

THE EFFECTS OF PEER AND TEACHER FEEDBACK: A
LONGITUDINAL, MULTI-METHOD STUDY

Rachael Ruegg

A thesis submitted in fulfillment of the requirements for the
degree of Doctor of Philosophy.

Department of Linguistics

Macquarie University

2015

Table of Contents

Abstract	i
Statement of candidate	ii

Chapter 1. Overall Introduction

1.1 Background to the studies	1
1.1.1 Source of feedback	1
1.1.2 Focus of feedback	2
1.1.3 Length of feedback treatment	3
1.1.4 Uptake of feedback	3
1.1.5 Effect of feedback	4
1.1.6 Confidence	5
1.1.7 Assessment of feedback processes	5
1.2 Disciplinary context of the studies	6
1.3 Physical context of the studies	7
1.4 Purpose of the studies and research methods	8
1.4.1 Purpose	8
1.4.2 Research questions	10
1.4.3 Data collection	10
1.4.4 Data analysis	14
1.5 Organisation of the dissertation	18
1.5.1 Chapter two	18
1.5.2 Chapter three	19
1.5.3 Chapter four	19
1.5.4 Chapter five	20
1.5.5 Chapter six	20
1.5.6 Chapter seven	20
1.5.7 Chapter eight	20
References	21

Chapter 2. The relative effects of peer and teacher feedback on improvement in EFL students' writing performance

2.1 Introduction	28
2.2 Review of previous studies	30
2.2.1 Improvement in writing ability	30
2.2.2 Focus of feedback	32
2.2.3 Length of treatment	34
2.2.4 Studies investigating feedback from different sources	35
2.3 Research method	37
2.3.1 Context	37
2.3.2 Epistemological orientation	37
2.3.3 Participants	38

2.3.4 Feedback treatment	39
2.3.5 Assessment instrument	43
2.3.6 Analysis of feedback points	44
2.3.7 Statistical analysis	47
2.4 Results	48
2.5 Discussion and conclusion	51
References	56
Appendix 2.A Feedback form	63
Appendix 2.B Writing test essay rating scales	64

Chapter 3. Differences in the uptake of peer and teacher feedback

3.1 Introduction	65
3.2 Previous studies	69
3.2.1 Research on the uptake of feedback	69
3.2.2 The influence of explicitness of feedback	69
3.2.3 Research on uptake of peer feedback	70
3.2.4 Research comparing uptake of peer and teacher feedback	70
3.3 Methods	72
3.3.1 Participants	72
3.3.2 Treatment	73
3.3.3 Data collection	75
3.3.4 Coding	76
3.3.5 Statistical analysis	81
3.4 Results	81
3.5 Discussion and conclusion	83
References	87
Appendix 3.A Peer feedback	92
Appendix 3.B Revisions after peer feedback	93
Appendix 3.C Teacher feedback	94
Appendix 3.D Revisions after teacher feedback	95

Chapter 4. The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy

4.1 Introduction	96
4.2 Literature review	97
4.2.1 Causes and effects of self-efficacy	98
4.2.2 The effects of feedback on self-efficacy	100
4.2.3 Praise and self-efficacy	102
4.2.4 Comparisons of the effect of peer and teacher feedback on self-efficacy	103
4.3 Research method	105
4.3.1 Context	105
4.3.2 Pedagogical treatment	106

4.3.3 Participants	108
4.3.4 Analysis of self-efficacy	109
4.3.5 Analysis of feedback	112
4.4 Results	114
4.4.1 Analysis of self-efficacy	114
4.4.2 Analysis of feedback	117
4.5 Discussion	117
4.6 Conclusion	119
References	120
Appendix 4.A Pre-treatment questionnaire	127
Appendix 4.B Post-treatment questionnaire (Peer feedback group)	129
Appendix 4.C Post-treatment questionnaire (Teacher feedback group)	131

Chapter 5. The effect of assessment of peer feedback on the quantity and quality of feedback given

5.1 Introduction and background	133
5.2 Research methods	138
5.2.1 Participants and data	138
5.2.2 Treatment	141
5.2.3 Data analysis	143
5.3 Results	147
5.4 Discussion and conclusion	151
References	155
Appendix 5.A Unhelpful feedback sheet	159
Appendix 5.B Helpful feedback sheet	160

Chapter 6. The effect of assessment of process after receiving teacher feedback

6.1 Introduction	161
6.2 Review of previous studies	162
6.2.1 Literature on classroom assessment	162
6.2.2 The assessment of process in L2 writing	165
6.2.3 Investigation of uptake of feedback	166
6.3 Method	168
6.3.1 Methodology	168
6.3.2 Participants and data	169
6.3.3 Pedagogical treatment	169
6.3.4 Uptake analysis	172
6.3.5 Analysis of improvement in writing performance	174
6.4 Results	176
6.5 Discussion and conclusion	178
References	182
Appendix 6.A Feedback form	188

Appendix 6.B Writing test essay rating scales	189
---	-----

Chapter 7. The Influence of Assessment of Classroom Writing on Feedback Processes and Product vs. on Product Alone

7.1 Introduction	190
7.2 Background	193
7.3 The current study	196
7.3.1 Participants	196
7.3.2 Feedback treatment	197
7.3.3 Assessment	199
7.3.4 Instrument	200
7.3.5 Data analysis	201
7.3.6 Suitability of the data for Principal Components Analysis	201
7.4 Results	202
7.4.1 Teacher feedback groups.....	202
7.4.2 Peer feedback groups	203
7.5 Discussion and Conclusion	205
References	208
Appendix 7.A Questions concerning students' perceptions of the feedback they received (Teacher feedback group)	214
Appendix 7.B Questions concerning students' perceptions of the feedback they received (Peer feedback group)	215

Chapter 8. Overall conclusion

8.1 Summary of findings	216
8.1.1 Source of feedback	216
8.1.2 Assessment of feedback processes	217
8.2 Contribution of the research	219
8.2.1 Source of feedback	219
8.2.2 Focus of feedback	220
8.2.3 Length of feedback treatment	221
8.2.4 Uptake of feedback	223
8.2.5 Confidence	224
8.2.6 Assessment of feedback processes	225
8.3 Implications	226
8.4 Limitations	228
8.5 Suggestions for further research	230
References	231
Appendix 8.A Confirmation of final ethics approval	237

List of tables and figures

Introduction

Table 1. Research questions	11
Table 2. Data collection matrix	13
Table 3. Overall participant information	14
Table 4. Types of data collected	15

Study One

Table 1. Descriptive statistics for writing pre-test scores	48
Table 2. Descriptive statistics for writing test gain scores	49
Table 3. Descriptive statistics for types of feedback received	50

Study Two

Figure 1 Coding system	79
Table 1. Descriptive statistics for uptake, revision outcomes and types of feedback	81

Study Three

Table 1. Descriptive statistics for changes in self-efficacy	116
Table 2. Descriptive statistics for feedback	117

Study Four

Table 1. Descriptive statistics for all drafts	149
Table 2. Overview of Mann-Whitney U test results	151

Study Five

Table 1. Descriptive statistics for results of feedback	176
Table 2. Descriptive statistics for writing pre-test scores	177
Table 3. Descriptive statistics for writing test gain scores	178

Study Six

Table 1. Descriptive statistics for perceptions of feedback in the teacher feedback groups	203
Table 2. Descriptive statistics for perceptions of feedback in the peer feedback groups	204

Abstract

Despite the considerable amount of research conducted in the area of feedback on English as an additional language (L2) writing, many aspects of the effects of feedback on developing writers are still unknown. Although previous research has compared peer and teacher feedback, there are no published studies comparing peer vs. teacher feedback groups over a period of an entire academic year. Although a number of writers have recommended that teachers tailor feedback to each learner's needs, there is little published about how one would go about doing this or what effect this might have. Although in the field of L2 writing it has been suggested that teacher feedback results in increased confidence, and that peer feedback has a more positive effect on learner confidence, these views have not been corroborated through research. There is also no research comparing the uptake of peer and teacher feedback longitudinally. In fact, even "...longitudinal research on student improvement as a result of teacher feedback has been virtually nonexistent" (Ferris & Hedgcock, 2005) and recently, several writers have stressed the need for more longitudinal studies on the effects of feedback on L2 writing.

This dissertation presents longitudinal research investigating the effects of feedback tailored to individual learners. The first purpose of the research was to address the issues of the effect of source of feedback on the effect of feedback, uptake of feedback, and learner confidence. The second purpose was to address the issue of assessment of feedback processes. This was done in three ways: comparison of the quantity and quality of feedback given by students who were assessed on the feedback they gave with that of students who were assessed only on the final draft of their own essays, comparison of the uptake of teacher feedback by students who were assessed on improvement between drafts with that of students who were assessed only on the final draft of their own essays, and comparison of the students' perceptions in all different feedback and assessment conditions of the feedback they had received.

Reference

Ferris, D. & Hedgcock, J. (2005). *Teaching ESL composition*. Mahwah: Lawrence Erlbaum Associates.

STATEMENT OF CANDIDATE

I hereby certify that the work reported in this thesis has not been submitted for a higher degree to any other university or institution, and also that the sources of information used, and the extent to which the work of others have been utilised, have been indicated in the thesis.

The research presented in this thesis was approved by the Macquarie University Ethics Review Committee. Ethics approval was granted by the Macquarie University Human Research Ethics Committee for this research and the reference number is 5201000118.

Rachael Ruegg (41718372)

2015

CHAPTER ONE

Overall introduction

This introductory chapter will explain the research background and the problems which prompted the six studies presented in this dissertation, before outlining the context in which the studies were conducted. It will then go on to introduce the purposes of the studies and the research questions. A brief sketch of the data collection and analysis procedures will be described, followed by a brief outline of the organization of the dissertation.

1.1 Background to the studies

Despite the considerable amount of research already conducted in the area of feedback on L2 writing, many aspects of the effects of feedback on learning writers are still not known.

1.1.1 Source of feedback

In both first language and second language writing classrooms, there has been a great deal of enthusiasm for peer feedback, especially as an alternative to teacher feedback (Ferris, 2003). A number of studies have compared the effect of written feedback from different sources (Caulk, 1994; Hedgcock & Lefkowitz, 1992; Nakanishi & Akahori, 2005 cited in Hirose, 2008; Yang, Badger & Yu, 2006). However, there appears to be only one published study comparing a peer feedback group and a teacher feedback group longitudinally.

Fox (1980) compared peer feedback with teacher feedback in terms of students' writing performance over a period of one semester. He found that there was no significant difference between the peer feedback group and the teacher feedback group in terms of their writing performance at the end of the semester and concluded that peer feedback is preferable

because the learners in the peer feedback group gained confidence over the one semester period. However, his study was not without methodological problems. Firstly, the study used a post-test only design, with no pre-test. It is therefore unclear whether the groups' performance would have differed at the beginning of the treatment period. Furthermore, the learners in the two groups carried out different writing assignments and it may have been the different writing assignments rather than the different feedback sources that lead to no significant difference in writing performance at the end of the semester.

1.1.2 Focus of feedback

The main thrust of the discussion on the issue of teacher feedback has been the “Grammar Correction Debate” between Truscott (1996, 1999, 2007) and Ferris (1999, 2004). As the name illustrates, this debate is limited to the effects of feedback on form on the improvement of learners' grammatical competence. Many studies (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Bitchener, Young & Cameron, 2005; Lalande, 1982; Fazio, 2001; Sheen, 2007) have emphasized teacher written feedback as an implicit form of grammar instruction and consequently targeted a particular grammatical form/s for feedback. There are also a number of papers which have compared different combinations of feedback on grammar and feedback on content (Fathman & Whalley, 1990; Fazio, 2001; Kepner, 1991; Semke, 1984; Sheppard, 1992). In these studies there has been a tendency for the learners who received feedback on content to improve more than those who received feedback on grammatical errors (Fathman & Whalley, 1990; Fazio, 2001; Semke, 1984; Sheppard, 1992). These studies, as well as Truscott's (1996) article show a shift of feedback away from errors and towards ideas.

There has also been debate about whether learners can attend to both language form

and content at the same time (Raimes, 1983). However, a meta-analysis of research (Biber, Nekrasova & Horn, 2011) found that a combined focus on both language form and content leads to greater gains than either focus on form or focus on content alone. A number of writers have recommended that teachers tailor their feedback to each learner's needs (Bitchener & Knoch, 2010b; Ferris, 2002; 2003; Ferris & Hedgcock, 2005; Goldstein, 2004; Hyland, K. 1996; Storch, 2010). On the other hand, little is published about how one would go about doing this, much less what effect this tailored feedback might have.

1.1.3 Length of feedback treatment

A number of researchers have investigated the effect of written feedback on writing by giving teacher feedback just once and then measuring its effect (Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Fathman & Whalley, 1990; Sheen, 2007; Sheen, Wright & Moldowa, 2009; Truscott & Hsu, 2008; Van Beuningen, De Jong & Kuiken, 2012). On the other hand, we know that interlanguage development (i.e. the development of a learner's internal grammatical system) happens gradually over a long period of time (Gass, 2003). Thus, intuitively, it appears that repeated feedback over a period of time is more likely to lead to language development. However, "...longitudinal research on student improvement as a result of teacher feedback has been virtually nonexistent" (Ferris & Hedgcock, 2005: 187) and recently, several writers have stressed the need for more longitudinal studies (Hyland, K. 2010; Storch, 2010; Van Beuningen, 2010).

1.1.4 Uptake of feedback

Many writers have claimed that feedback can only be effective if learners pay attention to it (Chandler, 2003; Chaudron, 1977) and uptake of feedback in revision is one sure sign that a learner has paid attention to that feedback. A number of studies have compared the uptake of

peer and teacher feedback when learners have received both teacher and peer feedback on the same assignment (Connor & Asenavage, 1994; Paulus, 1999; Tsui & Ng, 2000). All of these studies show that students pay more attention to teacher feedback than to peer feedback. However, these results are not surprising as intuitively, it seems clear that if students are faced with both peer and teacher feedback on the same essay, they will pay more attention to the 'expert' teacher than to their 'lay' peers. There appears not to be any research comparing the uptake of teacher and peer feedback longitudinally. It is therefore unclear when learners receive feedback from only one source over an extended period of time, whether there are any differences in the amount of attention they pay to that feedback.

1.1.5 Effect of feedback

A number of researchers (Ashwell, 2000; Bitchener, 2008; Fathman & Whalley, 1990; Ferris & Roberts, 2001) have demonstrated that teacher feedback had a significant effect on subsequent drafts of the same piece of writing. If we are to consider the effect of feedback on improvement in learners' writing performance we need to consider its effect on subsequent instances of writing, rather than on subsequent drafts of the same text (Bitchener, 2008; Ferris, 1999; 2004; 2010; Truscott, 1996; 1999; 2004). Few studies on written feedback in writing have used a pre-test, post-test research design. Most of the studies on written feedback that *have* used a pre-test, post- test research design have been experimental studies in which students received feedback only once before measuring its effect (Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Fathman & Whalley, 1990; Sheen, 2007; Sheen, Wright & Moldowa, 2009; Van Beuningen, De Jong & Kuiken, 2012). A longitudinal feedback treatment, which incorporates multiple rounds of feedback and uses a pre-test, post-test research design is likely to yield different results from such one-shot studies.

1.1.6 Confidence

Another observation in the field of L2 writing is that increased teacher feedback results in decreased learner confidence (Cleary, 1990; Cohen, 1987; Ferris, 2002; Krashen, 1982; Semke, 1982; Truscott, 1996) or that too much constructive teacher feedback leads to students feeling less confident (Andrade & Evans, 2013; Bitchener & Ferris, 2012; Cleary, 1990; Ferris, 2002; Hyland, F. 1998; Hyland & Hyland, 2001; Krashen, 1982; Semke, 1982). However, these views have not been corroborated through research. Some writers in the field of L2 writing have also suggested that peer feedback has a more positive impact on learners' confidence than teacher feedback (Chaudron, 1984; Ferris, 2003; Ferris & Hedgcock, 2005; Gungle & Taylor, 1989; Leki, 1990; Martinez, Kock & Cass, 2011; Mittan, 1989; Tsui & Ng, 2000; Zhang, 1995). Again, this view has not been corroborated through research.

1.1.7 Assessment of feedback processes

In this dissertation, 'process writing' refers to any situation in which writing receives feedback and there is an opportunity for the writing to be revised on the basis of that feedback. 'Feedback processes' refers to the provision of peer feedback and the revision of drafts after receiving teacher feedback. A great deal of the literature on classroom assessment of writing within the process approach describes ways of assessing the final drafts of essays (e.g. Archibald, 2011; Ferris & Hedgcock, 2005). On the other hand, it is extremely difficult to find any literature discussing ways of assessing the feedback processes used by learners during the writing process. Despite this, many books published on the instruction of writing for second language learners recommend that the process should be assessed. While the process approach to writing is now firmly entrenched in second/foreign language writing courses, it is often used for instruction but not for assessment. That is, students are asked to focus on revision

and editing, class time is devoted to training and practice in these skills and yet students' writing ability is assessed on the final draft alone. However, if the teacher and the curriculum place importance on revision skills, then this should be communicated to students through the assessment criteria.

