

The effect of assessment of peer feedback on the quantity and quality of feedback given

Rachael Ruegg

English for Academic Purposes Department, Akita International University

There has been a great deal of debate about the value of peer feedback in L2 writing classes. Different aspects of the way peer feedback is implemented have been found to contribute to its effectiveness. The purpose of the current study is to ascertain whether the assessment of feedback given by peers increases the quantity or quality of feedback given. The study investigated two intact classes at a Japanese university. Both groups used peer feedback on every preliminary draft for an entire year. One was assessed only on the final draft of each essay and the other on the feedback they gave to their peers in addition to the final drafts. The feedback given by students was analysed and compared between the two groups. It was found that the feedback-assessed group covered more points, wrote more comments, longer comments, more words overall, made more marks on partners' drafts, and made more specific comments than the product-assessed group. However, no significant difference was found between the accuracy of feedback in the two groups. The results suggest that if instructors want peer readers to give more feedback and to give more specific feedback, the feedback given by students should be assessed.

Key words: classroom assessment, peer feedback, peer review, peer response

Introduction and background

In the field of L2 writing there has been a great deal of debate over the effectiveness of peer feedback. As has been claimed by some concerning teacher feedback (e.g. Chandler, 2003; Ferris, 1999; Lee, 1997), it may not be a question of whether peer feedback is effective but rather of how it should best be implemented in the classroom to maximise its effectiveness.

For peer feedback to be effective it needs to be integrated into the curriculum as a usual activity rather than simply being used a few times throughout a course. This idea has been supported by some research. Kashimura (2007, as cited in Hirose, 2008) looked at the perceptions of lower-level Japanese students of peer feedback. However, rather than integrating peer feedback as a part of the writing course, he carried it out only three times over the course of a year. The students had negative perceptions of peer feedback, two thirds of them saying they didn't want to do it again. It seems that three times is insufficient for students to practice giving peer feedback and become confident in giving it. Furthermore, as an exception to the norm in class rather than a usual practice students would not have been likely to consider it as important or valuable as they would have if it had been an integral part of the course. On the other hand, in Hirose's (2008) study, also carried out in Japan, peer feedback was used every week as an integral part of the writing course. In Hirose's study students had positive perceptions of every aspect of peer feedback: reading their peers' writing, reading their peers' feedback, talking with their peers about their writing as well as giving their peers feedback. They considered the feedback they got from their peers to be as good as that which they got from the teacher. Furthermore, they enjoyed reading their peers' comments more than they enjoyed reading their teacher's comments.

Apart from the frequency of carrying out peer feedback, the level of support (or scaffolding) required by students to result in effective comments has also been discussed in the literature (Gere, 1987) with most agreeing that L2 learners require a high level of scaffolding to begin with (Berger, 1990). For example, it may be better to use non-autonomous feedback sheets initially so that students have a clear structure for the activity. Non-autonomous feedback sheets are explained by Gere (1987) as those which entail filling out a feedback checklist or editing sheet rather than giving open-ended feedback. In contrast to this, Hyland (2000) states that while teachers' intentions in using feedback sheets are good, believing that the sheets focus students' attention on appropriate aspects of writing, many researchers have stated that they turn peer feedback away from real communication and into yet another way to please the writing teacher. Indeed, the use of such non-autonomous feedback sheets may make

peer feedback into just one more hoop students are required to jump through in order to pass the writing course. They carry out peer feedback because they are required to rather than really considering what they are doing and why they are doing it.

Various different methods of peer feedback have been commonly used and it has been stated that written peer feedback is superior to oral peer feedback because reviewers have time to contemplate both the content of the feedback they give and the appropriate wording (Baker, McQuade, Sommers, & Tratner, 1989; Huff & Kline, 1987). Oral feedback, on the other hand, is very immediate, leading to not only reduced consideration of the effect the comments may have on the writer but also added cognitive pressure for less proficient L2 learners. Ferris (2010) adds that written feedback is more likely to have positive effects on revision because students have more time to consider the feedback and modify their output.

