Effective Use of Peer Assessment in a Graduate Level Writing Assignment: A Case Study

Ashley Landry[1], Shoshanah Jacobs[2] & Genevieve Newton[3]

[1] University of Guelph Humber, Canada

[2] Assistant Professor, Department of Integrative Biology, University of Guelph, Canada

[3] Assistant Professor, Department of Human Health & Nutritional Sciences, University of Guelph, Canada

Correspondence: Genevieve Newton, Assistant Professor, Department of Human Health & Nutritional Sciences, University of Guelph, Canada. E-mail: newton@uoguelph.ca

Abstract

At the undergraduate level, considerable evidence exists to support the use of peer assessment, but there is less research at the graduate level. In the present study, we investigated student perception of the peer assessment experience and the ability of graduate students to provide feedback that is comparable to the instructor and that is consistent between assessors on a written assignment. We observed that students were very supportive of the activity and that negative concerns related to inconsistent peer evaluations were not supported by the quantitative findings. Our quantitative analyses showed that the average grade of the student reviews was not significantly different from the average grade given by the instructor, although student reviewer reliability was not high. Students showed a significant grade improvement following revision subsequent to peer assessment, with lower graded papers showing the greatest improvement; greater grade change was also associated with an increased number of comments for which a clear revision activity could be taken. Design of the graduate peer assessment activity included several characteristics that have been previously shown to support positive findings, such as training, use of a clear assignment rubric, promotion of a trusting environment, use of peer and instructor grading, provision of directive and non-directive feedback, recruitment of positive comments, and use of more than one peer assessor. This study, therefore, builds on previous work and suggests that use of a carefully designed peer assessment activity, which includes clear direction regarding actionable comments, may provide students with useful feedback that improves their performance on a writing assignment.

Keywords: Peer assessment, Peer review, Assessment, Evaluation

1. Introduction

Peer assessment is a practice being used with increasing frequency in higher education. The process of peer assessment allows students to provide and receive feedback on their work with minimal time investment by the instructor, making it an attractive tool for use in today's era of ever increasing class sizes and faculty workloads. However, it is the advantages to the student that make peer assessment most appealing. For example, students receive formative feedback that is intended to enhance learning and improve academic performances (Gielen, Peeters, Dochy, Onghena & Struyven, 2010). As well, peer assessment promotes active involvement by students and allows them to engage with assignment rubrics that will subsequently be used by the instructor for grading (Van Gennip, Segers & Tillema, 2010; Cho & MacArthur, 2010). Consequently, peer assessment is being used across disciplines and with many types of assignments. Most commonly, though, peer assessment is used with written assignments, wherein students provide feedback - which may be formative only or both formative and summative - on such elements as clarity, organization, strength of argumentation, and grammar. Typically, students are given the opportunity to revise their assignment after the receipt of feedback and prior to submitting the assignment for instructor grading. In this way, it is intended that students will improve the quality of their final draft and learn as they go through the assessment and revision process.

Previous research shows the value of review to improve self-evaluation and better understanding of concepts being studied. Typically an instructor provides feedback on the final version of an assignment and students are rarely able to revise/improve their work before being graded. This is often an inevitability of time constraints – it simply is not

feasible for an instructor to assess and give feedback on multiple versions of an assignment to an entire class, despite the obvious educational benefits. In contrast, peer assessment allows students to give and receive feedback on an assignment and to revise the draft version prior to final submission. This is advantageous since research has shown that students learn by collaborating with others and having insight into their peers' ideas and opinions (Van Gennip et al., 2010). As well, students report that they experience a deeper understanding of the course material by examining the perspectives of their peers (Guilford, 2001). Performance is enhanced further when assessment procedures include feedback and opportunities for revision (Gibbs & Simpson, 2004). However, while peer assessment as an overall process yields positive academic improvements (Gielen et al., 2010), the type of feedback being given during the process can have an important and direct effect on performance outcomes. For example, Cho & MacArthur (2010) showed that non-directive feedback, which is feedback that refrains from focusing on specific errors but rather makes general improvement suggestions, was the most likely to yield the greatest revision improvements, while Gibbs & Simpson (2004) found that in order to confer a positive influence on learning, feedback must be sufficient in both frequency and detail. Thus, it is clearly relevant to consider the type of feedback being provided as well as the pedagogical impact of the receipt of feedback.

Since peer assessment is entirely student focused, the process relies heavily on student attitudes and perceptions, which have not been found to be consistently positive. For example, Venables & Summit (2003) compared students' attitudes toward peer assessment before and after completing a peer assessment project. The majority of students responded that they disliked or had reservations about other students assessing their work at the beginning of the study, and only 26% of students changed their comments from negative to positive following the exercise. Most students expressed that they had experienced considerable learning and that seeing their peers' perspectives was beneficial, yet this did not change their opinions about peer assessment as a whole. In contrast, a recent study by Mulder, Pierce & Baik (2014) showed a negative shift in student perception towards peer assessment following the experience in a cross-disciplinary investigation. In general, the influence of student attitudes and perceptions on learning is unclear; it has been observed that skepticism towards peer's ability to accurately criticize and give feedback can inhibit learning through the process of peer assessment (Bangert-Drowns, Kulik, Kulik & Morgan, 1991), although it has also been found that when students expressed negative perceptions of the peer-assessment process or were skeptical about the validity of the process, it did not affect their revision performance (Kaufman & Schunn, 2010). Consideration of students' attitudes and perceptions is therefore an important part of the evaluation of a peer assessment program.

Several studies have looked at the ability of students to act as accurate and reliable peer assessors. Research has shown that multiple peer assessments are the most accurate and yield evaluations that are on average, similar to the instructor (Marcoulides & Simkin, 1995), while also yielding more feedback (Cho, Schunn & Wilson, 2006). Having inconsistent evaluations can be bad for numerous reasons, most notably that students cannot learn from their mistakes if there is no clear indication or consensus as to what those mistakes were and how they can be corrected (Cho et al., 2006). In addition to accuracy, it is also important to consider whether evaluations are consistent, or reliable, between and among student assessors. Marcoulides & Simkin, (1995) showed that peer evaluation in an undergraduate writing exercise did not show a large margin of deviation, and that while students were not always consistent with the specificity of the errors they noted; these discrepancies were not enough to affect the overall grades of the paper. Topping, Smith, Swanson & Elliot (2000) reported similar reliability at the graduate level, although there is limited research regarding the reliability and accuracy of graduate students as peer assessors.