1.2 Disciplinary context of the studies

The research reported in this dissertation places itself within the broad field of Applied Linguistics, which can be defined as the teaching and learning of languages. More specifically, the research reported in this dissertation places itself within the field of Second Language Writing. Research in the field of Linguistics is often concerned with students learning the language alone (i.e. grammar, syntax and lexis) and such research would often be considered as research on Language Acquisition. However, the field of Second Language Writing is complex in that it is concerned with not only the teaching and learning of an additional language, but also the teaching and learning of the skill of writing. That is, rather than being concerned only with the teaching and learning of language, the field of Second Language Writing is additionally concerned with the learning of skills involved in writing, such as outlining, essay structure, providing support for ideas, referencing, giving and receiving feedback and revising. The largest number of studies referenced in this dissertation (around 60%) are from the field of Second Language Writing and therefore this is the discipline in which this dissertation belongs.

Perhaps the largest difference between the fields of Second Language Writing and Second Language Acquisition is that while a first language is usually learnt largely naturally, by simply being exposed to language from a young age, writing does not come naturally to anyone but rather is taught and learnt throughout one's education. It is this aspect of Second

Language Writing that creates common ground with the field of Composition and Rhetoric, which is concerned with the teaching and learning of writing in one's own language. A number of studies from the field of Composition and Rhetoric are also referred to in this dissertation.

Finally, as far as Second Language Writing is concerned with teaching and learning, irrespective of the subject being taught and learnt, it also has common ground with the field of Education. In this dissertation, mainly studies related to Educational Psychology and Educational Assessment were referred to, in relation to assessment practices and self-efficacy.

1.3 Physical context of the studies

The research reported in this dissertation was conducted in a private, foreign language university in central Japan. There is an incredibly wide range of academic levels of universities in Japan, and the university at which the research was conducted is roughly average in terms of the difficulty of entering. On the other hand, because it is a dedicated foreign language university, most of the students attend with very clear goals and a high level of motivation. In the first two years at the university, students take core skills classes in an 'English-only' environment in order to increase their second language proficiency. For these core classes, students are streamed into three proficiency tiers on the basis of Test of English as a Foreign Language (TOEFL) Institutional Testing Programme (ITP) listening and reading scores, as well as scores from the in-house speaking and writing tests. In the third and fourth year, they take a range of classes about language and culture and learn further languages.

The studies reported in this dissertation were conducted in the second year writing class. The participants had already received six years of English language education at the secondary level, as well as one year at the university level. The students were in four classes

in the bottom tier and, despite being in their second year of university level English language instruction, could be described as ranging from pre-intermediate to upper-intermediate proficiency level. Their (total) TOEFL ITP scores ranged from 370 to 500. Apart from a relatively low proficiency level, students in the bottom ability tier also have lower motivation levels. However, their motivation levels are still relatively high compared to university-level students who are not majoring in languages.

Japanese students have been claimed to be less accustomed to having autonomy over their learning than their counterparts from other cultural contexts (Benson, 2001; Dias, 2000) and the comparative power distance between the teacher and the student in Japan discourages students from asking too many questions (Hadley & Hadley, 1996). Because of this supposed lack of autonomy and strong educational hierarchy, they are often said to be passive, rather than proactive in the learning process. Japan has a so-called ‘collectivist’ culture and it has been suggested that peer feedback may be inappropriate for learners from collectivist cultures (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).

1.4 Purpose of the studies and research methods

The purpose of the six studies reported in this dissertation as well as the methods used in the studies to focus on the seven research gaps outlined above will be explained in this section followed by an overview of the data collection and analysis procedures and the research questions.

1.4.1 Purpose

The first purpose of this dissertation was to address the issues of the effect of source of feedback (i.e. whether it comes from the teacher or a peer, see section 1.1.1 above) on the

effect of feedback, as measured by improvement in students' timed writing performance over a period of one academic year (see section 1.1.5 above), uptake of feedback, as measured by uptake attempts and the relative success of those attempts (see section 1.1.5 above), and student confidence (see section 1.1.6 above). The second purpose of the studies was to address the issue of assessment of feedback processes (see 1.1.7 above). This was done in three ways: comparison of the quantity and quality of feedback given by students who were assessed on the feedback they gave with that of students who were assessed only on the final draft of their own essays, comparison of the uptake of teacher feedback by students who were assessed on improvement between drafts with that of students who were assessed only on the final draft of their own essays and comparison of the perceptions held by students in all different feedback and assessment conditions of the feedback they had received.

Although it has been suggested by writers that the focus of feedback should be tailored to individual learners (see section 1.1.2 above), there is very little research evidence of this practice being employed by instructors. In the research reported in this dissertation, feedback was tailored to individual learners and the method of tailoring is described in the main body of the dissertation. The dissertation describes an efficient and effective method of tailoring feedback to individual learners that promotes learner autonomy and can be implemented even with large numbers of students.

Length of feedback treatment has been another concern with feedback studies (see 1.1.3 above). Learning a language and learning to write are both incremental processes which take a long period of time. Accordingly, learning to write in a foreign language is likely to take even longer. It is unclear then, how providing feedback just once is likely to affect language learning or learning to write. Furthermore, since many instructors give feedback repeatedly over a long period of time to the same learners the results of such repeated

feedback over a long period of time are more likely to be relevant to them than the results of providing feedback only once. In the research reported in this dissertation, feedback was given 16 times over a period of one academic year and each draft received on average 7.5 feedback points each time, meaning that on average, each learner received 120 feedback points over the one year period.

1.4.2 Research questions

In order to fulfill these purposes, 14 research questions were posed and subsequently addressed in the six studies presented in this dissertation. Each study addresses one or two main research questions, while studies one, two, three and five (Chapters two, three four and six) also address one additional research question which further explores the results of the main question/s. Multimethod research entails the application of multiple research methods to investigate different but highly linked research questions. Since there are 14 different research questions addressed in this dissertation, all of which are linked, it was necessary to use a multimethod research design. Table 1 shows all 14 research questions, separated by study, with the main research questions for each study in bold type. The research gaps addressed by each study are also included in the table for reference.

1.4.3 Data collection

Four intact classes were used from which to collect data. Each class consisted of up to 30 students, which was a desirable number of students to have in each group. There were two different treatment variables investigated; the source of feedback and the method of assessment. Table 2 presents the two treatment conditions each class received.

Table 1. *Research questions*

Study	Research questions	Research gaps addressed
One. The relative effects of peer and teacher feedback on improvement in EFL students' writing performance	<ol style="list-style-type: none"> 1. Is there any difference between the amount of improvement in writing performance achieved by learners who receive systematic teacher feedback on their writing compared with those who give and receive systematic peer feedback? 2. Is there any difference in the types of feedback students received from teachers compared to from peers? 	<p>Source of feedback</p> <p>Length of feedback treatment</p> <p>Effect of feedback</p> <p>Focus of feedback</p>
Two. Differences in the uptake of peer and teacher feedback	<ol style="list-style-type: none"> 3. When receiving feedback from the same source for an extended period of time, is there any difference between the proportion of peer feedback and teacher feedback leading to revision attempts? 4. When receiving systematic feedback from the same source for an extended period of time, are students more successful at revising their texts on the basis of peer feedback or teacher feedback? 5. Are there any differences between the types of feedback given by peers and by an instructor that might explain differences in attention paid or revision success? 	<p>Source of feedback</p> <p>Length of feedback treatment</p> <p>Uptake of feedback</p> <p>Focus of feedback</p>
Three. The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy	<ol style="list-style-type: none"> 6. Are there any significant differences between the changes in self-efficacy of students who receive systematic teacher feedback on every preliminary draft of every essay over the period of a year compared to those of students who give and receive peer feedback on every preliminary draft of every essay over the same period? 7. What differences are there, if any, in the amount of constructive feedback and the amount of praise offered by the teacher and by peers that may account for differences in self-efficacy between the two groups? 	<p>Source of feedback</p> <p>Length of feedback treatment</p> <p>Confidence</p>
Four. The	8. Does the assessment of feedback given by peers	

effect of assessment of peer feedback on the quantity and quality of feedback given	<p>lead to an increase in the quantity of feedback given?</p> <p>9. Does the assessment of feedback given by peers lead to an increase in the quality of the feedback given?</p>	<p>Length of feedback treatment</p> <p>Assessment of feedback processes</p>
Five. The effect of assessment of process after receiving teacher feedback	<p>10. Does the assessment of process in addition to product lead to more revision attempts than the assessment of product alone?</p> <p>11. Does the assessment of process in addition to product lead to more successful revisions than the assessment of product alone?</p> <p>12. Does the assessment of process in addition to product affect the amount of improvement in writing ability achieved by learners over a one year period compared to the assessment of product alone?</p>	<p>Length of feedback treatment</p> <p>Assessment of feedback processes</p>
Six. The Influence of Assessment of Classroom Writing on Feedback Processes and Product vs. on Product Alone	<p>13. When receiving systematic teacher feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on both their use of the teacher's feedback and the final draft of their writing have different perceptions of the feedback they have received than those who are assessed on the final draft alone?</p> <p>14. When receiving systematic peer feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on quality and quantity of feedback they give in addition to the final draft of their writing have different perceptions of the feedback they have received than those who are assessed on the final draft alone?</p>	<p>Length of feedback treatment</p> <p>Assessment of feedback processes</p>

This method meant that in the studies that compare peer feedback and teacher feedback groups, class one and three could be grouped together and compared against class two and

four. On the other hand, when investigating the effect of the assessment of peer feedback on the feedback given (which was the definition of process-assessed for the peer feedback group), class 2 and class 4 could be compared and when investigating the effect of assessment of improvement between drafts (which was the definition of process-assessed for the teacher feedback group) on uptake of teacher feedback class 1 could be compared with class 3. In the one study that looked at learners' perceptions of the feedback they had received, data was collected from and compared between all four classes. In both the peer feedback and the teacher feedback only product-assessed groups, only the final written product was assessed.

Table 2. *Data collection matrix*

	Teacher feedback	Peer feedback
Only product-assessed	Class 1	Class 2
Process-assessed	Class 3	Class 4

The research reported in this article received ethics approval from the director of the department in which the data was collected. In order to satisfy the requirement of informed consent, a participant information sheet was created, which outlined what data would be collected from each participant and how the data would be used. The form included the information in both English and Japanese. In order to satisfy the requirement of voluntary participation, participants were asked to read the participant information sheet and then complete a consent form. The consent form had two options from which to choose – ‘consent’ and ‘do not consent’ – and was collected from every student in the two classes. Only the data from the students who had indicated that they consented to being included in the study were included in the analysis and reported in this article. Students were informed that they could withdraw from the study at any time during the semester with no questions asked. Table 3

shows the actual number of learners in each of the four classes and the number who gave consent to participate in the research.

Table 3. *Overall participant information*

Class number	Total <i>n</i> (Females: Males)	Consenting <i>n</i> (Females: Males)
Class 1	26 (18: 8)	25 (17: 8)
Class 2	26 (19: 7)	17 (13: 4)
Class 3	28 (19: 9)	17 (13: 4)
Class 4	26 (17: 9)	22 (15: 7)

Table 4 shows all the data which was collected from the consenting students for the research described in this dissertation and the number of students from whom each type of data was collected.

In addition to the data reported in table 4, structured interviews were conducted with three students from each class (12 students in total). Unfortunately, the structured nature of the interviews prevented the interviewer from rephrasing or clarifying questions when they were unclear. Not rephrasing or clarifying questions led to a lack of useful information being gleaned. Therefore, information from the interviews will not be reported in the dissertation. For future research with second language learners, it is suggested that semi-structured interviews be conducted if the ethics procedures allow them.

1.4.4 Data analysis

In study one, the scores of the peer and the teacher feedback groups on the in-house writing proficiency test at the beginning and end of the academic year were compared in order to measure improvement in writing performance between the two groups and thus find out which type of feedback is more effective pedagogically. First of all, gain scores were calculated by deducting the writing sub scores as well as the total writing score achieved at the beginning of

Table 4. *Types of data collected*

Type of data	Description	Class 1 <i>n</i>	Class 2 <i>n</i>	Class 3 <i>n</i>	Class 4 <i>n</i>
Pre-treatment test	Writing section of the in-house proficiency test, given to students at the end of the previous academic year (in January, 2010)	21	12	14	15
Post-treatment test	Writing section of the in-house proficiency test, given to students at the end of the academic year (in January, 2011)	22	14	14	14
Pre-treatment questionnaire	Given to the students on the first day of class (in April, 2010), to measure their confidence in writing	23	17	16	22
Post-treatment questionnaire	Given to the students on the last day of class (in January, 2011), to measure their confidence in writing and their perceptions of the feedback received	24	15	15	15
Assignment 1	Complete sample of the first assignment for the year (critical response), with three drafts and two feedback forms present	21	12	5	9
Assignment 2	Complete sample of the second assignment for the year (comparison/contrast essay), with three drafts and two feedback forms present	24	5	15	10
Assignment 3	Complete sample of the third assignment for the year (cause and effect essay), with three drafts and two feedback forms present	10	5	13	8
Assignment 4	Complete sample of the fourth assignment for the year (introductory section of a research paper), with three drafts and two feedback forms present	21	5	15	9
Assignment 5	Complete sample of the fifth assignment for the year (literature review section of a research paper), with three drafts and two feedback forms present	14	5	12	9
Assignment 6	Complete sample of the sixth assignment for the year (method section of a research paper), with three drafts and two feedback forms present	20	11	10	11
Assignment 7	Complete sample of the seventh assignment for the year (results section of a research paper), with three drafts and two feedback forms present	21	6	14	11
Assignment 8	Complete sample of the eighth assignment for the year (introductory section of a research paper), with three drafts and two feedback forms present	11	5	7	8

the academic year from those achieved at the end of the academic year. These gain scores were then compared between the two groups using Multivariate Analysis of Variance (MANOVA). MANOVA was used because it was necessary to compare the groups on the basis of five different variables simultaneously.

In study two, uptake was measured over a one year period with each group of students receiving feedback from just one source. The uptake of the two groups (peer feedback vs. teacher feedback) was analyzed in terms of the number of uptake attempts and the relative success of those attempts, enabling us to find out how much attention learners paid to each type of feedback when it is the only type of feedback available. A taxonomy was developed and previously trialled with students from the same population. The taxonomy involved coding each instance of feedback in terms of whether or not there was an uptake attempt (uptaken, not uptaken, non-specific) and the relative success of those uptake attempts (misunderstood, unsuccessful, successful). The results of this coding were then compared between the peer and teacher feedback groups using a Kruskal Wallis test. The Kruskal Wallis test is the non-parametric alternative to Analysis of Variance (ANOVA). In this study, the assumption of equal variances was not met for all variables; therefore a non-parametric test was required to find out whether there was any significant difference between the uptake attempts, or the success of those attempts, between groups.

Study three investigated the relative effects of peer and teacher feedback on learner self-efficacy, using self-report data from a pre-treatment and post-treatment questionnaire, conducted at the beginning and end of the academic year. The four questions designed to measure learner self-efficacy were answered on a six point Likert scale and were previously trialled with students in the same population. Principal Components Analysis with Varimax rotation was performed on both the pre-treatment and post-treatment questionnaires.

Cronbach's alpha reliability estimates were also calculated in order to determine the reliability of the questionnaire. The Likert scale responses were converted to numeric responses and then gain scores were calculated by deducting the self-efficacy scores at the beginning of the academic year from those at the end of the academic year for each student. These gain scores were then compared between the peer and teacher feedback groups using one-way ANOVA.

In order to investigate the effects of assessing the feedback given to peers in addition to the final draft, one intact class was assessed on the final draft alone (only product-assessed), while the other intact class was assessed on both the quantity and quality of feedback given to their peers and the final draft of their writing (process-assessed). The quantity and quality of feedback given to peers was analyzed and compared between the two groups using a Mann-Whitney U test. The Mann-Whitney U test is the non-parametric alternative to an independent samples t-test. In this study, the assumption of equal variance was not met for all variables. Therefore, it was necessary to employ a non-parametric test.

In order to investigate the effects of assessment of improvement between drafts in addition to the final draft, one intact class was assessed on their use of teacher feedback between drafts in addition to the final draft of their writing (process-assessed), while the other group was assessed only on the final draft of their writing (only product-assessed). The attempted uptake of teacher feedback and relative success of those attempts were then compared between the process-assessed and the only product-assessed groups using a Mann-Whitney U test, which is the non-parametric equivalent of an independent samples t-test. This test was used because the assumption of equal variance was not met for all variables.

In order to ascertain learners' perceptions of the feedback received, a post-treatment questionnaire was used. The questionnaire was previously trialled with students in the same

population. Principal Components Analysis with Varimax rotation was performed on the questionnaire. Cronbach's alpha reliability estimates were also calculated in order to determine the reliability of the questionnaire. The responses to the questionnaire were then compared between the only product-assessed and process-assessed peer feedback groups and the only product-assessed and process-assessed teacher feedback groups using non-parametric MANOVA.