Finally, it has been found that learners who are well prepared for the peer feedback activity are more effective peer readers than those who have been less well prepared (Berg, 1999; Min, 2006; Stanley, 1992; Zhu, 1995). Berg (1999) compared learners well trained in peer feedback with those less well trained and Min (2006) compared learners before and after receiving peer feedback training in terms of the revisions made by the writers after having received peer feedback. Both studies found that training positively influenced revision types and the quality of the revised texts; however, they did not attempt to compare the actual feedback given by learners in the two groups. Stanley (1992) and Zhu (1995), on the other hand, both compared the quantity and quality of the feedback given by learners who were well trained in peer feedback with that of learners who were less-well trained. Both of these studies found a significant positive effect for training learners well for the peer feedback process.

In Stanley's (1992) study, oral peer feedback was analysed in terms of the nature of each comment, the number of turns taken by each speaker and the mean length of turn. She stressed that shorter turns were preferable when giving oral feedback. However, in the context of written feedback, because negotiation of meaning occurs differently, it would seem that longer comments would be preferable. Longer comments provide more specific and detailed feedback than shorter ones and feedback which is more specific and more detailed is considered to be better than feedback which is vague or general (Grabe & Kaplan, 1996; Ferris, 1997, 2003; Hyland, 2001; Myers, 1997; Seror, 2009). Stanley (1992) states that during the training sessions, students were encouraged to be specific and that the trained students offered significantly more specific comments than their untrained counterparts.

Similarly, Zhu (1995) hypothesised that students who were trained for peer feedback would give more feedback, more feedback about global issues, and more specific feedback. She analysed the written feedback given by peer readers and found that trained students did indeed give more feedback, more feedback on global issues and more specific feedback than their untrained peers.

While some have stated that it is preferable to direct students to focus on rhetorical issues (such as content and organisation) and ignore language problems (Berg, 1999; Min, 2006; Stanley, 1992; Zhu, 1995) others consider that learning writers themselves should decide what they would like readers to provide feedback on (Hyland, 1996; Liu & Hansen, 2002). It seems fair to say that the focus of the feedback should vary depending on the aims and goals of the course. While feedback on rhetorical issues may be of primary importance to learners in academic ESL contexts, feedback on language use errors may be more helpful for EFL learners (Tsui & Ng, 2000) and those in general language courses (Ferris, 2010).

Often, literature written on the topic of peer feedback has recommended that peer feedback should be assessed (Ferris, 2003; Ferris & Hedgcock, 2005; Hyland, 1996; Leki, 1990; Mittan, 1989). It has been stated that if peer feedback is not assessed, learners can see the peer feedback process as a mere hoop to jump through rather than as a constructive pedagogical activity (Ferris, 2003), meaning that they may simply go through the motions of giving peer feedback without giving much consideration to the purpose and effects. The assessment of peer feedback is claimed to encourage learners to take the process more seriously (Ferris & Hedgcock, 2005), apply themselves more and in particular, assessment is claimed to lead to greater consideration of both the feedback given and received by learners. While some recommend assigning a grade for the quality and/or quantity of feedback given (Ferris, 2003; Ferris & Hedgcock, 2005; Hyland, 1996; Mittan, 1989), Leki (1990) suggests that it is effective for teachers to read the feedback and give learners their comments on the feedback.

The assessment of peer feedback has often been recommended and anecdotal evidence suggests that the assessment of peer feedback is now a fairly common practice, and that many teachers believe that it increases the effectiveness of the feedback given. However, to my knowledge there is no published research attempting to investigate the effects of the assessment of peer feedback on the feedback given and received by peers. Indeed, Min (2006) assessed the feedback given during peer feedback in her study on the effects of training on peer feedback and recommended a study to ascertain the effects of assessment of peer feedback on the feedback given. Accordingly, the purpose of the current

study is to ascertain whether assessment of the feedback given by learners during the peer feedback process increases the effectiveness of the feedback given. The research questions for the study are:

1. Does the assessment of feedback given by peers lead to an increase in the quantity of feedback given?
2. Does the assessment of feedback given by peers lead to an increase in the quality of the feedback given?