As described by Roger Graves (2013) in a recent article in University Affairs, many students across levels of higher education have difficulty with writing; however, Graves also states that student writing improves when students are given the opportunity to revise their work and when a rubric is used, as is the usual process during peer assessment. So, continued exploration of the use of peer assessment in writing is highly warranted and relevant, particularly at the graduate level, wherein most students are required to publicly disseminate their research findings as part of their academic program. Given the aforementioned benefits of peer assessment, and the relative paucity of research regarding the use of peer assessment at the graduate level, the aims of the present study are: (1) to examine the subjective student experience of peer assessment in a graduate writing exercise, (2) to determine whether graduate students assign grades that are accurate and reliable, and (3) to determine whether the use of peer assessment results in improvement in assignment quality, specifically as associated with actionable comments provided by student assessors. By answering these research questions, we will be addressing the gaps in the graduate level peer assessment literature.

2. Materials and Methods

2.1 Subjects

The subjects in this study were graduate students enrolled in a nutrition class in Fall 2012 and Fall 2013. Students could be at the MSc or PhD level and in any year of their program, but over 90% were MSc students in their first year of studies. All students in the course (n=37 Fall 2012, n=33 Fall 2013) were invited to participate in the study. In, total 44 students agreed to participate in the study through completion of the online survey. However, 10 students did not provide their name on the survey, so analysis of written papers could only be performed for 34 students. Participation in the study was voluntary, and no compensation or incentives were provided. This study was approved by the institutional Research Ethics Board and subjects provided informed consent to participate.

2.2 Assignment

The assignment was to write a critical assessment of a focused cluster of five to eight primary research publications, making clear both the original contribution represented by each piece of work discussed and the way in which the studies complement (and possibly build on) one another. The assignment was highly critical, requiring students to integrate the research findings and to discuss the strengths and the weaknesses of the various studies, with careful consideration of the methodological differences between studies.   The paper was a minimum of 2250 words and a maximum of 3250 words. The assignment rubric used with the assignment is presented in Table 1.

Table 1. Assignment rubric used by students and instructor to evaluate both the draft and final versions of the assignment

Rubric Categories	Marks
Introduction	Total /10
1) Is the main research question clearly defined, with the topic sufficiently focused to be covered by the scope of the paper?	/ 5
2) Are the papers that compromise the cluster of publications introduced and briefly described, with an outline of the contents to follow provided?	/ 5
Critique	Total /50
3) Is the original contribution represented by each piece of work discussed?	/ 10
4) Is it clear how the studies complement (and possibly build on) one another?	/ 10
5) Are there frequent associations made to show relationships between studies?	/ 10
6) Are the strengths and the weaknesses of the various studies identified?	/ 10
7) Does the author incorporate details, facts and other supporting evidence appropriately?	/ 10
Conclusion	Total /10
8) Does the student provide a brief summary or concluding remarks at the end of the paper?	/ 5
9) Are there suggestions for future research?	/ 5
Grammar and Organization	Total /20
10) Are there any grammar, spelling punctuation, etc. mistakes?	/ 5
11) Is the paper well organized and does it follow a logical train of thought?	/ 5
12) Is the writing concise? Are the sentences short and to the point, or long and convoluted?	/ 5
13) Is the language appropriate? Does the author use appropriate scientific and academic terminology? Were the important terms appropriately defined?	/ 5
References	Total /10
14) Do the selected references comprise a logical cluster of publications?	/ 5
15) Did the student reference all relevant citations, and use the appropriate citation style, throughout the paper (APA)?	/ 5

2.3 Peer Assessment

The PEAR (Peer Evaluation Assessment and Review) software was used as the peer assessment platform. PEAR manages each stage of the submission process, anonymously distributes papers to students for evaluation, maintains rubrics, and tracks activity. Each student submitted the first draft of their critical research assessment to the PEAR site and anonymously received and provided feedback to/from two of their peers. Students completed the quantitative rubric for the assignment (Table 1), assigning a grade /100, and provided qualitative feedback in the

form of responses to the following three categories: Commendations (what was done well), Recommendations (what could be improved), and Corrections (what was done incorrectly based on the assignment criteria). Students were then given a three-week period to make revisions to their papers and submit to the PEAR system for instructor grading. The grade assigned by the peer assessors contributed 2% each to the final course grade. In response to the assessments received, each student also provided feedback on the quality of the assessment provided, which contributed 1% each to the final course grade. Therefore, students received both quantitative (in the form of a grade) and qualitative (in the form of comments in each category) feedback, and there was a grade incentive to submit a high quality first draft as well as a high quality peer evaluation. The peer assessment activity occurred in week nine of the twelve-week semester, and it was the primary educational activity during that week. Students were oriented to the process of peer assessment with a three-hour training workshop in week three, in which they worked in small groups to use the assignment rubric to evaluate two anonymous sample assignments. This was co-facilitated by the instructor and a writing librarian, and was accompanied by a presentation regarding effective writing strategies that emphasized the critical nature of the course writing assignment.

2.4 Surveys

After submission of the final course grades, all students in the Fall 2012 and Fall 2013 classes were invited to complete an online survey that asked them questions about their experience with peer assessment. The survey also included an informed consent form that asked students for permission to analyze their writing samples and course performance. Out of 70 students enrolled in the courses, 44 students completed the online survey. The survey was used to measure the subjective experience of graduate students with the peer assessment activity. The survey included 24 questions that were ranked on a five point Likert scale and nine text response questions that asked about personal characteristics and open-ended opinions about the peer assessment experience.

2.5 Data Analysis

To determine the subjective experience of students with peer assessment, the percentage of total respondents in each point was determined for the Likert scaled questions. The open-ended opinion responses were analyzed for common themes, and the major disadvantages and advantages of the experience were categorized.

To determine the accuracy of grades provided by peer assessors, which compares grades from students to grades from professor, the average grade of each student assessment among the sections was compared to the professor grade on the draft version of the paper using an independent sample unpaired t-test. We performed a linear regression analysis on the instructor grade vs the average of the two student-assessors to derive the correlation coefficient with bootstrap analysis to determine the measure of accuracy within our sample estimates.   To address the question of reliability among reviewers, we conducted a similar analysis on the two reviews provided by students. A t-test was used to determine whether the two reviews were in agreement. A linear regression was performed to derive the correlation coefficient and a bootstap analysis to determine the accuracy of the estimator. The accuracy and reliability was determined for the overall grades exclusively because previous research has shown that a reviewer's assessment of each part of a rubric tend to agree more so than different reviews of that same part (Haaga, 1993). As well, the ranking of papers from lowest to highest was determined for the professor and for the assessors (with each paper as the average of the two assessors), as it is possible for grades to be different with similar rankings.