1.5 Organization of the dissertation

This doctoral research was carried out using the "thesis by publication" procedure offered by Macquarie University. The dissertation is organized into eight chapters, including this overall introduction and an overall conclusion chapter. Each body chapter is a self-contained journal article, which has been accepted for publication. Although the studies are presented as different papers, the separate studies were designed to make up a whole in terms of being a part of the same overall research project and furthering our understanding of the two main foci: the relative effects of peer and teacher feedback and the effects of assessment of feedback processes in addition to final products. A brief description of the subsequent chapters will be presented below.

1.5.1 Chapter two

Chapter two reports study one "The relative effects of peer and teacher feedback on improvement in English as a Foreign Language (EFL) students' writing performance", which addresses the issue of the relative effects of peer and teacher feedback on students' improvement in writing performance over a period of one year using a pre-test, post-test research design. In order to determine whether the differences in improvement in writing performance were the result of peers and the teacher focusing on different issues with their

feedback, the types of feedback given by peers and the teacher were also compared. A different version of this article was published in 2014 with the journal *Linguistics and Education*.

1.5.2 Chapter three

Chapter three reports study two entitled “Differences in the uptake of peer and teacher feedback on writing”, which addresses the issue of differences in the uptake of peer and teacher feedback. Both uptake attempts and the relative success of those attempts were compared. Differences in uptake may result from differences in the explicitness of the feedback given by peers and the teacher, so the explicitness of feedback was also compared between the two groups. A different version of this article has been published in 2015 with *Regional English Language Centre (RELC) Journal*.

1.5.3 Chapter four

Chapter four reports study three entitled “The effect of peer and teacher feedback on changes in EFL students’ writing self-efficacy”, which addresses differences in changes in learner confidence after receiving peer and teacher feedback. The amounts of constructive feedback and praise have been claimed to affect confidence, therefore, the relative amounts of constructive feedback and praise given by peers and the teacher were also compared. A different version of this article was published in 2014 in *Language Learning Journal*.

1.5.4 Chapter five

Chapter five reports study four entitled “The effect of assessment of peer feedback on the quantity and quality of feedback given”, which addresses the effect of assessment of peer feedback on the quality and quantity of feedback given by students to their peers. A different

version of this article was published in 2014 in *Papers in Language Testing and Assessment*.

1.5.5 Chapter six

Chapter six reports study five entitled “The effect of assessment of process after receiving teacher feedback”, which addresses the effect of the assessment of improvement between drafts on uptake of teacher feedback. It has been suggested that increased uptake of teacher feedback may also result in improved writing performance; therefore, relative improvement in writing performance over the year was also compared between the two groups. A different version of this article has been published in 2015 with the journal *Assessment and Evaluation in Higher Education*.

1.5.6 Chapter seven

Chapter seven reports study six, entitled “The influence of assessment of classroom writing on feedback processes and product versus on product alone”, which addresses whether the assessment of process affects learners’ perceptions of the feedback they have received. A different version of this article has been published in 2015 in the journal *Writing and Pedagogy*.

1.5.7 Chapter eight

Chapter eight is a brief overall summary of the research findings of all six studies presented in this dissertation, the contribution made by the dissertation, implications of the findings, limitations of the research and suggestions for further research.

References

- Andrade, M., & Evans, N. (2013). *Principles and practices for response in second language writing: Developing self-regulated learners*. New York: Routledge.
- Allaei, S., & Connor, U. (1990). Exploring the dynamics of cross-cultural collaboration in writing classrooms. *The Writing Instructor*, 10, 19-28.
- Archibald, A. (n.d.). Writing in a second language. Retrieved 16th September 2011 from the Centre for Languages, Linguistics and Area Studies website:
<http://www.llas.ac.uk/resources/gpg/2175>
- Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing* 9, 227-257.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow: Longman.
- Carson, J., & Nelson, G. (1994). Writing groups: Cultural issues. *Journal of Second Language Writing*, 3, 17-30.
- Dias, J. (2000). Learner autonomy in Japan: Transforming 'Help yourself' from threat to invitation. *Computer Assisted Language Learning*, 13, 49-64.
- Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *TOEFL iBT Research Report 14*.
- Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal*, 12, 409-431.

- Bitchener, J., & Knoch, U. (2009). The value of a focussed approach to written corrective feedback. *ELT Journal*, 63, 204-211.
- Bitchener, J., & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19, 207-217.
- Bitchener, J., & Knoch, U. (2010b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31, 193-214.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of feedback on ESL student writing. *Journal of Second Language Writing*, 14, 191-205.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28, 181-188.
- 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.
- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27, 29-46.
- Chaudron, C. (1984). Evaluating Writing: Effects of feedback on revision. *RELC Journal*, 15, 1-14.
- Cleary, L. (1990). The fragile inclination to write: Praise and criticism in the classroom. *English Journal*, 79, 22-28.
- Cohen, A. (1987). Student processing of feedback on their compositions. In A. Wenden and J. Rubin (Eds.), *Learner strategies in language learning* (pp. 57-69). Englewood Cliffs,

N.J.: Prentice-Hall.

- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.
- Ferris, D., & Hedgcock, J. (2005). *Teaching ESL composition*. Mahwah: Lawrence Erlbaum Associates.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161-184.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1-11.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: The University of Michigan Press.
- Ferris, D. (2003). *Response to Student Writing*. Mahwah NJ: Lawrence Erlbaum Associates.
- Ferris, D. (2004). The "Grammar Correction" debate in L2 writing: Where are we and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing*, 13, 49-62.
- 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.
- Fox, R. (1980). Treatment of writing apprehension and its effects on composition. *Research in the teaching of English*, 14, 39-49.
- Gass, S. (2003). Input and interaction. In C. Doughty and M. Long (Eds.). *The handbook of*

- second language acquisition* (pp. 224-255). Oxford: Blackwell.
- Goldstein, L. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, 13, 63-80.
- Gungle, B., & Taylor, V. (1989). Writing apprehension and second language writers. In D. Johnson and D. Roen (Eds.). *Richness in writing: Empowering ESL students* (pp. 235-148). New York: Longman.
- Hadley, G., & Hadley, H (1996). The Culture of Learning and the Good Teacher in Japan: An Analysis of Student Views. *The Language Teacher*, 20, 53-55.
- Hedgcock, J., & Lefkowitz, N. (1992). Collective oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing* 1, 255-276.
- Hirose, K. (2008). Peer Feedback in L2 English Writing Instruction. In K. Bradford-Watts, T. Muller and M. Swanson (Eds.). *JALT2007 Conference Proceedings* (pp. 543-552). Tokyo: JALT.
- Hyland, K. (1996). *Second language writing*. New York: Cambridge University Press.
- Hyland, F. (1998). The impact of Teacher Written Feedback on Individual Learners. *Journal of Second Language Writing*, 7, 255-286.
- Hyland, F. (2010). Future directions in feedback on second language writing: Overview and research agenda. *International Journal of English Studies*, 10, 171-182.
- Hyland, F., & Hyland, K. (2001) Sugaring the Pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10, 185-212.
- Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of Second language writing skills. *Modern Language Journal*, 75, 305-313.

- Krashen, S. (1982). *Principles and practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Lalande, J. (1982). Reducing composition errors: An experiment. *Modern Language Journal* 66, 140-149.
- 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.
- Martinez, C., Kock, N., & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students' writing anxiety and writing self-efficacy. *Journal of Adolescent and Adult Literacy*, 54, 351-360.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. Johnson and D. Roen (Eds.). *Richness in writing: Empowering ESL students* (pp 207-219). New York: Longman.
- Nelson, G., & Carson, J. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing*, 7, 113-131.
- Paulus, T. (1999). The Effect of Peer and Teacher Feedback on Student Writing. *Journal of Second Language Writing*, 8, 265-289.
- Raimes, A. (1983). *How English works: A grammar handbook with readings*. New York: Cambridge University Press.
- Semke, H. (1982). Effects of the Red Pen. *Foreign Language Annals*, 17, 195-202.
- Sheen, Y. (2007). The effect of focussed written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255-283.
- Sheen, Y., Wright, D., & Moldowa, A. (2009). Differential effects of focused and unfocussed written correction on the accurate use of grammatical forms by adult ESL learners.

- System*, 37, 556-569.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, 23, 103-110.
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10, 29-46.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning* 46, 327-369.
- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8, 111-122.
- Truscott, J. (2004). Evidence and conjecture on the effects of feedback: A response to Chandler. *Journal of Second Language Writing*, 13, 337-343.
- Truscott, J. (2007). The effect of error correction of learners’ ability to write accurately. *Journal of Second Language Writing*, 16, 255-272.
- Truscott, J., & Hsu, A. (2008). Error correction, revision and learning. *Journal of Second Language Writing*, 17, 292-305.
- Tsui, A., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9, 147-171.
- Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights and future directions. *International Journal of English Studies*, 10, 1-27.
- Van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62, 1-41.
- Yang, M., Badger, R. & Yu, Z. (2006). A comparative study of peer and teacher feedback in a

Chinese EFL writing class. *Journal of Second Language Writing*, 15, 179-200.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4, 209-222.

CHAPTER TWO

Study One

The relative effects of peer and teacher feedback on improvement in EFL students' writing performance

Abstract

The purpose of this study was to ascertain the relative effects of peer and teacher feedback on students' writing performance. One group received teacher feedback on every preliminary draft over one academic year while a second group gave and received peer feedback on every preliminary draft over the same period. Improvement was measured by gains between pre-treatment and post-treatment writing test scores. While there was no significant difference between gains for organisation, vocabulary, content or total essay scores, the teacher feedback group gained significantly more in grammar scores than the peer feedback group.

Investigation of the feedback given by peers and the teacher showed that significantly more of the teacher's feedback related to meaning-level issues and content. The findings of the study suggest that it may be better for teachers to provide feedback on grammar and content, while peers provide feedback on organization and academic style.

2.1 Introduction

There has been a great deal of debate about whether teacher feedback leads to improvement in L2 writing ability (e.g. Truscott, 1996; Ferris, 1999; Truscott, 2007).

Although this topic has been the focus of numerous studies, we are yet to reach a verdict about whether or not teacher feedback is beneficial in the improvement of writing ability. For many instructors, the most viable alternative to teacher feedback would be peer feedback which has

become almost as common as teacher feedback in writing classes.

The current study represents an investigation into the relative benefit of peer and teacher written feedback. The university and instructor in focus adopt a process approach to writing. According to Keh (1990), the process approach consists of pre-writing and the writing of three or more drafts. The writing of more than one draft by definition entails revision of the ideas expressed and/or editing of the language used. Therefore, in research on the effect of teacher feedback within the process approach to writing, it is difficult to justify a control group which receives no feedback between drafts. Guenette (2007) suggests that there is no point in using a group which does not receive feedback at all as a control group because no-one proposes that no feedback at all is one of the options open for consideration. However, the absence of a control group has been the main weakness of longitudinal studies on feedback in writing (Bitchener & Ferris, 2012). Therefore, to study the effect of teacher feedback, the learners in this study were divided into two groups, one of which received systematic teacher feedback on preliminary drafts of every essay for the period of one academic year while the other received no systematic teacher feedback, instead giving and receiving systematic peer feedback on the preliminary drafts of every essay for the same period. A pre-test, post-test research design was used in order to compare the improvement in writing performance made by the two groups over the one year period and ascertain the benefit, if any, of teacher feedback.

Peers and the teacher may focus on different aspects of writing in their feedback leading to differences in improvement in writing performance of students in the two groups. For example; it has been suggested that teachers may focus on surface-level issues while peers may attend more to meaning-level issues (Connor & Asenavage, 1994; Hedgcock & Lefkowitz, 1992; Yang, Badger & Yu, 2006). In order to determine whether differences in the

types of feedback given by peers and the teacher may have led to significant differences in improvement in writing performance, the kinds of feedback given by the teacher and peers were also compared.

The research questions for the current study are:

1. Is there any difference between the amount of improvement in writing performance achieved by learners who receive systematic teacher feedback on their writing compared with those who give and receive systematic peer feedback?
2. Is there any difference in the types of feedback students received from teachers compared to from peers?

2.2 Review of previous studies

A number of previous studies used various different definitions of improvement and various permutations of teacher feedback. Moreover, some studies had methodological problems making it difficult to draw any meaningful conclusions (see Ferris, 1999; Guenette, 2007). A brief overview of these past studies will be provided in order to clarify the methods and intentions of this study.

2.2.1 Improvement in writing ability

The most important distinction which must be drawn is between improvement of a piece of writing, through revision, and improvement in the writing ability of the learner, which can only be determined in subsequent instances of writing. A number of researchers (Ashwell, 2000; Bitchener, 2008; Fathman & Whalley, 1990; Ferris & Roberts, 2001) have demonstrated that teacher feedback had a significant effect on subsequent drafts of the same piece of writing. However, if we are to consider the effects of written feedback on improvements in

learners' writing ability we need to consider its effect on subsequent instances of writing, rather than on subsequent drafts of the same text (Bitchener, 2008; Ferris, 1999; 2004; 2010; Truscott, 1996; 1999; 2004).

In order to determine the effect of feedback on improvement in writing ability, a pre-test and a post-test need to be used. However, few studies on written feedback in writing have used a pre-test, post-test research design. Most of the studies on written feedback that have used a pre-test, post-test research design have been experimental studies in which students received feedback only once before measuring its effect (Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Fathman & Whalley, 1990; Sheen, 2007; Sheen, Wright & Moldova, 2009; Van Beuningen, De Jong & Kuiken, 2012). A longitudinal study with a pre-test, post-test research design is likely to yield different results from such one-shot studies.

If a pre-test and a post-test are to be used, it is important to consider the characteristics of the tests in order to measure improvement as accurately as possible. In evaluating studies on feedback in writing, Truscott (2007) states that post-tests should be given in a different context from the research treatment. He found that many studies in classrooms also used post-treatment tests which were carried out in the same class and instigated by the same teacher as the feedback treatment, whereas if the post-tests were given in a different context from the treatment we may expect to see less effect. This is because, for example, students who have become conditioned to focus on grammatical correctness in a certain class are also likely to focus on grammatical correctness when taking a test in that class. In addition to this, it would also seem that in order to determine the effect of feedback on the improvement of writing, a direct writing test should be used, for indirect assessment of writing may lack content validity and construct validity (Grabe & Kaplan, 1996). Furthermore, Bitchener (2008) explains that

we can only validly measure improvement if the pre-test and the post-test employ the same type of writing task.

2.2.2 Focus of feedback

The main thrust of the discussion on the issue of teacher feedback was the “Grammar Correction Debate” between Truscott (1996, 1999, 2007) and Ferris (1999, 2004). As the name illustrates, this debate is limited to the effects of feedback on form on the improvement of learners’ grammatical competence. Many studies (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Bitchener, Young & Cameron, 2005; Lalande, 1982; Fazio, 2001; Sheen, 2007) have emphasized teacher written feedback as an implicit form of grammar instruction and consequently targeted a particular grammatical form/s for feedback. In fact, in all of these studies combined only a limited number of grammatical forms have been investigated. Another point which illustrates that these studies have emphasized feedback as merely a form for grammar instruction is that when assessing the benefit of such feedback, in many studies (Bitchener, 2008; Bitchener & Knoch, 2008; 2009; Bitchener, Young & Cameron, 2005; Fazio, 2001; Kepner, 1991; Lalande, 1982; Polio, Fleck & Leder, 1998; Rob, Ross & Shortreed, 1986; Sheen, 2007) the learners’ ability to use that particular form/s was/were assessed, rather than assessing their writing performance more generally. It therefore remains unclear whether teacher feedback leads to improvement in writing performance more generally or even to increased acquisition of the many grammatical forms which have not been investigated in these studies.

While the finding that it is possible to implicitly teach a particular grammar form through the use of written corrective feedback adds to our knowledge about the instruction of grammar, many instructors may find that considering the time involved in providing learners

with written corrective feedback and considering the number of different grammatical errors that most learners make, written corrective feedback is simply not worth the time involved if its only purpose is to increase grammatical accuracy in writing.

Pienemann's (1989) theory of the order of acquisition illustrates that learners will only be able to deal with feedback for which they are developmentally ready. Indeed, this was one of the reasons why Truscott (1996) first stated that teacher feedback was ineffective. However, recently many teachers and researchers alike have realized that teachers need not focus their feedback on the same issues for every student. A number of writers have recommended that teachers tailor their feedback to each learner's needs (Bitchener & Knoch, 2010b; Ferris, 2002; 2003; Ferris & Hedgcock, 2005; 2014; Goldstein, 2004; Hyland, K. 1996; Storch, 2010). Likewise, for peer feedback some writers have recommended that students create their own feedback sheets (Hyland, K. 1996; Liu & Hansen, 2002). Although the practice of tailoring peer and teacher feedback to learners' needs has been recommended in the literature, it is unclear how widely it has been applied.