Method

Participants and data

The participants of this study were drawn from 30 students in two intact classes in the second year of study at a private university in Eastern Japan who gave their consent to being involved in the study. All students were majoring in English and classes in the first two years are conducted in English only. All students in the first and second year are placed in one of three tiers, based on their TOEFL scores as well as their scores on the writing and speaking sections of an in-house general English proficiency test given prior to the start of the academic year. Within the five classes which constitute the lowest ability tier, students are randomly distributed. The classes involved in this study were both in the lowest ability tier. They were considered to range from pre-intermediate to upper-intermediate level and their TOEFL scores ranged from 370 to 500. Students in the middle tier have a higher overall proficiency level and those in the top tier have typically either lived abroad or studied at international schools and are therefore of near-native proficiency level. In order to ascertain that the two classes were equal in terms of writing ability at the beginning of the academic year, an independent samples *t*-test was conducted, comparing the in-house writing test scores between the two groups. It was found that there was no significant difference between the writing test scores of the two groups at the beginning of the academic year; $t(25) = 1.533, p = 0.138$.

In their first year writing class, students were supposed to be introduced to multiple draft writing, including the processes involved in this, such as; giving and receiving feedback, and revising. However, out of the 30 students who gave consent to be involved in this study, three reported at the beginning of the year that they had never experienced giving or receiving peer feedback in either Japanese or English. Therefore, a small minority of the students were peer feedback novices.

Each semester lasted for 15 weeks and the writing classes met twice a week for 90 minutes. The data was collected for one academic year (two semesters),

meaning that altogether the classes met 60 times, for 90 minutes each time. The students were required to complete three drafts each of eight assignments over the data collection period and to carry out peer feedback on the first and second drafts of each assignment. A non-autonomous feedback sheet was used in this study because of the students' low level of English proficiency and the cultural context of the study. Japanese learners are said to be less accustomed to having autonomy over their learning than their counterparts from other cultural contexts (Benson, 2001; Dias, 2000). The feedback sheet was a checklist which contained blank lines where writers wrote four questions for the peer reader to answer. A fifth question asked the reader to give one piece of constructive feedback. The meaning of the word 'constructive' was explained to the learners as something that will help the writer to make the essay better; i.e. advice, as opposed to praise. Some examples of constructive feedback and praise were used in class to check that the learners had understood the concept. The purpose of the fifth question was both to allow the reader some freedom to point out something problematic that they had noticed in the draft and also to ensure that every preliminary draft received at least one constructive comment which the writer could use in revising their draft. This feedback sheet was designed by the instructor and had been used in similar classes for several years before the study took place.

Students were asked to work with a different partner each time they carried out peer feedback in order to experience working with as wide a range of partners as possible. Therefore, a motivated student who usually attended class would have carried out peer feedback 16 times during the data collection period and would have worked with at least 10 different partners over the one year period. During a peer feedback lesson the students would do a warm-up activity, be organised into pairs and then have most of the lesson to conduct peer review. If they finished early, they could start drafting the subsequent draft. Everyone was stopped shortly before the end of class to explain the homework and rearrange the desks. Therefore, each peer feedback session lasted around one hour. With the final draft of each assignment, students were required to submit all marked up preliminary drafts and completed feedback sheets.

All data, in the form of essay drafts and completed feedback sheets, was collected over a one year period and organised at the end of the year. Only essays for which all three drafts had been written and submitted and feedback sheets had been completed and submitted were included in the initial data pool. This led to a total of 13 to 22 complete sample essays for each assignment from the two groups. At least five essays from each group for each assignment were complete and included in the data pool. Therefore, five sample essays from each of the two groups were randomly selected from the pool for each

assignment. This resulted in a total of 80 essays (160 drafts) being selected and analysed for this study, including 40 essays (80 drafts) from each group.

This random selection process resulted in 14 students from each of the two classes being included in the analysis, while one from each class was excluded. In total, 78.6% (22) of the participants were female and 21.4% (6) were male. Although all the participants were studying in the same department at the same Japanese university, three of the participants were not Japanese nationals, two were Chinese and one was Thai. The participants ranged in age from 19 to 21 years at the beginning of the academic year.

Treatment

In order to ascertain the effect of the assessment of feedback on the quantity and quality of the feedback given, two intact classes were included in this study. Both classes used peer feedback exclusively for a period of one academic year; however, they were assessed differently. One class was assessed on the final draft of each essay they wrote (product-assessed group) while the other was assessed on the peer feedback they gave as well as the final drafts of their essays (feedback-assessed group). The score assigned for peer feedback was worth 25% of the grade they received for each assignment, while the remaining 75% was assigned for the quality of the final draft.