To determine whether the use of peer assessment results in improvement in assignment quality, the instructor grades on the draft version of the paper were compared to the final grades using a paired t-test. This was done for the overall grade as well as the grades for each section of the assignment. To determine whether the receipt of feedback related to specific elements of the writing assignment result in improvements to those elements in the revision stage, a comparison of the first and final drafts of the paper was made. The comments related to recommendations and corrections in each rubric were carefully considered and numbered (combining both categories), with each comment that could result in a change identified as a 'actionable comment'. For this evaluation, the feedback from both peers was combined since the final draft was influenced by both assessments. Very general feedback, such as "be more critical", was not counted as an actionable comment, since a direct revision action was unclear. Following the assignment of the number of actionable comments to each paper, a linear regression was performed to investigate the relationship between the number of actionable comments and the change in instructor assigned grade from the draft to the final version of the assignment. A linear regression was also performed between the change in grade and the instructor assigned draft grade, to determine the relationship between grade change and performance on the initial submission.

All statistical analyses were performed using SPSS version 21. The significance level was set at $p \leq 0.05$.

3. Results

3.1 Subjective student experience of peer assessment

The reported subjective experience of students with the peer assessment process is presented in Figure 1. The number of total survey respondents was 44; however, there were several questions that not all respondents answered, resulting in a range of n=42-44. The open-ended survey questions that were qualitatively analyzed for themes regarding common experiences with peer assessment are presented in Table 2 (positive) and Table 3 (negative).
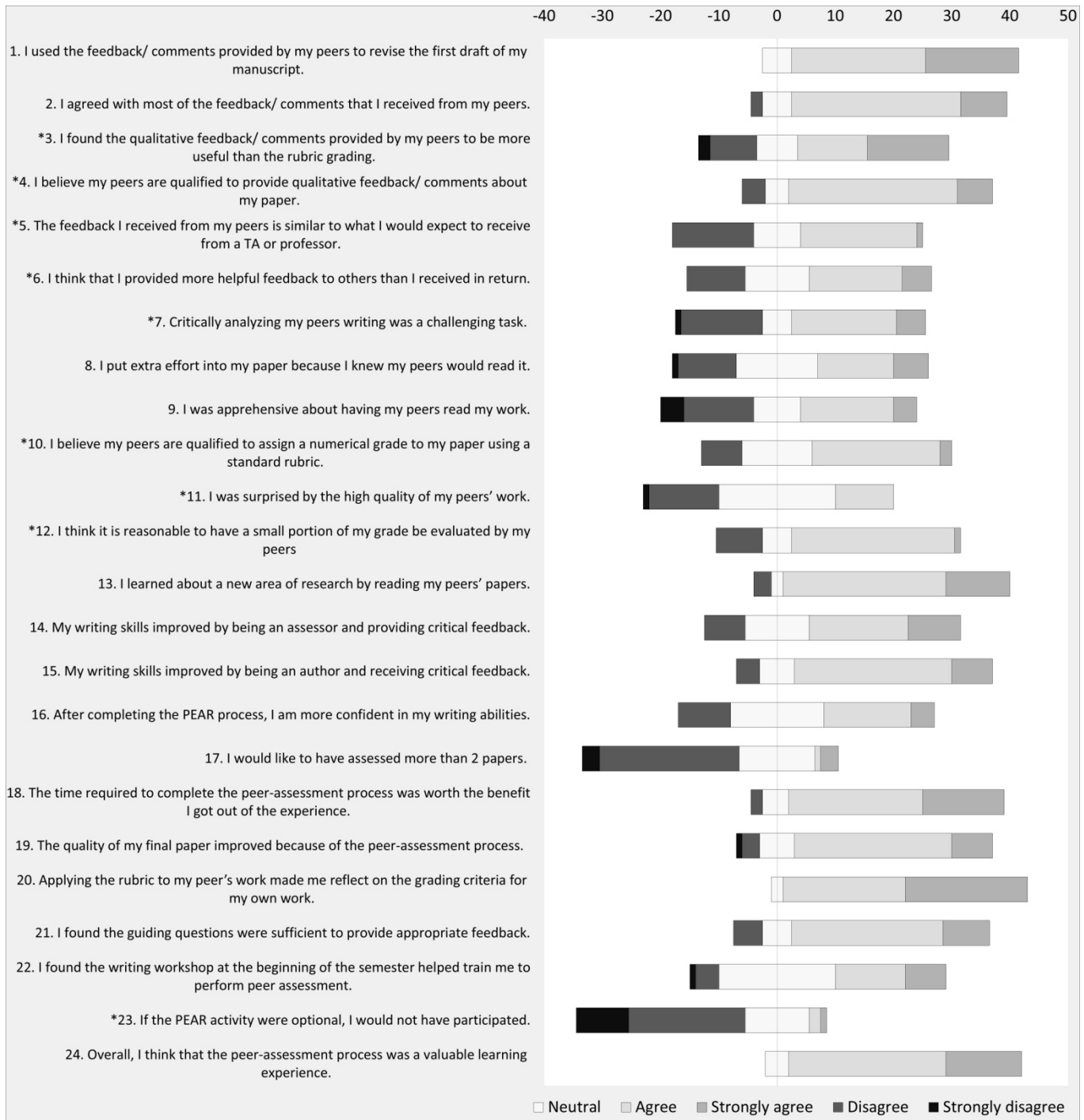

Figure 1. Distribution of responses (frequency) to survey questions regarding student perception of the peer assessment experience

* Indicates that all survey respondents did not answer the question, with a response rate between 42 and 43 students.

Table 2. Categories of positive qualitative feedback from surveys with examples of student comments

Category of Positive Feedback	Examples of Student Comments
Ability to improve writing before submitting the final draft.	- "It was great to get feedback from peers and having time to do edits before the final paper was due."
Learning about other scientific topics	- "I enjoyed reading the papers, learning new areas of research and gaining skills in critically analyzing and evaluating research."
	- "Using PEAR improved my writing skills, and allowed me to explore new research areas."
Time management	- "I feel that it was good to have a deadline for the rough draft as well to encourage your to start early."
	- "Having to submit a rough draft early for peer assessment was helpful in terms of getting started on the paper early."
	- "The specific deadlines for each aspect of the term paper although not directly related to PEAR was also helpful for keeping me on track."
Improving critical thinking	- "Being able to assessment someone else's paper was a major advantage for being able to think critically about your own paper."
	- "I enjoyed reading the papers, learning new areas of research and gaining skills in critically analyzing and evaluating research."
Self-evaluation	- "Using PEAR allowed me to see the caliber of writing of other students in the class. As a result, I was able to critically assess the quality of my own writing."
	- "Assessing other papers made me realize errors or insufficiencies in my own work and the peer comments were mostly helpful."