Some studies have investigated the benefit of providing written teacher feedback on a range of different language use errors, rather than grammatical errors alone (Chandler, 2003; Truscott, 2007). The results of such studies appear to show that some other errors, such as orthographical and word choice errors are more responsive to feedback than grammatical errors.

Other studies have compared the effect of written teacher feedback on grammatical errors with that on content (Fathman & Whalley, 1990; Fazio, 2001; Kepner, 1991; Semke, 1984; Sheppard, 1992). In these studies there has been a tendency for the learners who received feedback on content to improve more than those who received feedback on

grammatical errors (Fathman & Whalley, 1990; Fazio, 2001; Semke, 1984; Sheppard, 1992). This could be seen as a suggestion for teachers to shift the focus of their feedback away from errors and towards ideas, which has indeed occurred in recent years.

There has also been debate about whether learners can attend to both language form and content at the same time (Raimes, 1983). However, a meta-analysis of research (Biber, Nekrasova & Horn, 2011) found that a combined focus on both language form and content leads to greater gains than either focus on form or focus on content alone. Specifically, grammatical accuracy was found to be improved after learners received feedback on both form and content. On the other hand, feedback on form alone resulted in significantly less improvement.

2.2.3 Length of treatment

Another apparent problem with a large number of the studies on feedback in writing is the length of the treatment period. Truscott (1996) claimed that any knowledge which students gained through receiving feedback would be transient and superficial knowledge and it appears that studies which only focus on improvement between drafts are looking for this kind of transient knowledge, whereas, Second Language Acquisition (SLA) theory suggests that interlanguage development happens gradually over a long period of time (e.g. Gass, 2003). A number of researchers have investigated the effect of written feedback on writing by giving teacher feedback just once and then measuring its effect (Bitchener, 2008; Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Fathman & Whalley, 1990; Sheen, 2007; Sheen, Wright & Moldowa, 2009; Truscott & Hsu, 2008; Van Beuningen, De Jong & Kuiken, 2012). Intuitively, it appears that repeated feedback over a period of time is more likely to lead to language development.

“...Longitudinal research on student improvement as a result of teacher feedback has been virtually nonexistent” (Ferris & Hedgcock, 2005; 2014) and recently, several writers have stressed the need for more longitudinal studies (Hyland, F. 2010; Storch, 2010; Van Beuningen, 2010). Bitchener and Ferris (2012) separately discuss one-shot research and longitudinal research on written corrective feedback. According to their account there have been six longitudinal studies on written corrective feedback, however, none of those studies lasted for a period of longer than one semester and according to the information reported by Bitchener and Ferris (2012), each of these studies involved four or five writing assignments, with students receiving feedback as little as four times over that semester. Although there is no authoritative definition of a longitudinal study, I would argue that receiving feedback four or five times is closer in nature to the one-shot experimental studies than to a longitudinal research design.

2.2.4 Studies investigating feedback from different sources

In both first language and second language writing classrooms, there has been a great deal of enthusiasm for peer feedback, especially as an alternative to teacher feedback (Ferris, 2003). Leki (1990), points out that studies in the first language (L1) setting have found peer feedback to be overall more effective than teacher feedback.

A number of studies have compared the effect of written feedback from different sources (Caulk, 1994; Hedgcock & Lefkowitz, 1992; Nakanishi & Akahori, 2005 (Cited in Hirose, 2008); Yang, Badger & Yu, 2006). Most of the research done in this area seems to have found that the influence of peer feedback is comparable to that of teacher feedback (Caulk, 1994; Nakanishi & Akahori, 2005 (Cited in Hirose, 2008)), while Hedgcock and Lefkowitz (1992) and Paulus (1999) found that the final drafts of essays which had been

subject to peer feedback received higher scores than those which had been subject to teacher feedback.

Several studies have investigated the revisions made by learners after receiving peer or teacher feedback and found that peer feedback leads to more meaning-level revisions, while teacher feedback leads to more surface-level revisions (Connor & Asenavage, 1994; Hedgcock & Lefkowitz, 1992; Yang, Badger & Yu, 2006). However, it is unclear whether different types of revisions were a result of different feedback or some other factor, such as students valuing peer feedback on rhetorical issues and teacher feedback on language issues.

As far as I am aware, there are no published studies comparing a peer feedback group and a teacher feedback group over a period of one academic year. Fox (1980) compared peer feedback with teacher feedback in terms of writing performance over a period of one semester. He found that there was no significant difference between the peer feedback group and the teacher feedback group in terms of their writing performance at the end of the semester and concluded that peer feedback is preferable because the learners in the peer feedback group gained confidence over the one semester period. However, his study was not without methodological problems. Firstly, the study used a post-test but no pre-test. It is therefore unclear whether the groups differed in ability at the beginning of the treatment period. Furthermore, the learners in the two groups carried out different writing assignments and it may have been the different writing assignments rather than the different feedback sources that lead to no difference in writing performance at the end of the semester.

Ultimately it is clear that despite the substantial body of research on written feedback on writing there is far from enough research evidence for teachers to be certain how to facilitate effective feedback on their students' writing. It is hoped that the present study will

contribute to this ongoing debate.

2.3 Research method

2.3.1 Context

The present study was conducted in the English department at a private foreign language university in Japan. The study was carried out within the second year writing class for a period of one academic year, which consisted of two semesters (10 months). The Advanced Writing course met twice a week for ninety minutes and each semester lasted for 14 weeks. Apart from this course, the students were taking other core classes in English; however, very little writing is done within these classes. Other, elective classes which the students were taking were mostly conducted in Japanese, with reports being written in Japanese.

The researcher was also the instructor of all of the writing classes in which participants of this study were enrolled. She attempted to teach all her classes in the same way throughout the one year period to ensure that, apart from variables related to the feedback treatment, the writing instruction was as similar as possible.

As Ferris (2003) states, feedback is only one piece of the instructional puzzle, the area of focus for feedback should also be the area of focus for the instruction. When assigning grades for writing, the instructor in this study gave equal weight to ‘Grammar and Vocabulary’, ‘Essay Structure’ and ‘Content’. In line with this, she also dedicated equal class time to these three areas.

2.3.2 Design of study

This longitudinal, quasi-experimental research takes a post-positivist perspective, employing a pre-treatment writing test, a feedback treatment which was implemented throughout the period of one academic year (10 months) and a post-treatment writing test. The researcher was also

the instructor who taught all four classes involved in the research for the entire research period.

2.3.3 Participants

The participants were 51 second year students in four classes in the English department at the university. Thirty seven (73%) of the students were females and 14 (27%) were males. Of these students, 28 were in the teacher feedback group, while the remaining 23 were in the peer feedback group. While all of the participants were Japanese university students, not all of them were Japanese nationals. One was from Thailand. The students ranged from 19 to 21 years of age.

Students at the university are streamed into three ability tiers based on their overall English proficiency level. English proficiency is measured by students' reading and listening scores on the TOEFL as well as in-house speaking and writing tests. The speaking test consists of a 10 minute group discussion, while the writing test consists of a 30 minute timed essay. In order to ensure that the students involved in the study were all at roughly the same language proficiency level, the researcher requested to teach four classes of students from the same tier. The four classes involved in this study constitute four out of five classes in the low ability stream. Their proficiency was judged to range from pre-intermediate to upper-intermediate level and their TOEFL ITP scores ranged from 370 to 500. Although it would be ideal to be able to report the range of TOEFL scores for each group, individual scores for every student were not available to the researcher and she therefore relied on anonymous self-report data and knowledge of the university's placement system in order to report the score range for the entire research population.

2.3.4 Feedback treatment

The participants constituted four intact classes and were assigned to two different treatments. It would be ideal to randomly assign students in all four classes to the two treatments. However, for logistical reasons this was not possible. Thus, two classes were assigned to the teacher feedback group and two were assigned to the peer feedback group. This decision was made on the basis of scheduling. As the instructor would need to receive drafts from one class of 30 students and provide written feedback on all 30 drafts before the next lesson, the classes which had a longer space between lessons were assigned to the teacher feedback treatment while those that had lessons closer together were assigned to the peer feedback group.

The teacher feedback group received teacher feedback on all preliminary drafts of all their essays over the one year period. On the other hand, the peer feedback group gave and received peer feedback on all preliminary drafts of all their essays over the same period. All students wrote eight assignments over the academic year. Therefore, if a student completed all drafts of all the assignments, the learners in both groups would have received systematic feedback (either peer or teacher feedback) on 16 separate drafts (two drafts of each assignment).

Students were asked to write their own questions on a feedback form (See Appendix A) which the reader would respond to. In addition to this, the feedback form included a fifth question “Give one piece of constructive feedback.” In this way, the focus of the feedback was not determined by the teacher but by each learner themselves. This meant that not only was the feedback on form tailored to a learner’s individual needs, but also that they were free to determine to what extent the feedback focused on form and to what extent it focused on rhetorical issues. It was possible, for example, that a learner could continue asking for

feedback on content and organization for the entire academic year and never receive feedback on language issues at all. Likewise, the opposite was also possible.

At the beginning of the academic year, all learners were given a list of example questions. The list included roughly equal numbers of questions relating to content, organization and lexicogrammar, because these three aspects were also equally weighted in the course grading. An example of a question relating to content is “Does the introduction arouse your interest?” An example of a question relating to organization is “Is the order of ideas logical?” An example of a question relating to lexicogrammar is “Are verb tenses used correctly?” All students were also encouraged to use their own questions rather than just selecting from the list. In fact, at the beginning of the second semester the instructor did not distribute the example questions again, rather encouraging students to think of their own questions, since they already had one semester of experience with the feedback system. The peer or teacher reader was required to answer each of the writer’s questions by checking a box (either ‘yes’ or ‘no’). In addition to this, if the answer was ‘no’ the reader was required to mark the draft where they thought the problems lay. For example; if the question was ‘Is my vocabulary varied?’ and the answer was ‘No’, the reader should then mark words which had been used repeatedly in the draft. The teacher used a colour coding system, with four different colours, one for each question asked by the writer. Peer readers were also instructed to use this colour system; however, they were inconsistent in doing so.

In addition to the four questions written by the writer, every feedback form included one additional question ‘Give one piece of constructive feedback’. The purpose of this additional question was twofold. Firstly, if the writer had chosen their questions poorly and no constructive feedback resulted from the first four questions, this gave the writer the opportunity to receive some constructive feedback for every draft. Therefore, they would have

something to revise between every draft. Secondly, it was an opportunity for the reader to point out any problems that they felt were particularly apparent in the writing and thus it contributed to the writer's understanding of their weaknesses in writing and gave them ideas of possible questions to ask in subsequent feedback cycles. 'Systematic feedback' in this article, refers to feedback given using this system and does not include one-off questions asked by learners to individual peers or the teacher.

Learners in all four classes were introduced to the purpose of feedback and a range of questions they could ask the peer/teacher reader. For example, the teacher introduced the topic of grammar and then elicited questions from students that they might want to ask someone to check in their writing. This process was repeated for vocabulary, organization and content. Students were explained clearly that they were free to ask any specific questions, but that general directions such as "Please correct my grammar" would be ignored.

In addition, the peer feedback groups received training in how to provide peer feedback. The training did not include any discussion of what aspects to focus on, since the focus of the feedback was determined by the writer rather than the reader. Rather the training focused on where to find the information necessary to answer the questions. When checking for minor grammatical errors such as article and preposition errors, for example, peer reviewers were given advice about how to use their dictionaries to find the relevant information. Similarly, they were instructed to consult sample essays in order to answer questions about organization.

During the parallel lesson in the teacher feedback groups, learners were trained in how to interpret the teacher feedback and how to go about using the feedback in their revisions. Similar advice was given to these students as to the peer feedback groups. For example, when

receiving feedback on minor grammatical issues it was explained how to consult their dictionaries to find the relevant information and how to apply that information to correct their writing. They were also asked to consult the sample essays they had received in order to find models of well-organized writing.

The instructor gave indirect feedback, using four coloured highlighters, a different colour to highlight problems in the writing relating to each of the four questions asked. She also gave one further piece of constructive feedback. Students were also asked to use the four colour coding system when giving peer feedback; however, they were inconsistent in doing so. Feedback (both teacher and peer) consisted of checking a box (Yes or No) and marking up the draft (If there were problems in the draft that related to the question asked).

For the purpose of this paper, systematic feedback means reading a draft and giving feedback related to the four questions asked by the student in addition to one further piece of constructive feedback. All students were free to use the writing support services offered by the university. For the purposes of this study, the researcher checked all the Writing Centre records in order to determine whether students had used the Writing Centre/s continually. As there were eight assignments, it was considered that if a student in the peer feedback group had been to the Writing Centre/s more than four times, they should be excluded from the analysis. Going to the Writing Centre four times is the equivalent of receiving systematic teacher feedback once, as teacher feedback consisted of four questions each time whereas students were only permitted to ask one specific question each time they visited the writing centre. One student from the peer feedback group had used the Writing Centres 11 times and one had used the Writing Centres 29 times, so these two students were excluded from the study. Apart from these two students, the maximum number of visits to the Writing Centres made by a student in the peer feedback group over the one year period was four. A Peer

Writing Forum was also available which students could visit online to give or receive additional peer feedback on writing. No student in this study had used the Peer Writing Forum to receive additional peer feedback. In addition to this, students in all four classes were free to ask questions to the instructor. However, in line with the Writing Centres' policies and in the interest of developing learner autonomy, the instructor would only answer specific questions.

2.3.5 Assessment instrument

The writing section of the in-house proficiency test was used as both the pre-test and post-test. The second year writing curriculum requires that timed writing account for 20% of the second semester grade. For the students in these four classes, the writing section of the in-house proficiency test constituted the timed writing requirement. Therefore, for these students, the writing test score made up 20% of their second semester grade.

The writing section of the test consists of a 30 minute timed essay task. The essays are each rated by two raters from a pool of around 30 instructors using four analytic rating scales (see Appendix B). Scores are assigned for organisation, vocabulary, content, and grammar, with the scores being scaled using many-facet Rasch modelling to correct for the relative strictness or leniency of each rater on each scale. The Rasch adjusted fair scores were used to limit the effect of inter-rater variance on the results. The students in the present study took the pre-test at the end of January, two months before the treatment period started and the post-test at the end of January the following year, five days after the treatment period ended.

Because the same test was used as both the pre-test and the post-test, the pre-test and post-test both employed the same task with a similar type of prompt (Argument essay) and were rated in the same way. Unfortunately, test security prevented the researcher from being able to publish the prompts. Moreover, as a part of the institutional in-house proficiency test,

the context was completely different from that in which the feedback treatment was given. The writing test is a direct writing assessment, designed to measure writing performance.

2.3.6 Analysis of feedback points

The feedback given by peers and the teacher was analysed in order to ascertain whether the types of feedback received influence improvement in writing performance. In order to analyse the types of feedback given by peers and by the teacher, all three drafts of every student essay along with their corresponding feedback sheets were photocopied before being returned to the students. Any essay for which there were not three drafts was discarded. After cleaning the data in this way, there were differing numbers of essays from each class in the data pool. There were at least five sample essays from each class for each assignment therefore it was decided to randomly select five sample essays from each class for each assignment to analyse the types of feedback given. As there were four classes and eight assignments over the one year period, this amounted to 160 sample essays in total, consisting of 320 preliminary drafts.

The researcher coded each feedback point as relating to one of six categories: vocabulary, surface-level grammar, meaning-level, structure, style, or content. The coding categories were the ones suggested in a previous study by Ruegg (2015). Lexical and grammatical errors are notoriously difficult to distinguish. Indeed, corpus linguistics have argued for the inseparability of the two. However, in writing choosing which word to use is quite a different issue from using that word correctly. As the distinction between word choice and word use is a somewhat transparent distinction to make, it was decided to use such a distinction to code feedback in this study. Feedback on vocabulary was defined as that which related to word choice. Any feedback relating to word use was classed as grammar.

An example of a feedback point relating to word choice (vocabulary) follows:

“Especially, Hayao Miyazaki’s film often has difficult stories because they sometimes refer to the war.”

Complex

Two examples of feedback points relating to word use (surface-level grammar) follow:

“These factors cause that people tend to suicide.”

“And that action have intolerable power for people.”

Surface-level grammar feedback was defined as any feedback which related to grammar when the problem did not interfere with the communication of ideas. Examples of surface-level grammar feedback are those commenting on missing or incorrect prepositions or articles, incorrect verb tenses and word order errors.

An example of three feedback points related to surface-level grammar follows:

“If the engine moves when the car is stopping, automobile exhaust has been exhausted to in the air.”

Meaning-level feedback was defined as any feedback indicating that the reader could not understand the message.

An example of a feedback point related to meaning-level feedback follows:

“Now, it is called that now is the times we don’t buy things well.”

Structure feedback related to the order in which ideas were expressed or the organisation of the essay as a whole.

An example of a feedback point (comment) relating to structure follows:

“Paragraph 2 and 3 seem to be about the same topic. Please explain these 2 paragraphs more clearly.”