The peer feedback grade was determined by the instructor (who was also the researcher) based on the quantity and quality of the feedback given. As is common when assessing student work, the grades were determined subjectively based on the instructor's impression of the overall quality and quantity of feedback given. For example, a student who offered no constructive feedback, but rather praise alone, would receive a score of 1 out of 5, whereas a student who filled the entire feedback sheet with comments which were clear, detailed and considered by the instructor to be good advice, would receive a grade of 5 out of 5. If a student failed to carry out peer feedback on either preliminary draft of the essay and submitted the single, final draft, they would receive a score of 0 for peer feedback. Learners were encouraged to ask questions about the grades they received on their writing assignments, and they often did. If a learner asked about their peer feedback grade, the instructor would explain the grading criteria and would give the learner an example of peer feedback that received 5 out of 5. They would have a chance to read the feedback as a kind of model to see how much and what kind of feedback was expected.

The assessment was designed to be simultaneously formative and summative assessment. The students received their grades for each essay within one to two

weeks of having submitted it, in every case they received their grade for one essay before the subsequent essay was due to be submitted. This was summative due to the fact that each grade accounted for a proportion of the final course grade and formative due to the fact that the students received feedback about each essay before the next was due. They received not only a single grade, but three or four separate grades. This constituted feedback about their strengths and weaknesses so that they would know what to focus on in subsequent essays. Due to the formative aspect of the assessment process, the instructor expected students to improve upon their weak points from the start to the end of the academic year. In this way, it was considered that students who received relatively low scores for their peer feedback would consciously work on giving more and better feedback subsequently.

Both classes were instructed in exactly the same way by the same instructor. They participated in a 90-minute training session on peer review at the beginning of the year. During the training, the purpose of peer feedback and the benefits of giving and receiving peer feedback were discussed, following this the feedback sheet and process were explained to the learners and they participated in a practice peer-review session and received feedback from the teacher about their feedback. In addition to the training, the instructor monitored the class while they were giving peer feedback, both in order to answer questions which may have arisen and also to give advice about the peer feedback process. For example, if a pair of students finished giving peer feedback unusually quickly, the instructor would ask them about the process they used and remind them of any steps they had missed. This happened predominantly in the first one or two assignments when the students were still adjusting to the peer feedback process. At the beginning of the second semester, both classes had a refresher training session to remind them of the purposes and processes involved in peer feedback. This session lasted for around 45 minutes.

Data analysis

The peer feedback on all 160 drafts was evaluated by the researcher in terms of both quantity and quality. The quantity variables counted were: the number of different constructive points made by the reader, the number of marks made on the draft, the number of constructive comments written, the total number of words contained in constructive comments on both the feedback sheet and the draft and the average length of each comment (in words). Quality was determined by two variables: whether each point and mark made by the reader was accurate (good advice) and specific. For the purpose of this study praise and encouragement were excluded from the analysis because it was not

considered that differences in the quantity or quality of praise and encouragement would constitute more or less helpful feedback; however, student readers also praised and encouraged each other almost without fail.

The number of different points made by readers related to the number of different issues addressed in the feedback overall, this indicates the breadth of the feedback. For example, one learner could give a large amount of feedback, but all of it relating to the correct use of articles. On the other hand, another student might give less feedback, quantity wise, but address a larger number of different issues, showing more breadth. The number of marks on each draft related specifically to the act of marking up the peers draft and how many such marks were made. Marks referred to any visible feedback on a draft that was not written in word form. For example, underlining, circling, arrows and question marks were all considered marks. The number of constructive comments written relates to all comments written both on the draft and on the feedback sheet. The total number of words contained in the comments was counted in words, and to calculate the average length of comments, this figure was divided by the number of comments made.

The accuracy of each point and mark was scored on a three point scale whereby a score of 0 indicated that it was bad advice, a score of 0.5 indicated that it was a matter of personal preference (either the original or the suggested revision would be equally as suitable) and a score of 1 indicated that it was good advice. Feedback points were usually written on the feedback form, with corresponding marks on the draft in question. Points and marks were both included in the accuracy variable because sometimes the feedback point itself was correct but one or more of the places where the reader suggested using their feedback was inappropriate.