Table 3. Categories of negative qualitative feedback from surveys with examples of student comments

Category of Negative Feedback	Examples of Student Comments
Technology issues	- "[program] screens and displays could be more clear, it is necessary to click several times within a tab to get to the details, it would be better to have all in one main table displayed as to get a whole view at once. System takes too long to load."
	- "Sometimes [the program] runs slow."
Difficulty interpreting quantitative feedback	- "Questions are reasonable but it would be nicer to have space for added notes/comments for each question, to make the assessment more detailed. This would help the author understand exactly why a 4/5 was given over a specific entry."
	- "The quantitative portion gives the author feedback however I found it difficult to judge what numbers to give individuals, so I think that a larger portion of the marks (both for the author and assessor) should be devoted to the qualitative remarks given verse the quantitative."
Peers too harsh	- " I found that my peers were tougher markers than a professor or TA. Plus, the qualitative assessments were not helpful."
Not equal efforts	- "The downfall to the system is that while one person may put a lot of time and effort into their assessment, they may not receive as high quality of a assessment as they submitted."
	- " The major disadvantage was putting a lot of effort into someone else's assessment and not receiving comments about your own with equal effort."

3.2 Accuracy and reliability of student assessments

The overall accuracy of student assessments was determined by comparing the mean grades assigned by students relative to the instructor on the draft version of the assignment using an independent sample unpaired t-test on the

instructor grade vs. the average of the reviewer grades. None of the means associated with any of the assignment sections or the overall grades were significantly different (Table 4).

Table 4. Comparison between student and instructor assigned grades on individual components of the draft version of the assignment. Data are presented as mean +/- SEM.

Assignment Component	Instructor Grade	Average Student Reviewer Grade	T stat (df=66)	p value
Introduction (/10)	8.41 ± 0.26	8.25 ± 0.18	-0.51	0.61
Critique (/50)	41.62 ± 0.43	40.59 ± 0.68	-1.28	0.20
Conclusion (/10)	8.76 ± 0.17	8.41 ±0.15	-1.54	0.13
Grammar and Organization (/20)	15.89 ± 0.30	16.07 ± 0.31	0.43	0.67
References (/10)	9.21 ± 0.16	8.76 ± 0.18	-1.82	0.07
Overall (/100)	84 (±0.9)	82 (±1.1)	-1.44	0.16

The regression analysis on the average grade assigned by the student reviewers vs the instructor assigned grades had a correlation coefficient of r=0.72, and the coefficient was significant with bootstrapping (p=0.001), indicating that the instructor grade is a good predictor of the average of the reviewer grades and is therefore, a significant model. (Figure 2).
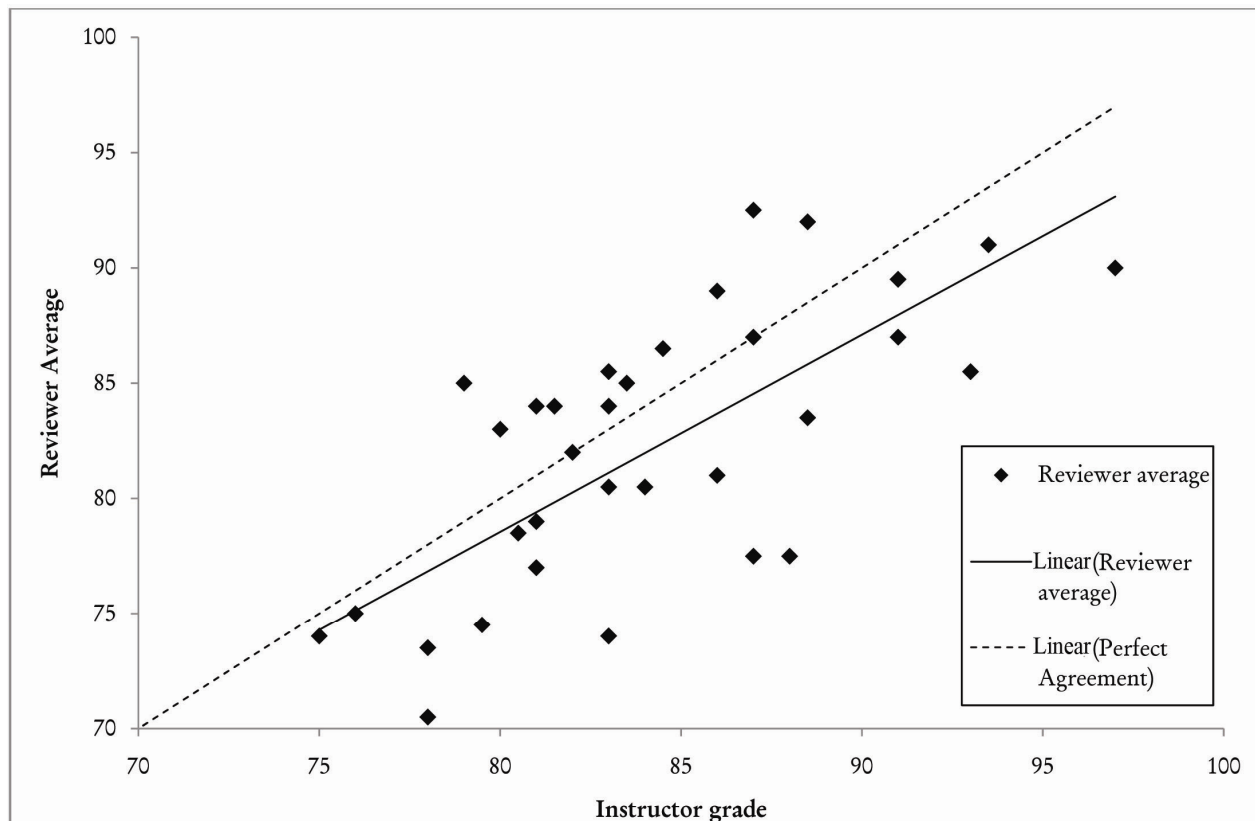

Figure 2. Regression analysis of instructor grades vs the average of two student reviewer grades

Dashed line represents the line of perfect agreement (x=y).

The reliability of grades provided by peer assessors was determined by comparing the grades assigned to each paper by the two reviewers, providing a measurement of the consistency of assessments between students on the same paper. The mean differences between assessments on the same paper are presented in table V. The regression analysis on the two grades assigned by the student reviewers had a correlation coefficient of r=0.26, and the coefficient was not significant with bootstrapping (p=0.13), indicating that the reviewer 1 grade is not a good predictor of the dependent variable reviewer 2 grade, and therefore, the model is non-significant. (Figure 3).

Table 5. Mean differences between pairs of student assessment on the draft version of the assignment. Data are presented as mean ± SEM.