Style feedback was any points relating to academic style which were not specific to the particular essay task. Examples are feedback on referencing, formatting and word processing issues.

An example of a feedback point related to style follows:

“Sarah, Rogers. (2010). What is international marriage? Retrieved September 28, 2010 from: http://www.ehow.com/facts_6722709_international_marriage_.html”

Content was defined as feedback relating to the actual ideas expressed.

An example of a comment relating to content follows:

“Please state how this information relates to your research.”

The essay drafts were coded by the researcher, who was also the instructor. Silva and Kei Matsuda (2008) have noted that feedback is often coded according to its intention, but coded by someone other than the instructor who gave the feedback. This can cause problems because the coder may misinterpret the intention of the instructor. They conclude that it is preferable for instructors to be the ones who code their own feedback. For this reason, one month later 20% of the data were randomly selected and recoded by the researcher. The Pearson correlation intra-rater reliability was 0.996.

Each draft of an essay was treated as one case and the results of all feedback points were calculated as proportions of all the feedback provided on the draft. The proportions ranged from 0 to 1. If, for example, 50% of the feedback was related to surface-level

grammar, then the resulting proportion would be 0.5, whereas if 100% of the feedback was on surface-level grammar points then the proportion would be 1.

2.3.7 Statistical Analysis

In order to ascertain whether there was any significant difference between the writing performance of the two groups at the beginning of the academic year, independent samples t-tests were conducted between the Rasch adjusted fair scores of the peer feedback group and those of the teacher feedback group. Independent samples t-tests rely on the assumptions of normal distribution and equal variance. Therefore, in addition to independent samples t-tests, Levene's test for equality of variances as well as Skewness and Kurtosis statistics were run. In order to ascertain whether the feedback treatments had any effect on improvement in the students' writing performance, gain scores were calculated by deducting the pre-test Rasch adjusted fair scores from the post-test Rasch adjusted fair scores. Some researchers in the field of linguistics have compared both pre-test scores and post-test scores between groups in order to find differences between the gains of groups of students. However, as stated by Anderson, Auquier, Hauck, Oakes, Vandaele and Weisberg (1980), when there are only two groups of students comparing pre-test scores and post-test scores separately is statistically equivalent to comparing gain scores. Subsequently, multivariate analysis of variance (MANOVA) was conducted to determine whether differences between the two groups were significant. As a parametric test, MANOVA also relies on the assumptions of normal distribution and equal variance. Therefore, Levene's test for the equality of variances as well as Skewness and Kurtosis statistics were run.

The coded data relating to the types of feedback received did not meet the assumption of normal distribution. Therefore, in order to ascertain whether there was any significant

difference in the types of feedback given by peers and by the teacher, a Mann-Whitney U Test was conducted.

2.3 Results

Although the results of a Rasch Analysis undertaken with raw data are presented in this phase of the research, the researcher was at that time unaware that further data would be required to report on the Rasch analysis itself. The files are no longer accessible due to the researcher's change of employer in the intervening years. Clearly, this might be a serious shortcoming in this particular paper. As a more experienced researcher now, the researcher certainly make sure to check carefully what data might be required and collect all the necessary data during the data collection period.

Descriptive statistics for the pre-test scores for the two groups can be seen in Table 1.

Table 1. *Descriptive Statistics for writing pre-test scores*

Score	<u>Peer feedback group (n = 23)</u>		<u>Teacher feedback group (n = 28)</u>	
	Mean	SD	Mean	SD
Organisation	2.6383	0.5672	2.8225	0.6701
Vocabulary	2.7423	0.4898	2.6950	0.5196
Content	2.9513	0.4799	2.7571	0.5284
Grammar	2.9309	0.4145	2.8039	0.4189
Total	11.2626	1.3270	11.0786	1.7077

No significant values were found by Levene's test for the equality of variances, thus the data were determined to meet the assumption of equal variances. Furthermore, all Skewness and Kurtosis values fell within the range -0.370 to 0.603. According to George and Mallery (2010), skewness and kurtosis values between -1 and +1 indicate excellent measures of normal distribution and measures of -2 to +2 indicate acceptable measures. Thus, these

data show normal distribution and equality of variance, meeting the criteria to employ parametric statistics.

Independent samples t-tests were conducted between the pre-test Rasch adjusted fair scores of the peer feedback group and the teacher feedback group and no significant difference was found between the pre-test organization scores ($t(49) = -1.046, p = 0.301$), vocabulary scores ($t(49) = 0.331, p = 0.742$), grammar scores ($t(49) = 1.082, p = 0.285$), content scores ($t(49) = 1.360, p = 0.180$) nor the total writing pre-test scores ($t(49) = 0.422, p = 0.675$). These results indicate that there was no significant difference in writing performance between the two groups at the beginning of the academic year.

Descriptive statistics for the gain scores for the two groups for each of the four rating scales, as well as the total essay scores can be seen in Table 2.

Table 2. *Descriptive Statistics for writing gain scores*

Score	<u>Peer feedback group ($n = 23$)</u>		<u>Teacher feedback group ($n = 28$)</u>	
	Mean	SD	Mean	SD
Organisation	0.7130	0.7359	0.5379	0.7246
Vocabulary	0.3670	0.7821	0.5770	0.5721
Content	0.1757	0.6600	0.4850	0.5713
Grammar	0.1154	0.6358	0.5313	0.6615
Total	1.3711	2.1142	2.1311	2.0334

In order to determine the suitability of the gain score data for parametric statistics, equality of covariance was tested using Box's M test and skewness and kurtosis were measured. Box's M test revealed no significant difference between the covariance of the groups ($p = 0.680$). Skewness and kurtosis values fell between -0.599 and 0.509. These results show that the gain score data meet the assumptions of equal covariance and normal

distribution. It is thus appropriate to conduct parametric analysis using these data.

RQ1: Is there any difference between the amount of improvement in writing performance achieved by learners who receive systematic teacher feedback on their writing compared with those who give and receive systematic peer feedback?

Multivariate analysis of variance results showed that the treatment group to which a learner belonged had a significant effect on their gain scores; Wilks' Lambda = 0.804, $F(4, 46) = 2.806$, $p = 0.036$. The tests of between subjects effects showed that the differences between the organization; $F(1) = 0.728$, $p = 0.398$, vocabulary; $F(1) = 1.224$, $p = 0.274$ and content gain scores; $F(1) = 3.219$, $p = 0.079$ as well as the total gain scores; $F(1) = 1.702$, $p = 0.198$ were not significant at the 0.05 level. On the other hand, the difference between the grammar gain scores was found to be significant at the 0.05 level; $F(1) = 5.166$, $p = 0.027$, with the teacher feedback group improving significantly more than the peer feedback group.

RQ2: Is there any difference in the types of feedback students received from teachers compared to from peers?

Table 3. *Descriptive Statistics for types of feedback received*

Type	Peer feedback group		Teacher feedback group	
	Mean	SD	Mean	SD
Lexis	0.0355	0.1178	0.0389	0.1359
Grammar	0.4323	0.3906	0.3695	0.3296
Meaning-level	0.0134	0.0582	0.2351	0.2242
Structure	0.0547	0.1701	0.0573	0.1417
Style	0.0836	0.1889	0.0938	0.1751
Content	0.1799	0.3179	0.2059	0.2555

$n = 160$

The descriptive statistics for the types of feedback received by students in the two groups can be seen in Table 3.

Skewness and kurtosis were tested for the data related to the types of feedback and they failed to meet the assumption of normal distribution. Skewness values ranged from 0.276 to 4.681, while kurtosis values ranged from -1.345 to 26.000. For this reason, a nonparametric Mann-Whitney U Test was conducted to determine whether the differences in the types of feedback received by learners in the two groups were statistically significant at the 0.05 level. The test revealed that the proportion of meaning-level feedback ($p = <0.001$) and that related to content ($p = 0.001$) were significantly different. The differences between the proportion of feedback on lexis ($p = 0.430$), surface-level grammar ($p = 0.185$), structure ($p = 0.129$) and style ($p = 0.064$) on the other hand, were not found to be statistically significant at the 0.05 level.

2.4 Discussion and conclusion

These results suggest that in the context of Japanese university students learning English, teacher feedback offers an advantage over peer feedback in terms of improvement in students' grammatical performance when writing in English.

This research investigated just 51 students within one course in one institutional context and from the same cultural background. Clearly, it is difficult to generalize these results to other contexts, specifically to other cultural contexts. Therefore, more longitudinal research is needed which investigates the relative benefits of peer and teacher feedback when given over an extended period. Ultimately, it will be beneficial to determine types of feedback which do not offer any benefit to particular types of students so that both peers and teachers

can focus their attention on those types of feedback that are beneficial to the learners in their classrooms. Nevertheless, this research does offer some conclusions and implications for the classroom.

Connor and Asenavage (1994), Hedgcock and Lefkowitz (1992) and Yang, Badger and Yu (2006) all studied students revisions after receiving feedback and found that they made more meaning-level revisions after receiving peer feedback and more surface-level revisions after receiving teacher feedback. In contrast to this, the present study investigated the types of feedback actually given by peers and by a teacher and it was found that significantly more of the teacher's feedback was related to the successful communication of ideas (both meaning-level feedback and feedback on content). Despite the fact that a higher proportion of the teacher's feedback related to meaning-level concerns, it was in the grammar scores that the teacher feedback group outperformed the peer feedback group. Biber, Nekrasova and Horn (2011) found that feedback on both content and form tended to result in increased grammatical accuracy, whereas feedback focusing on language form alone did not. In line with this, it is possible that the higher proportion of feedback on content received by the teacher feedback group in this study led to increased grammatical accuracy in their subsequent writing.

On the other hand, maybe the surface-level grammar feedback given by peers was highly inaccurate and therefore ineffective in improving grammatical accuracy. Although the proportion of teacher feedback dedicated to surface-level grammar was slightly less than that of peers, there may have been a significant difference in the accuracy of that feedback. It is outside of the scope of the present study to investigate the relative accuracy of peer and teacher feedback, but this is certainly a worthwhile avenue of future research.

The finding that significantly more of the teacher's feedback was meaning-level

feedback can be attributed to the fact that the teacher was a native speaker of English while the peers all shared a common first language. It is possible that because the students shared a common first language they could understand what their peers had written in English and in this way first language interference affected comprehensibility. For example; a sentence written in an essay in English may mean something significant when translated directly into Japanese, although in English it is difficult to understand. In this way, the shared first language could lead to students not recognizing each others' language errors as they make the same kinds of errors in their own writing. This suggests that teacher feedback is particularly important in EFL contexts, as in EFL contexts learners usually share a common first language and they may therefore be less able to provide meaning-level feedback. This feedback pointing out that parts of students' written texts could not be understood may also have given impetus for revision because students may have seen this as a more urgent problem to work on than other types of feedback, such as surface-level grammar feedback, which do not affect the successful communication of meaning.

Although much of the previous research on teacher written feedback in the field of L2 writing has found feedback on grammar to be ineffective in improving grammatical performance in subsequent writing (Fazio, 2001; Kepner, 1991; Polio, Fleck & Leder, 1998; Semke, 1984; Sheppard, 1992), it may be effective in some contexts. Much of that previous research focused on languages other than English or second language rather than foreign language contexts (Fazio, 2001; Kepner, 1991; Polio, Fleck & Leder, 1998; Semke, 1984; Sheppard, 1992). It has been stated by Tsui and Ng (2000) that EFL learners usually aim to improve their language ability in general whereas students in the English as a Second Language (ESL) context usually have the more specific aim of improving their writing. This may mean that EFL learners pay more attention to feedback on grammar while those in ESL

contexts pay more attention to feedback on rhetorical issues (Ferris, 1999; Hedgcock & Lefkowitz, 1994; Tsui & Ng, 2000). This increased attention would make feedback on grammar more effective for EFL learners than for their ESL counterparts. Context is a crucial factor when evaluating research on specific teacher interventions in the classroom. Since feedback is unlikely to be effective unless learners pay attention to it and since EFL learners pay more attention to feedback on grammar than their ESL counterparts, teacher feedback on grammar is likely to be more effective for EFL students than for ESL students.

This research suggests that repeated teacher feedback, given over an extended period, can be effective in students' efforts to improve the grammar in their subsequent language production. In terms of their grammatical performance in writing, the peer feedback group improved significantly less than the teacher feedback group. However, it should not be concluded from this that peer feedback should not be implemented as it has been found to have other benefits apart from the improvement of writing performance (e.g. Fox, 1980: Lundstrom & Baker, 2009; Mendonca & Johnson, 1994; Tsui & Ng, 2000) for example, increasing the amount of interaction and communication in the classroom as well as increasing learners' confidence and critical thinking skills. The findings of this study imply that the best suggestion may be for students to receive both peer and teacher feedback on their writing, with teachers focusing on both grammar and content in their feedback while peers focus on other aspects such as organization and academic style.

It can be concluded that teacher feedback can be more effective than peer feedback in terms of improvement in written grammar when students are given the freedom to decide what kind of feedback they would like. On the other hand, if feedback had been given using a predetermined feedback form, different results may have been found. The practice of students choosing the focus of the feedback they receive has been recommended in the literature

(Bitchener & Knoch, 2010b; Ferris, 2002; 2003; Hyland, K. 1996; Storch, 2010), and different versions of this practice have been described in the literature (e.g. Charles, 1990; Sommers, 1988; Storch & Tapper, 1997) it has received far less attention than the traditional ‘one-size-fits-all’ approach to teacher feedback. Nevertheless, limiting the feedback to a binary answer and related commentary, as done in this study, may be reduce the amount of meaningful feedback given by some students. Therefore, comparing this practice with feedback given based on a predetermined feedback form would be a worthwhile avenue for further research. One could investigate how successful the students were at selecting appropriate questions, how useful the resulting feedback was for revision or whether the students differed significantly in performance after receiving the two types of feedback.

It has been stated that too much of the recent research on feedback in writing has been done in an experimental context and both F. Hyland (2010) and Van Beuningen (2010) have called for more authentic classroom-based research on this topic. As stated by Ferris (2010), although laboratory research may show what the most effective feedback method is, in the context of an authentic classroom it may not be feasible for instructors to offer that particular type of feedback. This study was carried out in four real intact classes and both the instructor and the students were endeavouring to improve the students’ writing ability. Unfortunately, it is the reality of authentic classroom-based research that the researcher always needs to balance the requirements of the institution, the curriculum and his/her own beliefs about what is best for the students pedagogically with the requirements of the research project. The results are therefore difficult to generalize to other populations or contexts. However, in contrast to the lack of generalisability of the results of this study is the firm knowledge that the feedback practice used in this study can be implemented for an entire academic year from a practical viewpoint, taking into consideration the realities of curriculum requirements and teacher

workload. These kinds of practical considerations are something that has been missing from experimental studies.

A number of writers have recommended that teachers tailor their feedback to each learner's needs (Bitchener & Knoch, 2010b; Ferris, 2002; 2003; Ferris & Hedgcock, 2005; Goldstein, 2004; Hyland, K. 1996; Storch, 2010). On the other hand, Ferris (2002: 37) states that this "sounds challenging and cumbersome for teachers". This study has demonstrated one way in which feedback can be individualized for each learner. In addition to this, it has shown that teacher feedback alone may increase students' performance in written grammar whereas peer feedback alone may not be capable of this.

References

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9, 227-257.
- Biber, D., Nekrasova, T. & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *TOEFL iBT Research Report 14*.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, 17, 102-118.
- Bitchener, J. & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Bitchener, J. & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12, 409-431.
- Bitchener, J. & Knoch, U. (2009). The value of a focussed approach to written corrective feedback. *ELT Journal*, 63, 204-211.
- Bitchener, J. & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2

- writers with written corrective feedback. *Journal of second language writing*, 19, 207-217.
- Bitchener, J. & Knoch, U. (2010b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31, 193-214.
- Bitchener, J., Young, S. & Cameron, D. (2005). The effect of different types of feedback on ESL student writing. *Journal of Second Language Writing* 14, 191-205.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28, 181-188.
- 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.
- Charles, M. (1990). Responding to problems in written English using a student self-monitoring technique. *ELT journal*, 44(4), 286-293.
- Connor, U. & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Ellis, R., Sheen, Y., Murakami, M. & Takashima, H. (2008). The effects of focused and unfocussed written corrective feedback in an English as a foreign language context. *System*, 36, 353-371.
- Fathman, A. & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.
- Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing* 10, 235-249.

- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing* 8, 1-11.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: The University of Michigan Press.
- Ferris, D. (2003). *Response to Student Writing*. Mahwah NJ: Lawrence Erlbaum Associates.
- Ferris, D. (2004). The “Grammar Correction” debate in L2 writing: Where are we and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing* 13, 49-62.
- 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*. Mahwah: Lawrence Erlbaum Associates.
- Ferris, D. & Hedgcock, J. (2014). *Teaching L2 composition: Purpose, process and practice (3rd Edition.)*. Oxon: Routledge.
- Ferris, D. & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing* 10, 161-184.
- Fox, R. (1980). Treatment of writing apprehension and its effects on composition. *Research in the teaching of English*, 14, 39-49.
- Gass, S. (2003). Input and interaction. In Doughty, C. & Long, M. (Eds.). *The Handbook of Second Language Acquisition* (pp. 224-255). Oxford: Blackwell.
- George, D. & Mallery, P. (2014). *IBM SPSS statistics 21 step by step: A simple guide and reference*. New Jersey: Pearson Education.
- Goldstein, L. (2004). Questions and answers about teacher written commentary and student

- revision: Teachers and students working together. *Journal of Second Language Writing*, 13, 63-80.
- Grabe, W. & Kaplan, R. (1996). *Theory and practice of writing*. Harlow: Longman.
- Guenette, D. (2007). Is feedback pedagogically correct?: Research design issues in studies of feedback in writing. *Journal of Second Language Writing* 16, 40-53.
- Hedgcock, J. & Lefkowitz, N. (1992). Collective oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing* 1, 255-276.
- Hedgcock, J. & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing* 3, 141-163.
- 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.
- Hyland, K. (1996). *Second language writing*. New York: Cambridge University Press.
- Hyland, F. (2010). Future directions in feedback on second language writing: Overview and research agenda. *International Journal of English Studies*, 10, 171-182.
- Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44, 294-304.
- Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of Second language writing skills. *Modern Language Journal* 75, 305-313.
- Lalande, J. (1982). Reducing composition errors: An experiment. *Modern Language Journal* 66, 140-149.

- 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*. In D. Belcher & J. Liu (Eds.) Michigan Series on Teaching Multicultural Writers. 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.
- Mendonca, C. & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769.
- Paulus, T. (1999). The Effect of Peer and Teacher Feedback on Student Writing. *Journal of Second Language Writing* 8, 265-289.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 1, 52-79.
- Polio, C., Fleck, C. & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing* 7, 43-68.
- Raimes, A. (1983). *How English works: A grammar handbook with readings*. New York: Cambridge University Press.
- Ruegg, R. (2015). Student uptake of teacher written feedback on writing. *Asian EFL Journal*, 17(1), 36-56.
- Semke, H. (1984). Effects of the Red Pen. *Foreign Language Annals* 17, 195-202.
- Sheen, Y. (2007). The effect of focussed written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255-283.

- Sheen, Y., Wright, D. & Moldova, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37, 556-569.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal* 23, 103-110.
- Silva, T. & Kei Matsuda, P. (Eds.). (2008). *On Second Language Writing*. Mahwah: Routledge.
- Sommers, J. (1988). Behind the paper: Using the student-teacher memo. *College Composition and Communication*, 39(1), 77-80.
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10, 29-46.
- Storch, N. & Tapper, J. (1997). Student annotations: What NNS and NS university students say about their own writing. *Journal of Second Language Writing*, 6(3), 245-264.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning* 46, 327-369.
- Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing* 8, 111-122.
- Truscott, J. (2004). Evidence and conjecture on the effects of feedback: A response to Chandler. *Journal of Second Language Writing*, 13, 337-343.
- Truscott, J. (2007). The effect of error correction of learners' ability to write accurately. *Journal of Second Language Writing* 16, 255-272.
- Truscott, J. & Hsu, A. (2008). Error correction, revision and learning. *Journal of Second Language Writing*, 17, 292-305.
- Tsui, A. & Ng, M. (2000). Do Secondary L2 Writers Benefit from Peer Comments? *Journal of*

Second Language Writing 9, 147-171.

Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights and future directions. *International Journal of English Studies*, 10, 1-27.

Van Beuningen, C., De Jong, N. & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62, 1-41.

Yang, M., Badger, R. & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15, 179-200.

Appendix 2.A

Feedback form

Writer's name: _____

Reader's name: _____

Editing Checklist

	<u>Yes</u>	<u>No</u>
1. _____	<input type="checkbox"/>	<input type="checkbox"/>
2. _____	<input type="checkbox"/>	<input type="checkbox"/>
3. _____	<input type="checkbox"/>	<input type="checkbox"/>
4. _____	<input type="checkbox"/>	<input type="checkbox"/>
5. Give one piece of constructive feedback.		

Appendix 2.B

Writing test essay rating scales

	Organisation Think about: <ul style="list-style-type: none"> ▪ Coherence among ideas/sentences ▪ Overall structure 	Lexis Think about: <ul style="list-style-type: none"> ▪ Range ▪ Variety ▪ Preciseness 	Content Think about: <ul style="list-style-type: none"> ▪ Relevancy to the main idea ▪ Strength of support ▪ Convincingness 	Grammar Think about: <ul style="list-style-type: none"> ▪ Range ▪ Clarity ▪ Accuracy
1-1.5	No coherence or organisation, unconnected sentences which communicate little.	Demonstrates minimal word knowledge. Vocabulary is too limited to effectively complete task.	Information provides little or no support to the main idea. Ideas are all pretty much irrelevant. It's hard to find a main argument or if there is one it's hardly supported at all.	Many grammatical inaccuracies make message difficult to understand and/or very simple sentences with some errors.
2-2.5	Some attempts to organise information but with little connection between ideas and sentences apparent.	Uses a limited variety of vocabulary with little control. Has enough vocabulary to complete task in a very limited manner.	A little support for main idea, but mostly the information is not relevant or provides very weak support. The main argument is not convincing.	Inadequate range of grammar used to make ideas clear and/or much grammar is repetitive or inaccurate.
3-3.5	Obvious attempts to organise information though sometimes the lack of coherence between ideas and/or sentences creates ambiguity.	Uses an adequate variety of vocabulary with moderate control. Has enough vocabulary to complete task but not very effectively.	Most information is relevant but the support is quite weak. The main argument is a little convincing.	An adequate range of grammar is used, but some inaccuracies impede the clarity of the meaning of sentences.
4-4.5	The writing displays an organisational structure which enables the message to be followed. Ideas and sentences are connected but sometimes the lack of coherence might create ambiguity.	Uses a wide variety of vocabulary but there are some inaccuracies in word choice and/or formation. Has enough vocabulary to complete the task somewhat effectively.	Most or all information is relevant to the main argument. The support is quite strong. The main argument is somewhat convincing.	An adequate range of grammar is used and grammatical errors rarely impede the meaning of sentences. A fair number of errors which do not impede meaning are present.
5	The writing displays a coherent organisational structure which enables the message to be followed effortlessly at the sentence, paragraph and essay levels.	Uses a wide variety of vocabulary, quite accurately and with good control. Has enough vocabulary to complete task effectively.	All information is relevant and well-supported. The main argument is very convincing.	A wide range of grammar is used. Meaning is not negatively impacted by the grammar. A few errors which do not impede meaning may be present.

CHAPTER THREE

Study Two

Differences in the uptake of peer and teacher feedback on writing

Abstract

This study aimed to determine differences in the uptake of peer and teacher feedback after receiving feedback from only one source longitudinally. It also investigates the types of feedback given by peers and a teacher to help explain those differences. Data were collected from 64 Japanese university students in four classes. It was found that peer feedback was more often non-specific whereas teacher feedback was more often specific and more often uptaken. Moreover, while more of the teacher feedback led to misunderstandings or unsuccessful revisions, more of the peer feedback led to successful revisions. Furthermore, while the teacher gave more feedback using coded marks and general comments, peer feedback was made up of more uncoded marks and direct comments. It may increase the uptake of peer feedback if peer reviewers are trained to give specific feedback and general comments, rather than direct corrections.

3.1 Introduction

The process approach to writing consists of pre-writing and the writing of three or more drafts (Keh, 1990). The writing of more than one draft entails revision of the ideas expressed and/or editing of the language used. Usually, the revision of ideas and/or editing of language used happens as a result of feedback received. Readers who provide feedback on writing act as an audience and communicate to the writer which parts are unclear or difficult to follow. However, even when writing a single draft, many reread, revise and edit recursively rather

than writing from the beginning to the end of a draft in a linear way. The present study was carried out in a process writing class in order to determine the relative extent to which teacher and peer feedback offered between drafts are uptaken by learners when they write the subsequent draft.

Uptake is explained by Ellis (2012) as a move made by a student after receiving feedback from a teacher, whether or not that move is successful. Ellis, Basturkmen and Loewen (2001) introduce four criteria for determining whether a move can be defined as uptake: The move is student initiated, it is optional, it occurs after a student has demonstrated a gap in their knowledge and as a result of feedback from a teacher. In this study, feedback was provided by a peer or the teacher on some problem (gap) identified in their draft, students' revision attempts, which were optional and instigated by the students themselves in response to the feedback, were identified as uptake, whether or not they were successful. Therefore, the definition of uptake used in this study matches those provided by Ellis (2012) and Ellis, Basturkmen and Loewen (2001).

Literature in the field of Second Language Acquisition suggests that five factors are necessary for language acquisition: acculturated attitudes, comprehensible input, negotiated interaction, pushed output and attention (Ortega, 2008, pg. 63). Furthermore, studies on uptake of written corrective feedback have often considered the notion of noticing, because feedback is claimed to help learners notice features of the language. Schmidt's (2001) noticing hypothesis states that the more learners notice, the more they learn. However, this noticing hypothesis does not suggest that the noticing must be intentional, if learners notice something incidentally that is also considered to be 'noticing' (Ellis, 2015). An important point to note about the concept of noticing is that the means of noticing is input (Ellis, 2012, pp. 135 & 308). Similarly, Ortega (2008) discusses oral feedback and claims that 'noticing-

the-gap' occurs because learners compare what they have said with the input received from their interlocutor.

Because of the importance of input, the concept of 'noticing' would seem to be relevant to discussions of feedback when direct correction is used. However, in the present study the majority of the feedback was indirect feedback, which draws attention to learners' errors but does not provide input from which to notice. Therefore, although learners may notice some of the feedback and may fail to notice some, the linguistic concept of 'noticing' does not apply. Various different kinds of feedback were provided by the teacher and the peers in this study (direct correction, indirect feedback and commentary).

Furthermore, Schmidt (2001, pg. 5) stated that "the objects of attention and noticing are elements of the surface structure of utterances in the input". However, in the present study the feedback focussed on not only language use but also rhetorical features and even writing style. It is therefore clear that this concept would not encompass all of the features which the learners may have gained from the feedback they received in this study.

Intake is a concept with which uptake is commonly confused. Intake is defined by Reinders (2012, pg. 15) as "the crucial intermediary stage between input and acquisition". It can be seen from this definition that intake also relates directly to input and would therefore come into play in the context of direct error corrections, but not when indirect feedback is used, as in this study.

On the other hand, all of the feedback given in this study encouraged output modification. The functions of the teacher and peer feedback match the competence building functions for output that have been mentioned by Swain (2000): encouraging learners to notice problems with their current capability, to push them to revise their writing and thus to encourage

hypothesis testing.

One important area of research in L2 writing has been the relative effects of feedback from different sources. Most of these studies have compared peer feedback with teacher feedback (Caulk, 1994; Connor & Asenavage, 1994; Hedgcock & Lefkowitz, 1992; Paulus, 1999). Another important focus of research has been the relative effects of different methods of feedback. One question posed by previous studies has been whether feedback is more effective at certain different levels of explicitness than others (Bitchener, 2008; Bitchener & Knoch, 2010; Chandler, 2003; Ferris & Roberts, 2001).

Much of the research comparing the relative effects of peer and teacher feedback has been carried out after the learners received both peer and teacher feedback on different drafts of the same assignment (Connor & Asenavage, 1994; Paulus, 1999; Tsui & Ng, 2000), or even on the same draft (Caulk, 1994). It is not surprising that faced with both peer and teacher feedback, students would pay more attention to teacher feedback. After all, it is the teacher who will evaluate the writing once it is submitted.

Other research comparing peer and teacher feedback has been carried out in an experimental context, with students receiving feedback only once from a peer or the teacher and then revising the text according to the feedback received (Mendonca & Johnson, 1994). However, the results of experimental research such as Mendonca and Johnson (1994) may not hold if the same feedback method is used as a usual pedagogical practice over a longer period.

There is a need for research which examines uptake longitudinally, as most uptake research has been cross sectional in design. In addition, there is a need for research which compares feedback from different sources after students have received feedback from only one source. This will isolate the feedback source variable and make research results stronger.

Finally, there is a need for feedback which compares feedback from different sources longitudinally, as most revision behaviours take time to develop. The present study was designed in an attempt to address these research needs.

3.2 Previous studies

3.2.1 Research on the uptake of feedback

Researchers have argued against the validity of investigating the uptake of feedback because uptake has not been shown to lead to language acquisition (Mackey & Philp, 1998; Ohta, 2000). On the other hand, a number of researchers have used uptake of feedback as an indicator of the effectiveness of feedback (Loewen, 2004; Lyster & Ranta, 1997; Sheen, 2004) because uptake appears to represent one step in the acquisition process. Many writers have claimed that feedback can only be effective if learners pay attention to it (e.g. Chandler, 2003) and uptake of feedback in revision is one sure sign that a learner has paid attention to that feedback. This study is not aiming to ascertain whether peer feedback or teacher feedback leads to language acquisition, but rather whether feedback from one source leads to more revision attempts.

3.2.2 The influence of explicitness of feedback

Many researchers have questioned how explicit feedback needs to be in order to culminate in uptake (a revision attempt) or successful uptake (a successful revision attempt). Chandler (2003) and Ferris and Roberts (2001) found no significant difference between revisions made after receiving direct feedback and indirect feedback.

Caulk (1994) carried out a detailed comparison of peer feedback and teacher feedback and found that whereas the teacher's comments tended to be general, the peer comments were

often more specific. On the other hand, F Hyland (2000) found that in many cases peer feedback was so non-specific that it was fundamentally impossible to link it to any revisions. If there is a significant difference between the level of explicitness of feedback given by peers and the teacher, this could help to explain a significant difference between the amount of uptake or the success of revision attempts.

3.2.3 Research on uptake of peer feedback

Some studies have investigated the effect of peer feedback on uptake and on successful uptake. Mendonca and Johnson (1994) investigated uptake after learners had received only peer feedback and found that 53% of revisions made by learners were made in response to peer feedback they had received. Villamil and Guerrero (1998) found that when learners gave each other oral feedback on writing, 74% of that feedback resulted in revisions in the subsequent draft. It remains unclear what proportion of teacher feedback might result in revisions in the same population.

3.2.4 Research comparing uptake of peer and teacher feedback

A number of studies have compared the uptake of peer and teacher feedback when learners have received both teacher and peer feedback on the same assignment. Connor and Asenavage (1994) found that after receiving both peer and teacher feedback, about 5% of revisions were made in response to peer feedback, whereas around 35% were made in response to teacher feedback. Paulus (1999) found that while 32.3% of all revisions made by students after receiving peer feedback were based on that feedback, 56.7% of the revisions made after receiving teacher feedback were based on the teacher feedback. Tsui and Ng (2000) found that most learners paid attention to less than 50% of peer feedback received but more than 50% of teacher feedback received. All of these studies show that students pay more

attention to teacher feedback than to peer feedback. However, these results are not surprising as intuitively, it seems clear that if students are faced with both peer and teacher feedback on the same essay, they will pay more attention to the ‘expert’ teacher than to their ‘lay’ peers, since, as Leki (1990) points out, it is usually the teacher who will evaluate the final product.

In a study by Chaudron (1984), learners received peer feedback and teacher feedback on two different assignments. It was found that students paid equal attention to the different feedback sources when revising assignments. There is little or no research comparing the uptake of teacher and peer feedback longitudinally. It is therefore unclear when learners receive feedback from only one source longitudinally, whether there are any differences in the proportion of that feedback that results in revision attempts.

The present study compares uptake of peer and teacher feedback. One group of students received peer feedback on every preliminary draft of every writing assignment over a period of one academic year while another group received teacher feedback on every preliminary draft over the same period. Uptake was measured by the number of feedback points leading to revision attempts in the subsequent draft and the relative success of those attempts. In order to get a deeper understanding of why uptake differed, the types and explicitness of feedback given by peers and the teacher were also compared. The research questions for the present study are:

1. When receiving feedback from the same source for an extended period of time, is there any difference between the proportion of peer feedback and teacher feedback leading to revision attempts?
2. When receiving feedback from the same source for an extended period of time, are students more successful at revising their texts on the basis of peer feedback or teacher

feedback?

3. Are there any differences between the types of feedback given by peers and by an instructor that might explain differences in attention paid or revision success?

3.3 Methods

3.3.1 Participants

The 64 participants in this study were drawn from four intact classes in the second year of the English programme at a language university in Japan. Two of these classes received teacher feedback on all preliminary drafts of all assignments over a period of one academic year, while the other two received peer feedback. All four classes had the same instructor, who was also the researcher. The instructor attempted to teach all four classes in the same way excluding the feedback process.