An example of a feedback point that received a score of 0 for accuracy follows:

Original: 'A British non-profit group said to plan to air the country's first TV advertisement for abortion services.'

Comment: 'I couldn't understand "air" meaning in this case without dictionary. I suggest that you change it another word, example "announce".'

Although the first sentence in this comment is valid, the suggestion to change the word to "announce" would decrease the quality of the draft and therefore, this comment received a score of 0.

An example of a feedback point that received a score of 0.5 for accuracy follows:

Original: 'The air-conditioning in the American house is a central heating method.'

Suggestion: 'The air-conditioner in the American house is a central heating method.'

As it would be equally acceptable to use 'air-conditioning' or 'air-conditioner' in this context, this comment received a score of 0.5.

An example of a feedback point that received a score of 1 for accuracy follows:

Comment: 'I can't know which sentence is quoted or paraphrased. You should delete ' ' from paraphrased sentence.'

The point that paraphrased sentences should not be enclosed in quotation marks is accurate and therefore, this comment received a score of 1.

Feedback was determined to be specific if it would be possible to link it with its intended revision in the subsequent draft. Feedback which was unspecific received a score of 0, while specific feedback points were scored as 1.

An example of a feedback point that received a score of 0 for specificity follows:

Comment: 'Look at the grammar again more carefully.'

This comment was considered not to be specific because the reader did not indicate what kinds of grammar problems there were in the essay or where in the essay they were located.

An example of a feedback point that received a score of 1 for specificity follows:

Comment: 'Childbearing in an advanced age is accompanied by danger. That's why if you agree [with] it I think you have to write [a] more persuasive draft, for example; leading medical care.'

Although there are clearly language problems, this comment was considered specific because it gave specific advice about what the writer should do to improve their draft.

The scores for each feedback point were averaged so that the score for each draft constituted a proportion, from 0 to 1 of the number of comments and marks that were accurate. The data was coded by the researcher, then after an interval of two months, 20% of the data was recoded in order to assess the intra-rater reliability of the coding. The Pearson correlation intra-rater reliability was

0.996. An example of a very unhelpful peer feedback sheet and a very helpful peer feedback sheet in terms of these measures of quantity and quality can be seen in the appendices (Appendices A and B respectively).

T-tests are a parametric test to determine whether the means of groups are equal (Huizingh, 2007). When using t-tests two assumptions need to be met: the assumption of normal distribution and the assumption of equal variance (Huizingh, 2007). The Mann-Whitney U test is the non-parametric alternative to an independent samples t-test, which can be used when these two assumptions are not met (Huizingh, 2007). Because the assumption of equal variance was not met for all variables in this study, all variables were compared between the two groups using a Mann-Whitney U Test in order to ascertain whether there was any significant difference between the quantity or quality of the feedback given in the two groups. Subsequently, in order to determine whether the assessment of feedback led to greater increases in quantity or quality of feedback given by the feedback-assessed group over time compared to the product-assessed group, the data from the first semester (constituting 3 assignments, 6 drafts) and the second semester (constituting 5 assignments, 10 drafts) was compared separately using Mann-Whitney U Tests.

Results

Table 1 shows the descriptive statistics for all seven variables on all drafts for three different time periods: semester one, semester two and the whole year. Both groups decreased the number of different points made on each draft between semester one and semester two. However, while the product-assessed group also decreased the number of marks they made on each draft, the number of comments they wrote on each draft, the number of words written and the accuracy of the points and marks between semester one and semester two, the feedback-assessed group increased their number of marks, number of comments, the number of words written and accuracy of points and marks made on each draft. In terms of the length of comments, and the proportion of specific comments, both groups increased between semester one and semester two.