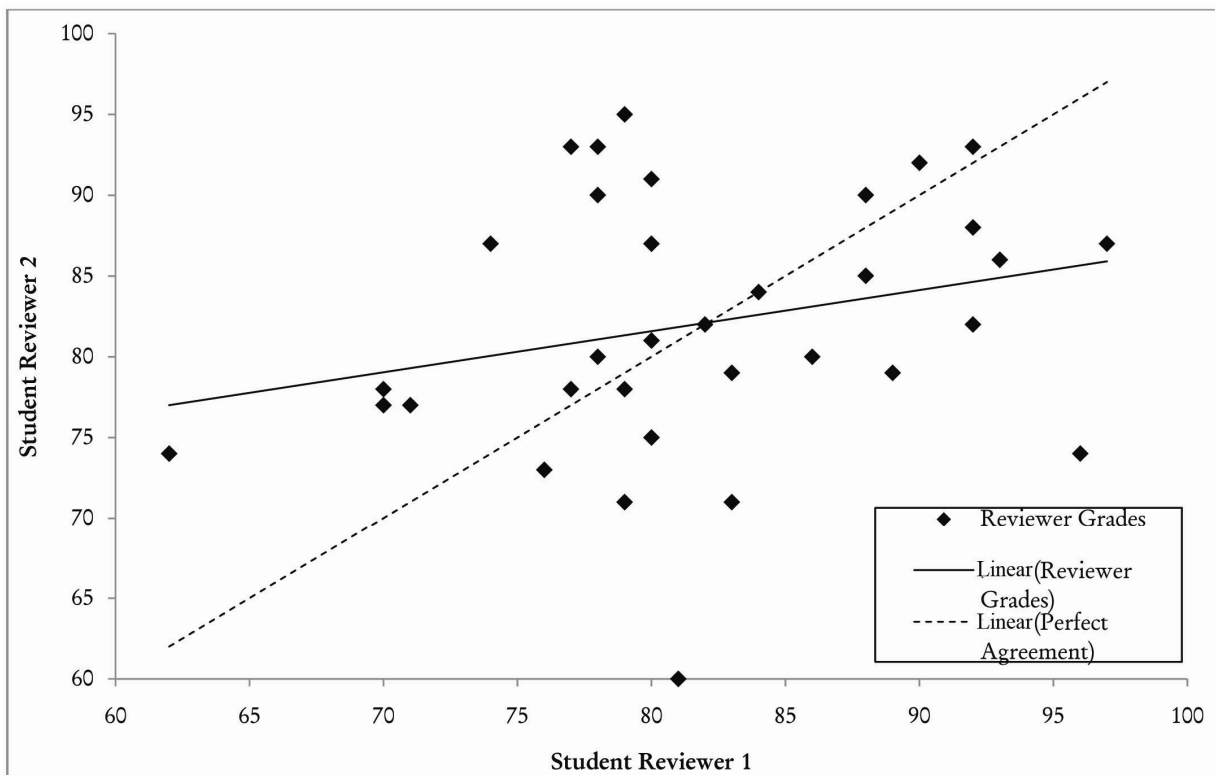| Assignment Component	Mean Difference
Overall Paper (/100)	7.58 ± 1.02
Introduction (/10)	1.44 ± 0.20
Critique (/50)	4.05 ± 0.56
Conclusion (/10)	1.24 ± 0.15
Grammar and Organization (/20)	2.21 ± 0.27
References (/10)	1.24 ± 0.20

Figure 3. Regression analysis of the two student reviewer grades

Dashed line represents the line of perfect agreement (x=y).

A comparison of the professor and student rankings on the draft version of the assignment showed differences across 32 of 34 ranks. The mean differences between student and professor rankings for papers ranked by the instructor in the class were considered as quintiles, although it should be noted that the bottom quintile contained six, rather than seven samples. The results for quintiles one through five are as follows: Q1 mean difference -7.29 +/- 3.41, Q2 mean difference -1.57 +/- 4.29, Q3 mean difference 1.14 +/- 2.87, Q4 mean difference 4.71 +/- 1.77, and Q5 mean difference 3.5 +/- 2.86. Therefore, it appears that student and professor rankings were most closely approximated in Q2 and Q3, followed by Q4 and Q5, with the top quintile having the greatest deviation.

3.3 Peer assessment and assignment quality

There was a significant improvement in the paper grades between the draft version and the final version (+3.4%, t_{32} = -6.5, p<0.001, Figure 4). It was further observed that the change in grade from the draft to final version was directly associated with the number of actionable comments provided by students as part of the assessment process (r=0.237, R^2=0.176, p=0.01) (Figure 5). The change in grade from the draft to final version was also indirectly associated with the instructor assigned grade on the draft version of the paper (r=-0.337, R^2=0.360, p<0.001), with lower graded draft papers showing the greatest change following receipt of peer assessment (Figure 6).
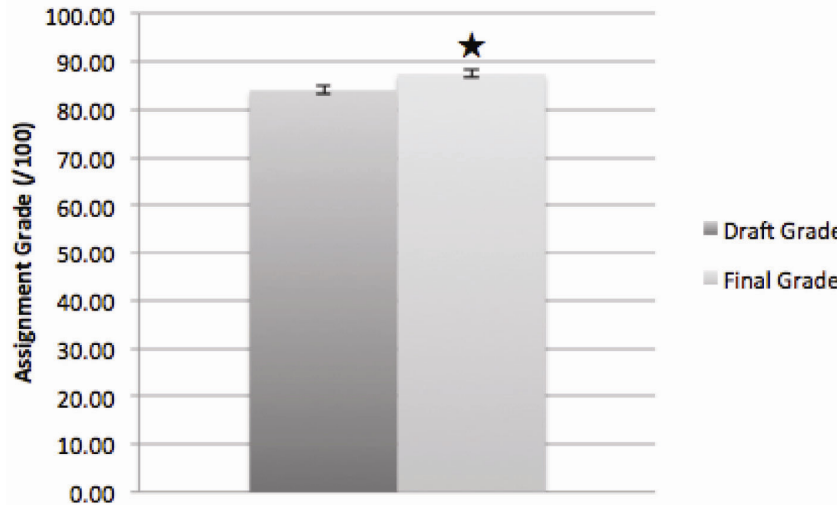
Figure 4. Comparison between instructor assigned grades on the draft and final version of the assignment
Data are presented as mean +/- SEM. The grade on the final version was significantly (p<0.0001) higher than the draft version.
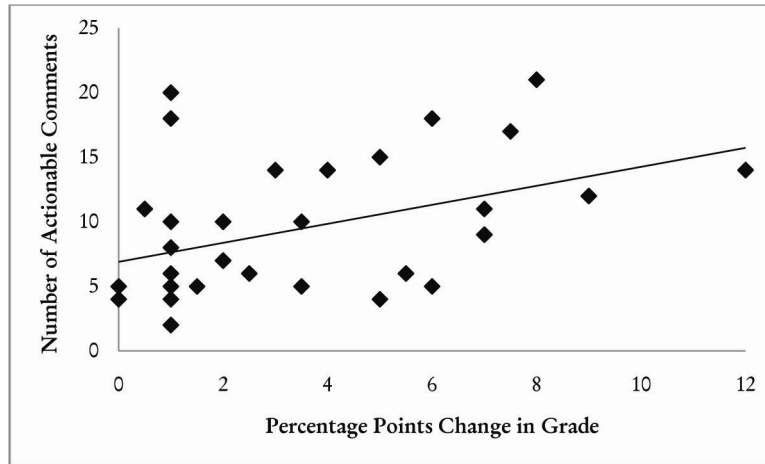

