the difference in students’ performance: $F(1,32)=0.000, p=1.00, \eta^2=0.00$. This finding does not support the hypothesis. Table 6 shows the summary of repeated measures ANOVA with within-subjects effects of anonymity and peer accountability.

Further, there was no statistically significant, $F(1)=0.336, p=.566, \eta^2=.010$, effect of peer accountability on the difference in students’ performance from critique 1 to critique 2.
However, there was a statistically significant, \( F(1) = 4.360, p = .045, \eta^2 = .120 \), effect of anonymity on the difference in students’ performance from critique 1 to critique 2. 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 (\( M = 7.96, SD = 0.47 \)). Figure 2 shows means’ of the students’ performance from critique 1 to critique 2 in the anonymity groups.

Finally, there was no statistically significant difference, \( F(1) = 0.121, p = .730, \eta^2 = .004 \), in the instructor-assigned students’ marks from critique 1 (\( M = 8.00, SD = 0.65 \)) to critique 2 (\( M = 8.04, SD = 0.57 \)). Table 7 shows the means of the instructor-assigned marks on the students’ critique 1 and critique 2.

**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 8 shows the students’ response on learning benefits derived from peer assessment process followed in this study.

Table 9 shows the students’ views on the peer assessment process followed in this study.
Summary of the Results

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

1. As predicted, the number of peer over-marking (i.e., peer-assigned a mark higher relative to the instructor) was greater in the named group (13 of 18, i.e., 72%) compared to the anonymous group (7 of 18, i.e., 39%).

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.

## Conclusion

This chapter 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 the conclusions drawn are:

First, anonymous online peer assessment reduced the number of peer over-marking, with a medium effect size. Notably, there was no significant effect of anonymity (anonymous and named) on the number of peer assessors under-marking and assigning identical marks. There was no significant relationship between peer accountability and the number of peer assessors over-marking, under-marking, and assigning identical marks.

Second, anonymous online peer assessment enhanced critical comments in peer feedback. However, the effect size was small. Further, 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.

Third, in online peer assessment two variables, namely peer accountability and anonymity, significantly affect the quality of peer comments. The results in this study showed that within 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), although the effect size was small. It was interesting to note that within 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.

Fourth, there was no significant improvement in students’ performance from critique 1 to critique 2 (see Table 7). 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 2). The improvement in the students’ performance in the named group may 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). 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.

According to Webb (1995), there is some evidence in the literature to indicate that the level of elaboration of comments is related to achievement. Webb’s (1989) extensive review of the empirical studies on peer interactions in small groups 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 is required.

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 of 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, Howe, Soden, Halliday, & Low, 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:** The data on the students response to the questionnaire showed that the students found the peer assessment exercise beneficial (see Table 8). 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 9).

---

**Limitations of the Study**

The first limitation of this study was that no information was obtained on 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. There were two reasons for the absence of the instructor’s qualitative comments on the students’ critiques. First, literature 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). Therefore, in the presence of the instructor’s feedback, students may not take peer comments seriously. Second, it was opined that when both the instructor and the peers provided qualitative feedback, it

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
would be 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 and 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; Slujsmans 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.

Summary

This 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 and 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 chapter. It is hoped that these findings provide some direction for researchers and educators about peer assessment in an online learning environment.
References


Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.


### 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**

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

A view of the “View Peer Assignment” page indicating each uploaded assignment with an instructor-assigned student number

Table D1. A sample of uploaded student critique in the anonymous group

<table>
<thead>
<tr>
<th>Assignment 1 By Student (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>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. .............</td>
</tr>
</tbody>
</table>

Appendix D

A sample of uploaded student critique

Table D1. A sample of uploaded student critique in the anonymous group

<table>
<thead>
<tr>
<th>Assignment 1 By Student (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>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. .............</td>
</tr>
</tbody>
</table>
Appendix E

A sample of compiled peer assessment (peer-assigned marks and qualitative comments) sent to a student in the named group (Fictitious names)

Student (10) – John Doe

How would you grade (1/10) John Doe’s assignment? Please provide comments

<table>
<thead>
<tr>
<th>User ID</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jane Doe</td>
</tr>
<tr>
<td>6</td>
<td>Leslie</td>
</tr>
<tr>
<td>9</td>
<td>Beverely</td>
</tr>
</tbody>
</table>
Appendix F

An example of “positive” and “critical” comments provided by the peer assessors on the assessed critique