Of the 64 participants in this study, 28 were from the two peer feedback classes. These included 22 females (78.6%) and 6 males (21.4%). The other 36 belonged to the two teacher feedback classes. These included 26 females (72.2%) and 10 males (27.8%). Although female participants greatly outnumbered male participants, the numbers of participants of each gender are in proportion to the numbers usually found in English departments in Japan. However, in different contexts such as compulsory English classes for science majors, different results may be found because of the lower motivation levels of such students as well as the different gender distribution.

Students in the English department at the university where the study was conducted are randomly placed into one of three tiers based on their overall English proficiency level, as measured by the reading and listening sections of the TOEFL ITP combined with scores on in-

house writing and speaking tests. All four classes were part of the lowest ability tier in the department. The students TOEFL ITP scores ranged from 370 to 500. It would be ideal to include learners of different levels in order to determine to what extent proficiency effects uptake. However, in order to exclude the instructor as a variable, this research investigated students who received feedback from the same teacher. In the academic year investigated, the teacher (who was also the researcher) only taught students in the lowest proficiency tier and therefore it is beyond the scope of this study to investigate the effects of proficiency level on uptake.

3.3.2 Treatment

All learners completed eight writing assignments during the academic year, writing two preliminary drafts as well as a final draft. In the first semester, they wrote a critical essay, compare and contrast essay and a cause and effect essay. In the second semester they wrote one, long research paper which was divided into five sections, each of which was written, revised and submitted before work began on the following section. Students were taught what kinds of topics were appropriate for each rhetorical mode. Following this, each student decided what topic to focus on for each essay. In addition to submitting each section individually, students were also required to compile the sections and submit the entire paper at the end of the semester.

At the beginning of the academic year, a 90 minute training session was conducted in all classes, explaining either how to provide useful constructive feedback to peers or how to use teacher feedback in revisions. Peer reviewers were trained to use simple tools such as dictionaries and thesauruses to provide feedback on grammar and vocabulary and how to use the model essays provided by the teacher to offer feedback on essay structure. The teacher

feedback group was trained to use the same simple tools to find the information necessary to correct the problems identified by the teacher. At the beginning of the second semester, a follow-up 45 minute training session was conducted. In addition to this, the teacher monitored the classes closely during peer review sessions and provided one-to-one oral advice as needed.

Disregarding absences and failure to do homework, all learners should have received feedback 16 times during the one year period. In the peer feedback classes, students were asked to work with a different partner each time they received feedback and those partners were chosen randomly by the instructor.

In the interest of encouraging learner autonomy, all writers wrote their own four questions, which they gave to the peer or teacher reader to respond to. At the beginning of the academic year, all learners were given a list of example questions. The list included roughly equal numbers of questions relating to content, organization and lexicogrammar, because these three aspects were also equally weighted in the course grading. An example of a question relating to content is “Does the introduction arouse your interest?” An example of a question relating to organization is “Is the order of ideas logical?” An example of a question relating to lexicogrammar is “Are verb tenses used correctly?” All students were also encouraged to use their own questions rather than just selecting from the list. In fact, at the beginning of the second semester the instructor did not distribute the example questions again, rather encouraging students to think of their own questions, since they already had one semester of experience with the feedback system. The peer or teacher reader was required to answer each of the writer’s questions by checking a box (either ‘yes’ or ‘no’). In addition to this, if the answer was ‘no’ the reader was required to mark the draft where they thought the problems lay. For example; if the question was ‘Is my vocabulary varied?’ and the answer was ‘No’, the reader should then mark words which had been used repeatedly in the draft. The teacher used

a colour coding system, with four different colours, one for each question asked by the writer. Peer readers were also instructed to use this colour system; however, they were inconsistent in doing so.

In addition to the four questions written by the writer, every feedback form included one additional question ‘Give one piece of constructive feedback ‘. The purpose of this additional question was twofold. Firstly, if the writer had chosen their questions poorly and no constructive feedback resulted from the first four questions, this gave the writer the opportunity to receive some constructive feedback for every draft. Therefore, they would have something to revise between every draft. Secondly, it was an opportunity for the reader to point out any problems that they felt were particularly apparent in the writing and thus it contributed to the writer's understanding of their weaknesses in writing and gave them ideas of possible questions to ask in subsequent feedback cycles. Both preliminary drafts of every essay received feedback in the same way, while the final draft was submitted to the teacher for grading.

3.3.3 Data collection

Upon submission of the final draft for grading, students submitted all three drafts and both feedback forms (for drafts 1 and 2). Copies of all drafts of every assignment as well as the two feedback forms were made before the graded originals were returned to the students.

Following this, every assignment was checked to ensure that it included three drafts and two feedback forms. Inevitably, some students failed to write three drafts or receive feedback two times on some assignments because of absence or failure to complete their homework. Any assignment which did not include three drafts and two feedback forms was excluded from the data pool. This resulted in an uneven number of samples of each

assignment from each class being included in the data pool. As there were at least five complete samples of each assignment from each class in the data pool, it was decided to randomly select five samples of each assignment from each class for analysis; this yielded 10 samples of each assignment from the teacher feedback group and 10 from the peer feedback group.

The data used in this study therefore comprises 20 complete samples of each assignment. As learners completed eight assignments over the one year period, the complete data set includes 160 assignments (320 preliminary drafts), each with its corresponding feedback form.

3.3.4 Coding

The definition of uptake used in this study is drawn from the literature on oral feedback. Lyster and Ranta (1997) refer to two types of uptake in their study: ‘repair’ and ‘needs repair’. ‘Repair’ indicates successful uptake, while ‘needs repair’ indicates that there was an uptake attempt, rather than the learner ignoring the feedback altogether. On the other hand, Sheen (2004) discussed written feedback and distinguished only whether the learner had made some attempt at uptake or had ignored the feedback completely. Loewen (2004) discussed oral feedback and added a third one to Sheen’s (2004) two categories: no opportunity. He did this because he perceived that sometimes a teacher gave feedback to a learner then carried on talking without giving the learner a chance to react to the feedback. In line with Lyster and Ranta’s (1997) categories, Loewen (2004) further categorized uptake into ‘successful uptake’ and ‘unsuccessful uptake’.

This study distinguished three categories of uptake, similar to Loewen’s (2004) categories: ‘attempt’, ‘no attempt’ and ‘non-specific’. Attempt refers to any feedback which

led to a change in the portion of text in the subsequent draft. No attempt refers to any portion of text which received feedback but remained unchanged in the subsequent draft. The category ‘non-specific’ refers to the kinds of feedback mentioned by F. Hyland (2000), representing those comments which are so non-specific that it would be impossible to link them to any specific revisions in the subsequent draft. An example of such a feedback point follows: “You have many grammar mistakes, please proof-read your draft again.” Although such feedback may lead to the revision of the writing in terms of grammar errors, it is impossible to determine whether these revisions happened as a result of the feedback received.

Like Loewen (2004), this study further categorized the revision attempts into three outcomes: ‘misunderstood’, ‘unsuccessful’ and ‘successful’. ‘Misunderstood’ refers to revision attempts which were completely different in nature from the intention of the feedback. An example of a misunderstood feedback point follows:

Original: “It became with a factor of country music.”

Revised: “It become with a factor of country music.”

This was a lexical error, but the student misunderstood it to be a verb tense error and therefore revised the text in a way that was inconsistent with the instructor’s intention.

‘Unsuccessful’ refers to situations in which the writer attended to the feedback point but failed to attend to it completely. An example of an unsuccessful revision follows:

Original: “There are some reasons for a tendency to read a book less and less such as; communication is developed, internet is extended, and people don’t have time to read a book.”

Revised: “There are some reasons for a tendency to read a book less and less such as;

communication developed, internet extended and people don't have time to read a book."

This feedback identified verb tense errors. Although the student understood that the verb tense was incorrect, the revised version with a different verb tense is still incorrect.

'Successful' indicates that the feedback point was successfully revised in line with the intention of the feedback. An example of a successful revision follows:

Original: "I think they should precede national human rights."

Revised: "I think they should put national human rights before the law of the nation."

Although the revised sentence is not perfect, it is clearer and easier to understand than the original and it was therefore deemed to be a successful revision.

In this study, the explicitness of feedback was divided into four categories: 'coded mark', 'uncoded mark', 'direct comment' and 'general comment'. The four categories in this study meant that it was possible for a feedback point to be categorized twice, for example; a problematic portion of text may be underlined (uncoded mark) with a corresponding comment in the margin (general comment), or a code may be used in the text (coded mark) and then a suggested revision supplied in the margin (direct comment). Figure 1 illustrates the coding system used.

Appendix A shows a draft that received peer feedback. The two feedback points were coded as both 'coded feedback' and 'direct comment'. This student used the colour coding system suggested by the teacher, so the colour indicates which question (on the feedback form) the feedback relates to. The direct corrections suggested on the draft are examples of direct comments. Direct comments include both longer comments with the correct form

Figure 1. *Coding system*

Feedback	Uptake	Outcome
Uncoded mark Coded mark Specific comment General comment	No attempt	-
	Attempt	Misunderstood Unsuccessful Successful
	Non-specific	-

embedded in them, and these kinds of direct corrections. Appendix B shows the revised version of the essay in appendix A. Both of the feedback points were coded as ‘attempt’ and ‘successful’. Appendix C shows a draft that received teacher feedback. The two feedback points were provided using the instructor’s colour coding system. The colour indicates which question (on the feedback form) the feedback relates to. Both of these feedback points were coded as ‘coded feedback’. In addition to this, the teacher had written two comments on the feedback form. The first one, “The recession in Japan has only been going on for the last two years and the declining population has been a problem for about 35 years. In fact, even during Japan’s best economic time (The Bubble), declining population was a problem.” was coded as a general comment. The second one, “No-one is saying that having few immigrants is the cause of the declining population, they are saying that immigration is one solution to the problem.” was also coded as a general comment. Appendix D is the revised version of the essay in appendix C. The first coded feedback point was coded as ‘attempt’ and ‘unsuccessful’, while the second coded feedback point was coded as ‘attempt’ and ‘successful’. In terms of the comments, both were coded as ‘attempt’ and ‘successful’.

Eleven essay drafts received no constructive feedback after going through the peer feedback process. For this reason, 11 drafts were excluded from the peer feedback group and 149 essay drafts remained. The teacher feedback group included 160 drafts as every draft had received at least one constructive feedback point. These 309 drafts were used for the

evaluation of feedback and uptake. However, a further 18 drafts in the peer feedback group and 1 draft in the teacher feedback group resulted in no revision attempts even after receiving constructive feedback. For the evaluation of outcomes, these 19 drafts were excluded and the remaining 131 drafts in the peer feedback group and 159 drafts in the teacher feedback group were evaluated using the coding system described above.

The percentages of feedback points, revision attempts and outcomes falling into each category in the coding system for each draft were used as the data for statistical analysis. This practice meant that regardless of how many feedback points one draft received, each draft had the same weighting in the data. It has been stated that weaker students have different revision behaviours from stronger students (Hall, 1990; Lee & Krashen, 2002; Porte, 1996). Furthermore, it has also been suggested that weaker students may receive more feedback than stronger students (Cohen & Cavalcanti, 1990; Ferris, Pezone, Tade & Tinti, 1997). Therefore, if each feedback point was included as one case rather than each draft, the revision behaviours of weaker students would be overrepresented in the data and the stronger students underrepresented. Thus, inputting the data as a percentage for each draft ensures that each draft receives equal representation in the data.

Silva and Kei Matsuda (2008) have noted that feedback is often coded according to its intention, but coded by someone other than the instructor who gave the feedback. The data being coded by someone other than the instructor can cause problems because the coder may misinterpret the intention of the instructor. They conclude that it is preferable for instructors to be the ones who code their own feedback. For this reason, the researcher (who was also the instructor) coded the feedback herself. One month later, the researcher randomly selected and recoded 20% of the data (62 drafts) in order to determine intra-rater reliability. Intra-rater reliability was analysed using Pearson correlation and a reliability of 0.996 was found.

3.3.5 Statistical analysis

The assumption of equal variance was not met for all variables, therefore a Kruskal-Wallis test was employed to ascertain whether there were any significant differences between the number of revision attempts made, the outcomes of those attempts or the explicitness of the feedback provided by peers and by the teacher.

3.4 Results

Table 1 shows the descriptive statistics for the uptake attempts, the success of those attempts and the types of feedback given by teacher and peer reviewers.

Table 1. *Descriptive Statistics for uptake, revision outcomes and types of feedback*

Category	Teacher feedback group			Peer feedback group		
	Mean	SD	N	Mean	SD	N
No attempt	0.1529	0.1996	160	0.2416	0.3146	149
Attempt	0.8365	0.1993	160	0.6686	0.3398	149
Non-specific	0.0106	0.0418	160	0.0898	0.1913	149
Misunderstood	0.0149	0.0715	159	0.0000	0.0000	131
Unsuccessful	0.2247	0.2157	159	0.0647	0.1796	131
Successful	0.7605	0.2204	159	0.9353	0.1796	131
Coded mark	0.8504	0.2182	160	0.4343	0.3876	149
Uncoded mark	0.0018	0.0226	160	0.1610	0.2650	149
Direct comment	0.0355	0.1243	160	0.3347	0.3582	149
General comment	0.3808	0.2959	160	0.3249	0.3671	149

difference between the two groups in terms of the proportions of the following uptake

categories: attempts; ($p = <0.001$) and non-specific; ($p = <0.001$). While significantly more peer feedback was not specific enough to link to any particular revision attempts, significantly more teacher feedback led to revision attempts. However, there was no significant difference in the proportion of feedback points that led to no revision attempt despite being specific ($p =$

0.975).

RQ2: When receiving feedback from the same source for an extended period of time, are students more successful at revising their texts on the basis of peer feedback or teacher feedback?

A Kruskal-Wallis Test was also run to determine whether there was any significant difference between the relative success of revision attempts between the peer and teacher feedback groups. It was found that there was a significant difference between the number of feedback points that were misunderstood; ($p = 0.002$), unsuccessful; ($p = <0.001$) and successful ($p = <0.001$). While significantly more of the teacher feedback led to misunderstandings and unsuccessful revision attempts, significantly more of the peer feedback led to successful revision attempts.

RQ3: Are there any differences between the types of feedback given by peers and by an instructor that might explain differences in attention paid or revision success?

A further Kruskal-Wallis Test was employed to investigate whether there were any significant differences between the types of feedback given by peers and by the teacher that might explain differences between the uptake of their feedback. It was found that there was a significant difference between the proportions of uncoded marks; ($p = <0.001$), coded marks; ($p = <0.001$), general comments; ($p = <0.001$) and direct comments; ($p = <0.001$). While the teacher used significantly more coded marks and general comments than peers when giving feedback, peers used significantly more uncoded marks and direct comments in their feedback.

3.5 Discussion and conclusion

From the results of this study, it seems that teacher feedback results in more uptake than peer feedback in terms of attempting to revise their writing based on the feedback received. Although teacher feedback appears to push students to modify their output and, as mentioned by Swain (1995), this kind of pushed output may be necessary for language development, teacher feedback also led to more revision attempts after having misunderstood the intention of the feedback. One possible reason for this may be the comparative power distance between the teacher and the writer, which discourages students from asking too many questions (Hadley & Hadley, 1996). Although the learners in this study were encouraged to ask the teacher questions about the feedback they received, they may have felt that it would burden the teacher to answer too many questions and may have thus not asked even when they were unsure about the feedback they had received. Furthermore, although there was some time for revision in class, inevitably students finished their revision at home outside of class time. Although they were encouraged to e-mail the teacher or visit her in her office with questions, the relationship between teachers and students in Japan is such that they may have hesitated to do so.

After a student has received feedback from a peer there is probably more chance for discussion about the feedback leading to less instances of misunderstanding. One of the advantages of peer feedback mentioned by Chaudron (1984) is that peers are more socially supportive and therefore may be more easily approachable with questions. Another possible explanation for this difference is that if students misunderstood feedback from a peer they may have simply ignored it whereas they may have felt obliged to try to use the teacher's feedback regardless of whether they understood it or not.

Teacher feedback also led to more unsuccessful revision attempts. One possible reason for this is that the students in this study were in streamed ability tiers. Because they were at a

somewhat similar level to each other, they may have given feedback at a level at which their peers were developmentally ready to uptake. Pienemann (1989) has argued that there is a certain order of acquisition and that if a learner is not developmentally ready for the input they receive, they will not benefit from it. Similarly, Vygotsky (1986) argues that learners are more likely to learn when the input is in the 'zone of proximal development'. Another advantage of peer feedback mentioned by Chaudron (1984) is that peers give feedback at an appropriate level. In contrast to this, the students may have received feedback from the teacher for which they were not ready.

On the other hand, peer feedback led to more successful revision attempts. Peer feedback has the advantage of allowing learners to feel that they are free to critically evaluate each piece of advice and only attend to those which they feel improve the writing. However, this may also mean that they only pay attention to those pieces of advice which they are sure are good advice. Therefore, peer feedback may lead them to correct mistakes based on knowledge that they already have but not push them beyond what they already know to take risks with feedback they are unsure of. Indeed, Sengupta (1998) conducted interviews with 12 students and found that not one of them paid attention to peer feedback unless it pinpointed problems that they had already identified in their own writing. In this way, a feeling that they should pay attention to the teacher feedback may actually make it more useful in terms of language development and the development of revision skills, although it results in a lower proportion of successful revision attempts.