Figure 5. Relationship between the change in paper grade from draft to final version and the number of actionable comments provided by student assessors. There was a significant direct association (r=0.237, R^2=0.176, p=0.01) between the number of actionable comments provided and the change in paper grade.
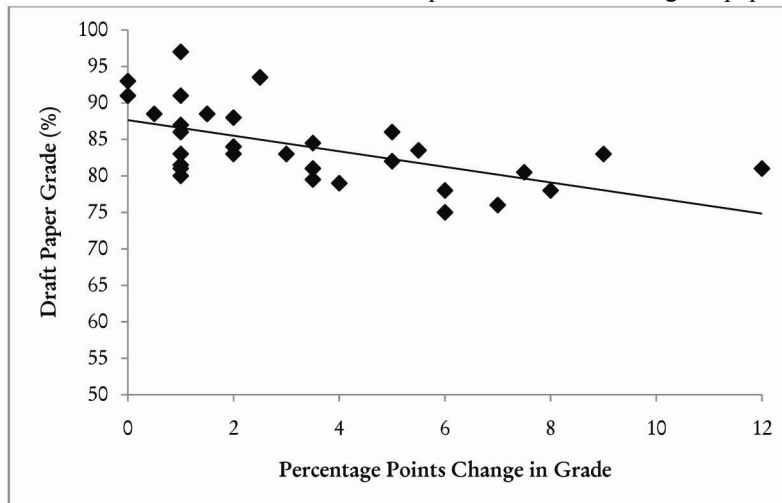

Figure 6. Relationship between the change in paper grade from draft to final version and the draft grade from instructor provided by student assessors. There was a significant indirect association (r=-0.337, R^2=0.360, p<0.001) between the draft grade and the change in paper grade.

4. Discussion

The objective of this study was to investigate the use of peer assessment in a graduate writing assignment, and to measure student perception, accuracy, and reliability. It was observed that student perception of the peer assessment process was very positive, and that there were no significant differences on the overall grades assigned by the student reviewers (as the average of two reviewers) and the instructor. Our data do suggest that multiple student reviewers are necessary in order to ensure a more accurate assessment since our analysis revealed disagreement between student reviewers.   Moreover, there was a significant improvement in assignment quality following revision after peer assessment. Cumulatively, these findings strongly support the use of peer assessment as a tool for use in graduate education.

In this study, there was overwhelming support for the use of peer assessment as a learning tool based on student perception. Overall, over 90% of students either agreed or strongly agreed that the peer assessment process was a valuable learning experience. These findings are consistent with previous research; although there is some variation among studies, published reports of student perception of the peer assessment process are generally very high across levels of education, with a majority of students reporting a positive experience at the graduate (Haaga, 1993; Topping et al., 2000) and undergraduate (Guillford, 2001; Orsmond, Merry & Reiling, 1996; Venables & Summit, 2003; Vickerman, 2009) levels.   The studies by Haaga (1993) and Topping et al. (2000) are important contrasts to the present study as there are very few published reports of peer assessment in graduate education. While Haaga (1993) found that the majority of graduate psychology students found the peer assessment process to be educational, as in the present study, they reported that the process of assessing one's own paper was a more valuable experience, with the peer assessment process being ranked at 7.9/10 while personal assessment was ranked at 8.9/10 (scaled as 10 being highly educational). However, it is important to note that the peer assessment process supports both activities because the receipt of peer assessment feedback is then used to support the revision of one's own paper. Similarly, in the graduate level study by Topping, Smith, Swanson & Elliot, (2000) it was found that the majority of students (83%) perceived the peer assessment process to be effective. It therefore appears that there is broad support by students across the levels of education for peer assessment, though student perception may be influenced by characteristics of the peer assessment process such as trust (Van Gennip et al., 2009) and positive feedback (Kaufman & Schunn, 2008).

Although students were required to dedicate time to assessing their peers' work, most students felt that the benefits were worth the time invested. This is in contrast to the study by Venables et al. (2003) wherein the students reported that the time required for the peer assessment process was too excessive. This difference in perception may be attributable to the differences in student populations in the two studies, as it is expected that graduate students are more motivated to participate in peer assessment given the greater relevance of the process to their current academic activities (that is, peer assessment of their own research). As well, it appears that the time investment for peer assessment was greater in the study by Venables & Summit (2003), which extended across much of the course, whereas the peer assessment process in the present study was limited to one week. Students in the present study were also given a reduced workload during the peer assessment week in order to complete the assessments. Though the design of the present study did not directly investigate this issue, it suggests that perception of the peer assessment process may thus be influenced by the time commitment required, where shorter, dedicated activities may be preferred to longer ones that co-occur with other learning activities.

An important learning outcome in education, particularly at the graduate level, is the ability to think critically and student responses to the survey in this study suggest that participation in the peer assessment process improved their critical thinking skills. Several students provided descriptive feedback regarding this impression, citing an improved ability to think critically about research and to be critical of their own papers following the critique of their peers' work. This observation is supported by Orsmond et al. (1996), who found that undergraduate students described improved critical thinking ability following peer assessment of scientific posters. Relatedly, Vickerman (2009) found that the majority of undergraduate students reported that peer assessment of written annotated bibliographies improved their critical writing and analytical abilities. A 2001 review of peer assessment in teaching and learning similarly describes an enhancement of critical thinking skills as an outcome of the peer assessment process (Morris, 2001).

While the observation of improved critical thinking skills has been frequently made, it is usually limited by subjectivity, as in the student reported improvement noted in the present study. However, the present study also makes a more objective observation of improved critical thinking ability due to the nature of the peer assessment assignment. In this assignment, students were required to critically evaluate current scientific research, considering

elements such as study strengths and weaknesses, integration between related studies, and the original contribution of each piece of work, with 50% of the assignment grade being conferred by the critique. The observation of a significant grade improvement following receipt of peer feedback was in part due to positive modifications in this area of the rubric, which suggests an improvement in critical analysis.