Table 1. Descriptive Statistics for all drafts

Variable	Semester One				Semester Two				Whole year			
	Product-assessed		Feedback-assessed		Product-assessed		Feedback-assessed		Product-assessed		Feedback-assessed	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Points	3.6000	4.3040	4.4667	3.4214	2.6600	2.5202	4.3000	2.7124	3.0125	3.3090	4.3625	2.9777
Marks	3.1000	4.9644	2.6667	3.0324	2.4600	2.7495	4.5200	3.9083	2.7000	3.7193	3.8250	3.6967
Comments	1.1000	1.3734	1.7667	1.1943	0.9400	0.8184	3.0800	2.3198	1.0000	1.0554	2.5875	2.0666
Words	10.9333	12.0829	21.2333	13.9177	10.7600	11.4759	39.3800	25.8322	10.8250	11.6312	32.5750	23.7310
Length	7.2627	7.4599	13.2278	9.8718	8.3800	8.7510	14.6998	10.6159	7.9610	8.2598	14.1478	10.3048
Accuracy	0.7262	0.3955	0.7953	0.2815	0.7139	0.3979	0.8441	0.2509	0.7185	0.3945	0.8259	0.2621
Unspecific	0.1519	0.2364	0.1247	0.2723	0.0683	0.1412	0.0332	0.0886	0.0997	0.1859	0.0675	0.1846

n = 80

It can also be seen that in the product-assessed group there was much more variation between students in terms of the quality and quantity of feedback given than in the feedback-assessed group, as demonstrated by the proportionally higher standard deviations of the product-assessed group. This shows that the quality and quantity of feedback given in the product-assessed group varied greatly from person to person which suggests variation in the amount of effort made during the peer feedback activity. In the feedback-assessed group, on the other hand, after the first semester, the only variable with a higher standard deviation than mean is 'unspecific', showing that perhaps learners need to be trained more in how to give specific feedback.

The Mann-Whitney U Test between the groups for the first semester revealed significant differences (at the 0.05 level) between the number of comments given in the two groups; (0.005), the total number of words written in the two groups; (0.003), and the average length of comments made; (0.006), while the number of different points made by the readers in the two groups; (0.094), the number of marks made on drafts; (0.812), the accuracy level of the comments and marks; (0.720) and the proportion of specific comments made by the two groups; (0.173) were not significantly different at the 0.05 level.

The Mann-Whitney U Test between the groups for the second semester revealed significant differences (at the 0.05 level) between the number of different points made in the two groups; (0.000), the number of marks made on drafts in the two groups; (0.001), the number of comments written in the two groups; (0.000), the total number of words written in the two groups; (0.000) and the average length of comments in the two groups; (0.000). However, no significant difference was found (at the 0.05 level) between the accuracy level of the comments and marks; (0.318), nor the proportion of specific comments; (0.075) in the two groups.

The Mann-Whitney U Test between the two groups including the data from the whole year revealed significant differences (at the 0.05 level) between the number of points made in the two groups; (0.000), the number of marks made on drafts in the two groups; (0.007), the number of comments written in the two groups; (0.000), the total number of words written in the two groups; (0.000), the average length of comments; (0.000), and the proportion of specific comments made in the two groups; (0.026), while the accuracy level of the comments and marks in the two groups was not significantly different at the 0.05 level; (0.319).

Discussion

It seems clear that assessing the quality and quantity of peer feedback did lead to feedback that was better, in terms of both quality and quantity, in the feedback-assessed group. Taking the data from the entire academic year into consideration, the feedback-assessed group made more different constructive feedback points, wrote more comments, more words overall than did the product-assessed group and they also wrote longer comments and a significantly higher proportion of specific comments than the product-assessed group.

In the feedback-assessed group the minimum number of constructive feedback points given was one in both the first semester and the second semester, meaning that every draft received at least one suggestion for improvement. In the product-assessed group on the other hand, in both semesters there were drafts that received no constructive feedback whatsoever, leaving the writer with nothing to revise after the peer feedback session. In fact, there were five such drafts in the first semester and five in the second semester, a total of 10 out of the 80 drafts analysed.

Taking the data from the first semester into consideration, the feedback-assessed group wrote more comments than the product-assessed group, wrote more words overall than the product-assessed group and wrote on average longer comments than did the product-assessed group; however, there was no significant difference between the number of different constructive feedback points offered, the number of marks on drafts or the accuracy of the feedback. The grading policies were outlined clearly at the beginning of the academic year, so this difference appears to show that the feedback-assessed group took the peer feedback more seriously from the outset because the importance of the feedback was demonstrated to them through the fact that it was to be worth 25% of each essay grade. In contrast, students in the product-assessed group knew peer feedback would not have any weight in their final grades.