<table>
<thead>
<tr>
<th>Component</th>
<th>Sub-Component</th>
<th>Positive Comment</th>
<th>Critical Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
<td>Problem Statement &amp; Research Question</td>
<td>“The problem statement was clearly defined.”</td>
<td>“The statement of purpose is vaguely stated.”</td>
</tr>
<tr>
<td></td>
<td>Literature Review</td>
<td>“I felt you gave a good summary of the literature review.”</td>
<td>“Your introduction should have focused on the literature review.”</td>
</tr>
<tr>
<td></td>
<td>Hypothesis</td>
<td>“You also gave a very thorough overview of the hypothesis.”</td>
<td>“You should to discuss whether the hypothesis was in testable form.”</td>
</tr>
<tr>
<td>Method</td>
<td>Sampling Procedure &amp; Subjects characteristics</td>
<td>“You raised a valid point when discussing the methodology of how participants were selected.”</td>
<td>“The information on selection of the participants was provided but you mentioned otherwise.”</td>
</tr>
<tr>
<td></td>
<td>Variables</td>
<td>“The choice and explanation of your variables were accurate and well written.”</td>
<td>“Control or intervening variables were not identified.”</td>
</tr>
<tr>
<td></td>
<td>Validity – Internal</td>
<td>“Issues concerning the internal validity were covered in detail.”</td>
<td>“In the ‘Internal Validity’ section, I differ with you on maturation.”</td>
</tr>
<tr>
<td></td>
<td>Validity - External</td>
<td>“Very good points on threats to external validity.”</td>
<td>“There was no mention made to external validity.”</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
<td>“You raised some valid points concerning the operations trainer and reliability.”</td>
<td>“Analysis on the reliability of these measures would help to strength the critique.”</td>
</tr>
</tbody>
</table>
Appendix G

Examples of social and cognitive (surface and in-depth) comments made by the peer assessors on a student’s critique.

<table>
<thead>
<tr>
<th>Category of Comments</th>
<th>Definition</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social</strong></td>
<td>Statements not related to formal content of subject matter.</td>
<td>“The assignment I printed had some spelling errors and unfinished sentences. Not sure if it was my printer or the paper. However I know quality is better than quantity.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“The student has shown clear understanding of the concepts. She has raised many points from the text. Good job.”</td>
</tr>
<tr>
<td><strong>Quality (Cognitive - Surface level processing)</strong></td>
<td>Statements indicating the strengths and weaknesses in students’ work.</td>
<td>“Independent and dependent variables were well defined. However there was no mention of other variables.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“The purpose of the experiment, type of hypothesis and the variables were well defined. The section on research design and validity needs elaboration.”</td>
</tr>
<tr>
<td><strong>Quality (Cognitive In-depth level processing)</strong></td>
<td>Statements indicating the strengths and weaknesses in the student’s work that contained supporting arguments, suggestions for improvement, and reasoned responses.</td>
<td>“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.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“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.”</td>
</tr>
</tbody>
</table>
### Appendix H

<table>
<thead>
<tr>
<th>Activity and Purpose</th>
<th>WebCT Tool Used</th>
<th>Category</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>WebCT Tools Used</td>
<td>Course Content Tools</td>
<td>Options</td>
</tr>
<tr>
<td>Submit</td>
<td>Students submit their critique as per instructions</td>
<td>Content Module</td>
<td>Option</td>
</tr>
<tr>
<td>View</td>
<td>Students view other students critiques</td>
<td>Course Content Tools and Pages</td>
<td>Option</td>
</tr>
<tr>
<td>Assess</td>
<td>Students assess other critiques and submit their assessments</td>
<td>Evaluation and Activity Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Feedback</td>
<td>The instructor compiles peers' critiques and sends it to the individual student</td>
<td>Evaluation and Activity Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Collect student response on the peer assessment process followed in this study</td>
<td>Evaluation and Activity Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Grades</td>
<td>The instructor assesses each student critique and updates students grades</td>
<td>Student Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Course Mail</td>
<td>Mail</td>
<td>Communication Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Discussions</td>
<td>Discussions</td>
<td>Communication Tools</td>
<td>Option</td>
</tr>
<tr>
<td>My Grades</td>
<td>My Grades</td>
<td>Student Tools</td>
<td>Option</td>
</tr>
<tr>
<td>Quizzes</td>
<td>Quizzes</td>
<td>Evaluation and Activity Tools</td>
<td>Option</td>
</tr>
</tbody>
</table>