In contrast to the aforementioned positive perceptions, students were less supportive of the provision of grades by their peers, and many felt that the feedback that they provided was more helpful than that which they received, a finding which is supported by previous research. Sluijsmans, Moerkeke, Merrienboer & Dochy, (2001) found that students generally felt uncomfortable with awarding grades to their peers and preferred just giving qualitative feedback, while Cassidy (2006) found that the majority of students were generally uncomfortable assessing other students' work. However, it appears that these concerns may be unfounded: while over 30% of students in this study felt that the feedback they received from their peers was not comparable to that from a TA or professor, there was less than a 3% mean grade difference between the average of the student assessments and the instructor's, suggesting a high degree of accuracy by students (although student rankings of papers in the top quintile showed greater deviation from instructor ranking relative to the other quintiles). As well, even though the mean student grades were slightly lower than the mean instructor grades, these results do not strongly support students' concern that peer assessments were harsher than the instructor. This high degree of accuracy also does not support students' concerns about the usefulness of the quantitative feedback they received, citing difficulty with interpretation, since this feedback was very close to that which they would have received from the instructor.   These findings are not necessarily surprising at the graduate level, where the population is usually comprised exclusively of high-performing students, suggesting both that their assignments will be of high quality and that they will be better able to critically evaluate their peers' work. The study by Topping et al. (2000), which considered peer assessment using purely qualitative feedback likewise found good reliability between peer and instructor assessments. Similarly, Haaga (1993) found a good correlation between graduate student pairs when they assessed the same paper. Discussion of these research findings with future cohorts of students may help to mitigate concerns regarding discrepancies between student and instructor feedback.

Responses to the student survey in this study also revealed that the majority of students felt that the quality of their paper improved as a result of the experience and that their writing skills improved both by giving and receiving critical feedback. This perception is supported by the finding that there was a significant grade improvement of close to 4% in instructor assigned grades between the draft and final version of the assignment. Students with lower instructor assigned grades on the draft assignment showed a greater change in grade following peer assessment, which is consistent with previous research (Gielen et al., 2010). Although it may be assumed that experts are superior in providing effective feedback to students, this may not be the case: it is hypothesized that experts may give advice that is ambiguous or inconsistent, while students – who unlike experts do not possess an abundance of subject-matter knowledge in a specific discipline – are able to more effectively communicate appropriate feedback (Cho & MacArthur, 2010). The most important role of the assessor, whether it is a student or an instructor, is to provide in-depth, constructive feedback which helps to promote the development of higher order thinking skills (Bostock, 2000).

Analysis of the type of feedback that yields positive outcomes has been the focus of several recent investigations (Cho & MacArthur, 2010; Gielen et al., 2010; Strijbos, Narciss & Dunnebier, 2010; Van Steendam, Rijlaarsdam, Sercu & Van den Bergh, 2010). As summarized by Topping (2010), the most effective feedback is that which is non-directive, meaning that the comments were non-specific to that particular paper, although directive comments, which are specific to that individual paper, were also positively associated with improvements (Cho & MacArthur, 2010). This is in contrast to other types of feedback such as praise comments, which are positive or encouraging observations, or critical comments, which are negative evaluations in the absence of suggestions for improvement (Cho & MacArthur, 2010). Student peer assessors have been observed to provide a variety of feedback types, in contrast to instructor feedback, which tends to be directive (Cho & MacArthur, 2010). In the present study, each type of feedback was openly solicited from students in the form of "Commendations" (praise), "Recommendations" (directive and non-directive feedback), and "Corrections" (although these comments appear as criticism, they were accompanied by suggestions for improvement, which would re-classify them as directive comments). Rather than distinguishing between types of feedback, we focused instead on comments for which an action could be taken in the revision process, which we termed "actionable comments", that could fall into either the directive or non-directive categories observed a strong direct association between the number of "actionable comments" and the change in grade from draft to final version of the assignment. While the provision of praise may not be associated with improvements in assignment quality, positive student perceptions of the peer assessment process are associated with

positive and useful feedback (Kaufman & Schunn, 2008), so their inclusion is likely still of value. The practical implication of these findings is that students can easily be trained to provide feedback that brings about a meaningful change in assignment quality, using concise and straightforward instructions to phrase feedback so that it has clear consequences for revision.   Although Topping (2010) also notes that the impact of peer feedback is influenced by the academic competency of the assessor and assessee, this effect is likely to be considerably less in a graduate student population, which is largely comprised of high performing students. The range of marks on the final assignment, as graded by the instructor, was only 17%, and consisted of all A and A+ grade levels. For this reason, differences in performance between the students giving and receiving feedback were not given consideration in the present study.

The design of the present study included several characteristics of peer assessment that have been previously shown to be associated with positive outcomes. As already mentioned, inclusion of positive and useful feedback and the assignment of both a peer and instructor grade to student work results in more positive perceptions of the peer assessment process (Kaufman & Schunn, 2008), and use of non-directive and directive feedback is positively associated with improvements in assignment quality (Cho & MacArthur, 2010; Topping, 2010). Students in the present study were instructed to provide positive comments regarding the areas of the assignment that were done well in the qualitative component of their peer assessment, which also solicited non-directive and directive feedback through request for recommendations for improvement and suggestions for corrections. Moreover, while the draft version of the assignment was evaluated by students, there was a minimal contribution to their overall course grade, and the final paper was evaluated by the instructor. The instructor also considered rebuttals from students if they disagreed with their peer assessment grades, further minimizing the contribution of the peer grade in negative cases. As well, a review article by Van Zundert, Sluijmans & van Merrienboer, (2010) describes that training and practice have a positive influence on peer assessment, and use of a rubric, or clear grading criteria, is also predicted to improve outcomes (Graves, 2013; Orsmond et al., 1996; Mulder et al., 2014). Cho & Wilson (2006) use the term "scaffolded peer assessment" to describe this general process. In the present study, students took part in a three-hour writing workshop prior to engaging in peer assessment of the writing assignment in which they were instructed how to use the rubric, applied the rubric to two sample papers, and received feedback from the instructor on how their qualitative and quantitative comments aligned with the instructor grades on the same papers, all of which likely contributed to a level of comfort and proficiency with the rubric and assignment criteria that may have enhanced their abilities to provide accurate and reliable peer assessments. Moreover, training and clarity may have promoted a trusting environment in the classroom, which has similarly been shown to confer positive outcomes with peer assessment (Van Gennip et al., 2009). And lastly, research suggests that there should be a minimum of two assessors for each paper (Marcoulides & Simkin, 1995), with this minimum criteria being met in the present study, although higher numbers of peer assessors are recommended for better accuracy (Cho & Wilson, 2006). The positive findings of student perception of the peer assessment experience and the good accuracy of graduate student peer assessors in the present study therefore likely results from inclusion of a number of characteristics into the peer assessment process.