In the second semester, on the other hand, we see that the feedback-assessed group wrote more different feedback points, made more marks on drafts, wrote more comments, more words overall and longer comments; however, the difference between the accuracy level of the feedback in the two groups did not reach the level of significance. It is interesting to note, however, that while the accuracy level of the product-assessed group decreased slightly between the first semester and the second, from 72.63% accurate feedback to 71.39%, the accuracy level in the feedback-assessed group actually increased a little, from 79.53% to 84.41%. If a larger sample size had been used, this difference may

have reached the level of significance. It may have been the significantly larger quantity of feedback given by the feedback-assessed group that led to increased accuracy. Alternatively, this may be an indication of the feedback-assessed group considering their feedback more carefully, putting more effort into it as an assessed component of the course.

It has been recommended to assign a grade for peer feedback in order to encourage learners to take the peer feedback process more seriously (Ferris, 2003; Ferris & Hedgcock, 2005; Hyland, 1996; Mittan, 1989) but previously this recommendation had not been corroborated through research. When participating in peer feedback, it is to be encouraged that students give as much feedback as they can and that they try to make that feedback as specific and accurate as possible. If it can be confirmed that the assessment of peer feedback encourages students in these ways, then assessing that feedback would seem to be a worthwhile use of instructor time.

This study shows that the assessment of feedback given by peers does appear to improve both the quantity and quality of the constructive feedback learners give. In this study, the instructor explained the importance and benefits of peer feedback to both groups. However, it seems that telling students that peer feedback is an important and beneficial activity might not be enough to motivate them to really make an effort. On the other hand, assessing the feedback they give as one of the evaluation criteria may be enough to give them additional motivation necessary to try to give the best feedback they can. If teachers believe that peer feedback is important and beneficial, then it is suggested to assess the feedback given during peer review in order to motivate learners to make the most of the activity.

It has been found previously (Lundstrom & Baker, 2009) that peer reviewing others' writing improves the peer reviewer's own writing more than the writer's. In this case, assessing students on the feedback they give to their peers as well as the final draft of their own writing encourages them to make more effort in the giving of feedback and thus improves their writing skills. The purpose of this study was to ascertain whether the assessment of feedback actually makes a difference to the quantity or quality of feedback given and it was outside of the scope of the present study to measure improvement in the writing ability of the two groups. However, as it appears that assessing peer feedback does make a difference to the quantity and quality of feedback given; further research is recommended to confirm whether the assessment of feedback makes any measurable difference in terms of writing ability.

This is a small-scale study, including just 28 learners at one educational institution, within one cultural context. As such, these results may not be generalisable to learners at different kinds of institutions or in different cultural contexts. However, it has been suggested that peer feedback may be inappropriate for learners from collectivist cultures such as Japan (Nelson & Carson, 1998; Tsui & Ng, 2000) because of the cultural necessity to attend to relationships in interactions with others (Allaei & Connor, 1990; Carson & Nelson, 1994). From the data in this study it does appear that Japanese learners are able to provide helpful constructive feedback to each other in a peer situation and assessing the feedback given can serve to increase the amount and specificity of that feedback. Further research, investigating other effects of the assessment of peer feedback, would help us to better understand pedagogical effects that the assessment of peer feedback may have. Clearly, it would also be desirable for others to verify the findings of this study in other cultural contexts, with learners at different proficiency levels or with a larger number of participants.

References

- Allaei, S. & Connor, U. (1990). Exploring the dynamics of cross-cultural collaboration in writing classrooms. *The Writing Instructor*, 10, 19-28.
- Baker, M., McQuade, D., Sommers, N., & Tratner, M. (1989). *Resources for teaching student writers at work and in the company of other writers: The Bedford prizes* (pp. 102-107). New York: St. Martin's Press.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow: Longman.
- Berg, C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215-241.
- Berger, V. (1990). The effects of peer and self-feedback. *The CATESOL Journal*, 3, 21-35.
- Carson, J. & Nelson, G. (1994). Writing groups: Cultural issues. *Journal of Second Language Writing*, 3, 17-30.
- Chandler, J. (2003). The efficacy of various different kinds of feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267-296.
- Dias, J. (2000). Learner autonomy in Japan: Transforming 'Help yourself' from threat to invitation. *Computer Assisted Language Learning*, 13(1), 49-64.
- Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315-339.

- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1-11.
- Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah: Lawrence Erlbaum Associates.
- Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 32, 181-201.
- Ferris, D. & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process, practice*. Mahwah: Lawrence Erlbaum Associates.
- Gere, A. (1987). *Writing groups: History, theory and implications*. Carbondale and Edwardsville: Southern Illinois University Press.
- Grabe, W. & Kaplan, R. (1996). *Theory and practice of writing*. Harlow: Longman.
- Hirose, K. (2008). Peer feedback in L2 English writing instruction. In K. Bradford-Watts, T. Muller & M. Swanson (Eds.), *JALT2007 Conference proceedings* (pp. 543-552). Tokyo: JALT. Available at <http://jalt-publications.org/archive/proceedings/2007/E072.pdf>.
- Huff, R. & Kline, C. (1987). *The contemporary writing curriculum: Rehearsing, composing, and valuing*. New York: Teachers College Press.
- Huizingh, E. (2007). *Applied statistics with SPSS*. London: Sage.
- Hyland, K. (1996). *Second language writing*. New York: Cambridge University Press.
- Hyland, F. (2000). ESL writers and feedback: giving more autonomy to students. *Language Teaching Research*, 4(1), 33-54.
- Hyland, F. (2001). Providing effective support: Investigating feedback to distance language learners. *Open Learning*, 16(3), 233-247.
- Min, H. (2006). The effects of trained peer-review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15, 118-141.
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System*, 25(4), 465-477.
- Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp 57-68). Cambridge: Cambridge University Press.
- Liu, J. & Hansen, J. (2002). *Peer response in second language writing classrooms*. Ann Arbor: University of Michigan Press.
- Lundstrom, K. & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30-43.

- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In Johnson, D. and Roen, D. (Eds.). *Richness in writing: Empowering ESL students* (pp. 207-219). New York: Longman.
- Myers, S. (1997). Teaching writing as a process and teaching sentence-level syntax: Reformulation as ESL composition feedback. *TESL-EJ*, 2(4). Retrieved 6 June, 2013, from: <http://www.tesl-ej.org/wordpress/issues/volume2/ej08/ej08a2/?wscsr=>.
- Nelson, G. & Carson, J. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing*, 7(2), 113-131.
- Seror, J. (2009). Institutional forces and L2 writing feedback in higher education. *The Canadian Modern Language Review*, 66(2), 203-232.
- Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1(3), 217-233.
- Tsui, A. & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9, 147-171.
- Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication*, 12(4), 492-528.

Appendix A: Unhelpful feedback sheet

Writer's name: _____ Reader's name: _____

Feedback Checklist

	<u>Yes</u>	<u>No</u>
<u>Do you understand everything clearly?</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>Is their [sic] enough information?</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>Are prepositions used correctly?</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>Does every sentence have a subject, verb and object?</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Please give one piece of constructive feedback.

I can understand everything clearly. It's easy to understand and has a lot of information. I wanna go to Bangkok!!

Appendix B Helpful feedback sheet

Writer's name: _____ Reader's name: _____

Feedback Checklist

	<u>Yes</u>	<u>No</u>
<u>Do you understand everything clearly?</u> _____	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>Is the organisation of ideas clear?</u> _____	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>Are the verb tenses correct?</u> _____	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<u>Are all words spelled correctly?</u> _____	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Please give one piece of constructive feedback.

It's interesting result. Especially Q. 2. I like this topic!! But....you have some careless mistakes. Please check these!!

Q. 1

L. 4. Same results for men and women → of

The reason for answered yes people was cool,....

→The reason which people answered yes were cool,....(Does it look strange?)

Q. 2

L. 1. Which part of the body does you like the most? → do

L. 2. the most popular among man → for ↑

L. 3. forearm and upper arm is popular among woman

→are

Q. 3

L. 2. Most of the man knows....but most woman don't know....

→ Most of the men know.....but most of the women don't know....

Q. 4

L. 5. Other answers was....→were

L. 6-7. the answer of woman....→answers of women....

L. 7. Other answer was....→Other answers were....

☞ In the last paragraph, I think you should change verb form.

→trying, playing, pushing, running, lifting, stretching, pushing, dieting