There are several limitations to the present study that should be considered. First, at 44 students for the subjective investigation, and 34 for the investigation of accuracy and reliability, the number of subjects is low. However, class sizes at the graduate level are typically very small, and the present study included two classes of students enrolled across two years. A further limitation to subject enrollment is that only approximately 50% of eligible students participated, but with no compensation or incentive being provided, this is not unexpected. Notably, the students who participated in the study did not show any differences related to performance or gender relative to those who did (data not shown). The study was also gender biased, with only 18% male participants, though this course consistently has a female majority and similar proportions of eligible male and female students participated in the study. A small effect of gender was observed in peer assessment by Langan et al. (2005), albeit on an oral, rather than written, assignment. And lastly, generalization of these results is limited by the inclusion of students in a single discipline, although use of peer assessment in a general writing assignment – rather than an assignment exclusive to the discipline of nutrition – may be relevant. When Cho & Wilson (2006) considered peer assessment at the graduate level across four graduate courses simultaneously, variability among the courses was observed suggesting that peer assessment may not confer the same findings in all contexts, although an important consideration when comparing results between studies is that most did not consistently incorporate peer assessment characteristics that are known to confer positive outcomes. Greater consistency among the designs of peer assessment activities across courses is likely to yield more consistent research findings. However, more research is needed to determine whether the

findings of student perception, accuracy, and reliability in the present study are consistent between males and females and across different disciplines of graduate education.

In conclusion, the present study strongly suggests that peer assessment is a valuable tool to use in graduate peer writing assignments. Students were very supportive of the activity, and their negative concerns related to inconsistent peer evaluations were not supported by the quantitative findings, since students provided averaged grades that were very close to those from the instructor. The low student-reviewer reliability suggests simply that multiple student reviewers should be used during such an exercise. These findings were consistent across all categories of the assignment rubric. Students showed a significant grade improvement following revision subsequent to peer assessment, with lower graded papers showing the greatest improvement; greater grade change was also associated with an increased number of comments for which a clear revision activity could be taken. This study did note student concern with the technology used to manage the peer assessment activity, and while this could be mitigated by use of a non-electronic peer assessment method, the technology allows for clear management of all stages of the process (such as recording the dates/times of student assessments and summaries of grades), so may be worthwhile regardless. While there is limited research regarding use of peer assessment in graduate education, the findings of the present study are consistent with previous research, although the strength of student support and degree of grading accuracy herein is slightly higher. This may be due to the incorporation of characteristics that have been previously shown to support positive findings, such as training, use of a clear, promotion of a trusting environment, use of peer and instructor grading, provision of directive and non-directive feedback, recruitment of positive comments, and use of more than one peer assessor. This study therefore builds on previous work and suggests that use of a carefully designed peer assessment activity, which includes clear direction regarding actionable comments, may provide students with useful feedback that improves their performance on a writing assignment.

References

Bangert-Drowns, R., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Assessment of Educational Research*, 61, 213-238. http://dx.doi.org/10.3102/00346543061002213

Bostock, S. (2000). Student peer assessment. *The Higher Education Academy*. Accessed online at: http://www.reading.ac.uk/web/FILES/engageinassessment/Student_peer_assessment_-_Stephen_Bostock.pdf. Accessed 18 August 2014.

Cassidy, S. (2006). Developing employability skills: peer assessment in higher education. *Education and Training*, 48(7), 508-517. http://dx.doi.org/10.1108/00400910610705890

Cho, K., &MacArthur, C. (2010). Student revision with peer and expert reviewing . *Learning and Instruction*, 20, 328-338. http://dx.doi.org/10.1016/j.learninstruc.2009.08.006

Cho, K., Schunn, C.D., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891-901. http://dx.doi.org/10.1037/0022-0663.98.4.891

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3-31. Accessed online at: https://www.open.ac.uk/fast/pdfs/Gibbs%20and%20Simpson%202004-05.pdf

Gielen, S., Peeters, E., Doucy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning . *Learning and Instruction*, 304-315. http://dx.doi.org/10.1016/j.learninstruc.2009.08.007

Graves, R. Why students struggle with writing. University Affairs, September 11, 2013. Accessed online at: http://www.universityaffairs.ca/why-students-struggle-with-writing.aspx. Accessed 18 August 2014.

Guilford, W. (2001). Teaching peer review and the process of scientific writing. *Advances in Physiology Education* , *25*, 167-175. http://dx.doi.org/10.1152/advan.00071.2009

Haaga, D.A.F. (1993). Peer review of term papers in graduate psychology course. *Teaching of Psychology*, 20(1), 28-32. http://dx.doi.org/10.1207/s15328023top2001_5

Kaufman, J.H., & Schunn, C.D. (2011). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*, 39(3), 387-406. http://dx.doi.org/10.1007/s11251-010-9133-6

Langan, M.A., Wheater, C.P., Shaw, E.M., Haines, B.J., Cullen, W.R., Boyle, J.C., et al. (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of

assessment criteria. *Assessment and Evaluation in Higher Education*, 30(1), 21-34. http://dx.doi.org/10.1080/0260293042003243878

Marcoulides, G.A., & Simkin, M.G. (1995). The consistency of peer review in student writing projects. *Journal of Accounting Education*, 14(4), 515-533. http://dx.doi.org/10.1080/08832323.1995.10117753

Morris, J. (2001). Peer assessment: A missing link between teaching and learning? An assessment of the literature. *Nurse Education Today*, 21, 507-515. http://dx.doi.org/10.1054/nedt.2001.0661

Mulder, R.A., Pierce, J.M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education*, http://dx.doi.org/10.1177/146987841452739

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education*, 21(3), 239-250. http://dx.doi.org/10.1080/0260293960210304

Sluijsmans, D.M., Moerkeke, G., Merrienboer, J.J., & Dochy, F.J. (2001). Peer assessment in problem-based learning: Students' views. *Studies in Educational Evaluation*, 27,153-173. Accessed online at: http://www.iaea.info/documents/paper_2b7122b93.pdf

Strijbos, J.W., Narciss, S., & Dunnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291-303. http://dx.doi.org/10.1016/j.learninstruc.2009.08.008

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. http://dx.doi.org/10.1080/713611428

Topping, K.J. (2010). Methodological quandaries in studying process and outcomes in peer assessment, *Learning and Instruction*, 20, 339-43. http://dx.doi.org/10.1016/j.learninstruc.2009.08.003

Van Gennip, N.A.E., Segers, M.S.R., & Tillema, H.H. (2010). Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learning and Instruction*, 20(4), 280-290. http://dx.doi.org/10.1016/j.learninstruc.2009.08.010

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327. http://dx.doi.org/10.1016/j.learninstruc.2009.08.009

Van Zundert , M., Sluijmans, D., & van Merrienboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20, 270-279. http://dx.doi.org/10.1016/j.learninstruc.2009.08.004

Venables, A., & Summit, R. (2003). Enhancing Scientific Essay Writing Using Peer Assessment. *Innovations in Education and Teaching International* , 40, 281-290. http://dx.doi.org/10.1080/1470329032000103816

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning?. *Assessment and Evaluation in Higher Education*, 34(2), 221-230. http://dx.doi.org/10.1080/02602930801955986