This study corroborates the finding of F. Hyland (2000) that peer reviewers give more non-specific feedback than a teacher. In this study, 1.1% of the teacher's feedback was non-specific, while 5.3% of peers' feedback was. Because the learners in the peer feedback group were encouraged to give as much constructive feedback as possible, in cases where they could

not find anything specific to comment on, they may have given non-specific advice in order to offer something constructive to the writer. If we take two examples from the drafts in this study we can see that this may or may not actually be useful for the writer in developing their revision skills. Non-specific comments such as the following were often given by peer reviewers: "Don't forget to proof-read your draft." These kinds of comments serve to reinforce advice the teacher has also given in class and may therefore support emergent writers in developing familiarity with the writing process. On the other hand, there were several comments similar to the following: "There are a few sentences I cannot understand." These comments were sometimes offered without any further information, such as an indication of which sentences the peer reviewer was referring to. In such cases, it is difficult to see any way in which the writer could use the feedback although the peer reviewer may have considered this to be constructive feedback.

Investigation of the types of feedback given by the teacher and peer reviewers also gives us insight into why students may pay more attention to teacher feedback than to peer feedback. Peer reviewers more often gave uncoded marks whereas the teacher more often gave coded marks. Coded marks give more information from which to revise and this offers one possible explanation as to why teacher feedback led to a larger proportion of revision attempts than peer feedback. Furthermore, the teacher gave significantly more general comments, whereas peer reviewers gave more direct comments. Direct comments are more likely to change the meaning of the text and as such may be considered appropriation (Goldstein, 2004). An unclear sentence written by a student in the peer feedback group: "They work on the computer when they eating." Was directly corrected by her peer to read: "Sometime they are not suppose to eat because they prefer to work." These two sentences appear to be very different in meaning and it would seem reasonable for the writer to feel that

the peer reviewer had changed the meaning from what she intended. This too may be one reason why a smaller proportion of peer feedback led to revision attempts.

Whether receiving peer or teacher feedback, students need to be told explicitly that the feedback is just a suggestion and that it is completely up to the writer him or herself whether he or she adopts each suggestion or not. Furthermore, in order to encourage students to give due consideration to peer feedback, it would be useful to introduce them to a range of other resources from which they can obtain a second opinion about the feedback they have received. Whether or not the peer feedback received turns out to be good advice, the process of consulting other resources is an opportunity for learning. Finally, teachers would do well to ensure that they come across as approachable and rather than simply expecting that students will approach them, it is always a good idea to make expectations as explicit as possible, especially when dealing with students from a different cultural and educational background.

Additionally, students need to be trained prior to giving peer feedback and some of the findings of this study could represent fruitful areas of training for peer reviewers. Examples of specific and non-specific feedback could be used in the training process so that students realize how difficult non-specific comments are to apply to their writing. In addition to this, direct comments and general comments could be used in order for students to see how a direct comment may change the writer's intended meaning, whereas a general comment allows the writer to maintain control of their writing. Furthermore, it may be desirable to use examples of both coded and uncoded marks during training so that students can decide for themselves which is more useful in the revision process. Ultimately, if writers find the peer feedback they receive to be useful, they are more likely to pay attention to it when revising their essays.

The participants in this study were just 64 students at a similar language proficiency

level from one educational institution in a single cultural context. Therefore, different results may be found if a similar study were conducted with a different student population.

Furthermore, there are various different ways in which feedback can be implemented and this study examined just one of those methods. Clearly, different pedagogical methods are highly likely to result in different results. However, unlike in experimental research (e.g. Bitchener, 2008; Bitchener & Knoch, 2010; Sheen, 2007), the students in this study received systematic feedback from only one source for a period of one academic year. Although receiving feedback from only one source for a period of a whole year may be unusual in real classroom contexts, the feedback and revision behaviours shown in this study are real behaviours exhibited by real learners in an authentic context. From an ecological standpoint, the practices discussed in this study are feasible, whereas some practices suggested in experimental studies would not be feasible in an authentic classroom environment because they are too time intensive for teachers and/or students.

References

- Ashwell, T. (2000). Patterns of Teacher Response to Student Writing in a Multi-Draft Composition Classroom: Is Content Feedback Followed by Form Feedback the Best Method? *Journal of Second Language Writing* 9, 227-257.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, 17, 102-118.
- Bitchener, J. & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31(2), 193-214.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28(1), 181-197.
- 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.
- Chaudron, C. (1984). Evaluating Writing: Effects of feedback on revision. *RELC Journal*, 15(2), 1-14.
- Cohen, A. & Cavalcanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.) *Second Language Writing*. Cambridge: Cambridge University Press.
- Connor, U. & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Ellis, R. (2012). *Language Teaching Research and Language Pedagogy*. Chichester: Wiley-Blackwell.
- Ellis, R. (2015). *Understanding Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R., Basturkmen, H. & Loewen, S. (2001). Doing focus on form. *System*, 30, 419-432.
- Ferris, D., Pezone, S., Tade, C. & Tinti, S. (1997). Teacher commentary on student writing: Descriptions and implications. *Journal of Second Language Writing*, 6(2), 155-182.
- Ferris, D. & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing* 10, 161-184.
- Goldstein, L. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, 13, 63-80.
- Hadley, G. & Hadley, H. (1996). 'The Culture of Learning and the Good Teacher in Japan: An Analysis of Student Views'. *The Language Teacher*, 20(9), 53-55.
- Hall, C. (1990). Managing the complexity of revising across languages. *TESOL Quarterly*, 24,

43-60.

- Hedgcock, J. & Lefkowitz, N. (1992). Collective oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing* 1(3), 255-276.
- Hyland, F. (2000). ESL writers and feedback: giving more autonomy to students. *Language Teaching Research* 4(1), 33-54.
- Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294-304.
- 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.
- Lee, S. & Krashen, S. (2002). Predictors of success in writing in English as a foreign language: Reading, revision behavior, apprehension and writing. *College Student Journal*, 36(4).
- Loewen, S. (2004). Uptake in incidental focus on form in meaning focused ESL lessons. *Language Learning*, 54(1), 153-188.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66.
- Mackey, A. & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses and red herrings. *Modern Language Journal*, 82(3), 338-356.
- Mendonca, C. & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769.
- Ohta, A. (2000). Re-thinking recasts: a learner-centred examination of corrective feedback in the Japanese language classroom. In J. Hall & L. Verplaeste (Eds.) *The construction of second and foreign language learning through classroom interaction* (pp. 47-71).

Mahwah: Erlbaum.

Ortega, L. (2008). *Understanding Second Language Acquisition*. Oxon: Routledge.

Paulus, T. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing* 8, 265-289.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 1, 52-79.

Porte, G. (1996). When writing fails: How academic context and past learning experiences shape revision. *System*, 24(1), 107-116.

Reinders, H. (2012). Towards a definition of intake in second language acquisition. *Applied Research on English Language*, 1 (2), 15-36.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Sengupta, S. (1998). Peer evaluation: 'I am not the teacher'. *ELT Journal*, 52(1), 19-26.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8(3), 263-300.

Sheen, Y. (2007). The effect of focussed written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255-283.

Silva, T. & Kei Matsuda, P. (Eds.). (2008). *On Second Language Writing*. Mahwah: Routledge.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.) *Principles and practices in the study of language* (pp. 125-144). Oxford: Oxford University Press.

Tsui, A. & Ng, M. (2000). Do Secondary L2 Writers Benefit from Peer Comments? *Journal of Second Language Writing* 9, 147-171.

Villamil, O. & Guerrero, M. (1998). Assessing the impact of peer revision on L2 writing.

Applied Linguistics, 19(4), 491-514.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: Massachusetts Institute of Technology Press.

Peer feedback

Whaling

According to the article in The Japan Times (2010, May 21) entitled “No progress with Australia on whaling, FTA”, Australia and Japan told about whaling’s FTA (that is, free-trade agreement). Japan said that eating whale meat is Japanese food culture, but Australia requested that Japan should to stop whaling in the Antarctic Ocean. Also, Australia said this issue is better through negotiations at the International Whaling Committee. However, Australia promised that reserved right to bring the this issue to the International Court of Justice, and Australia told that we hopes to conclude the accord as soon as possible to ensure the both countries’ request. This negotiation about FTA has been continued since 2007, but it doesn’t still solve.

I disagree with this problem. Solving this problem is impossible because Australia is against whaling very strong. ^{Therefore} So, it is very difficult for Japan to persuade Australia. Now, Japan was allowed whaling to search. However, if this agreement concludes, the price of whale meet would decrease because it would become no tax. ^{meat} Then, everyone may forget whales are precious. The number of whales will be decreasing more and more.

Revision after peer feedback

Whaling

According to the article in The Japan Times (2010, May 21) entitled “No progress with Australia on whaling, FTA”, Australia and Japan told about whaling’s FTA (that is, free-trade agreement). Japan said that eating whale meat is Japanese food culture, but Australia requested that Japan should to stop whaling in the Antarctic Ocean. Also, Australia said this issue is better through negotiations at the International Whaling Committee. However, Australia promised that reserved right to bring the this issue to the International Court of Justice, and Australia told that we hopes to conclude the accord as soon as possible to ensure the both countries’ request. This negotiation about FTA has been continued since 2007, but it doesn’t still solve.

I disagree with this problem. Solving this problem is impossible because Australia is against whaling very strong. Therefore, it is very difficult for Japan to persuade Australia. Now, Japan was allowed whaling to search. However, if this agreement concludes, the price of whale meat would decrease because it would become no tax. Then, everyone may forget whales are precious. The number of whales will be decreasing more and more. Also, whale meat is not popular in Japan now. Japan doesn’t need to get whales so much. So, Australia and Japan should not conclude this agreement.

Teacher feedback

Immigrants can buoy Japan

According to the article on The Japan Times, entitled “Immigrants can buoy Japan”, If Japan take many immigrants, there are many benefit. Having much immigrants will not only stops declining of population, but also increase Japanese learner and diplomatic power. Therefore, world will pay more attention to Japan and the economic will be better. In Europe, many immigrants were naturalized and the demographic condition had been much better. But Japan has a few immigrants, so population is declining. Thus, immigration is one of the most important tool for Japan.

I'm disagree with this opinion, Because Japan has no room for immigration. Recently, people are difficulty in find a job even Japanese, so it is less important than improve Japanese economic. Declining of population is caused by depression, and not having a few immigrants.

Reference

Robert, D. (2010). Immigrants can buoy Japan. Retrieved May 20, 2010 from The Japan Times Website:
<http://search.japantimes.co.jp/cgi-bin/eo20100520a1.html>

Revisions after teacher feedback

Immigrants can buoy Japan

According to the article on The Japan Times, entitled “Immigrants can buoy Japan”, If Japan take many immigrants, there are many benefit. Having much immigrants will not only stops declining of population, but also increase Japanese learner and diplomatic power. Therefore, world will pay more attention to Japan and the economic will be better. Even European states that are in better demographic condition have many immigrants, and they contribute for society. Thus, immigration is one of the most important tool for Japan.

I'm disagree with this opinion, because Japan has no room for immigration. Recently, people are difficulty in find a job even Japanese who graduated famous university. Many people takes dozens or hundreds of interview for a job, but still they can't find employment. If Japan takes immigrants now, those people will increase and it causes new problem. So immigration won't be way to solution. However, accept other country's people will be good effect. Because to increase diplomatic power, Japan have to attract the world. If Japan regarded as active diplomatic country, world will pay more attention to Japan and it connect to increase Japanese learner. But government should solve the employment problem before thinking about it. In conclusion, immigrants has possibility to buoy Japan, but Japanese government is in distress about other matter.

Reference

Dujarric, R. (2010). Immigrants can buoy Japan. Retrieved May 20, 2010 from The Japan Times Website:
<http://search.japantimes.co.jp/cgi-bin/eo20100520a1.html>

CHAPTER FOUR

Study Three

The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy

Abstract

Literature from education, educational psychology and Second Language Acquisition has consistently found that self-efficacy is a key factor leading to increased language learning success. Students with more self-efficacy are claimed to have higher motivation and to expend more effort in the learning process. On the other hand, those with less self-efficacy give up readily when they encounter difficulty. Teacher feedback on L2 writing is one of the culprits that has been claimed to decrease student L2 writing self-efficacy, whereas some have suggested that peer feedback has a more positive impact on learners' self-efficacy. The present study compares changes in self-efficacy over a period of one academic year, between two groups of Japanese university students. One group received teacher feedback on every preliminary draft for the one year period while the other group gave and received peer feedback on every preliminary draft over the same period. It was found that the teacher feedback group increased in writing self-efficacy significantly more than the peer feedback group.

4.1 Introduction

A great deal of research has been done on the issue of feedback in L2 writing. However, the majority of the research has focused solely on the effectiveness of the feedback in terms of revisions or changes in writing performance without taking into account other effects feedback might cause. There are many effects, both positive and negative, that can be caused by

feedback apart from the pedagogical effect. This study investigates the differential effects of peer feedback and teacher feedback on Japanese university students' writing self-efficacy.

Traditionally, in feedback research effects such as self-efficacy have been ignored and the effects of feedback have been evaluated only on the extent to which they increase learners' performance. It has probably been assumed that there is a one-way relationship between self-efficacy and achievement and that increased achievement will lead to increased self-efficacy. However, few would argue that increased effort leads to improved performance and more recent studies (such as Hyland, F. 1998) have found that self-efficacy plays a critical role in increasing motivation, which is important for learners to make the effort required to attain improved performance. Self-efficacy is a powerful motivational construct (Pajares, 1997). However, while many studies have been devoted to the issue of motivation, few have been devoted to self-efficacy (Gardner, 1997).

Although many in the field of L2 writing state that increased teacher feedback results in decreased confidence (e.g. Cleary, 1990; Ferris, 2002; Truscott, 1996), this position has not been corroborated through research. Since low self-efficacy limits the improvement a student can achieve in terms of academic performance and since students often lack self-efficacy in terms of writing, knowing what teaching practices lead to increased self-efficacy should be just as important to teachers of writing as knowing what teaching practices lead to increased performance in writing. The present study is intended to contribute to L2 writing research by determining the relative effects of peer and teacher feedback on Japanese university students' English language writing self-efficacy.

4.2 Literature review

Self-efficacy beliefs are metacognitive beliefs learners hold in their capability to accomplish a

task (Graham, 2007). Bandura (1986) asserts that self-efficacy beliefs are an important part of personal agency and a filter through which people interpret the world and mediate their behaviour. This same construct has been called ‘efficacy’, ‘self-confidence’, ‘self-perception of ability’, ‘perceived competence’, ‘confidence’ and ‘a subjective measure of proficiency’. For the purpose of this study, the terms ‘self-efficacy’ and ‘confidence’ will be used as synonyms. As stated by Bandura (1986) and Pajares (1996), self-efficacy is a context-specific evaluation of one's capability to perform certain tasks at a certain point of time. Therefore, the words ‘self-efficacy’ and ‘confidence’ in this paper do not refer to general self-confidence but rather to one's confidence in their abilities in English language academic writing at a specific point in time.

4.2.1 Causes and effects of self-efficacy

As stated by Aida (1994), there is a need to investigate which variables lead to increased or decreased success in language learning. Literature from the fields of L1 writing education (e.g. Daly, 1978; Martinez, Kock & Cass, 2011; McCarthy, Meier & Rinderer, 1985), educational psychology (e.g. Bong, 2001; Prat-Sala & Redford, 2010; Robbins, Lauver, Le, Davis & Langley, 2004) and Second Language Acquisition (e.g. Bown & White, 2010; Gardner, Tremblay & Masgoret, 1997; Saito & Samimy, 1996) has consistently found that self-efficacy is a key factor leading to increased success in terms of course grades, standardised test results and general proficiency measures. Because of its effect on success in language learning, self-efficacy is an important avenue of investigation for language education researchers. Indeed, Bandura (1986) has stated that teachers would be well served by paying as much attention to students' perceptions of competence (i.e. self-efficacy) as to actual competence. According to Bandura (2006) students with the same level of actual ability seem to perform differently depending on

their perception of their ability (i.e. self-efficacy). Ferris, Liu, Sinha and Senna (2013) found that confidence seemed to influence students' ability to benefit from teacher feedback and from instruction more generally. Similarly, McCarthy, Meier and Rinderer (1985: 466) state that

If writing difficulties result not only from an inability to solve writing problems but also from one's own decision that one is unable to solve them, then one important step in improving writing would be to strengthen individuals' efficacy expectations about their writing ability.

The higher level of success achieved by more self-efficacious students has been attributed to greater effort being expended by more self-efficacious learners than by their less self-efficacious counterparts (Bandura, 1989; Bong, 2001; Schunk 1991). Their success has also been attributed to higher levels of motivation amongst more self-efficacious learners (Bandura, 1986; 1993), self-efficacious learners setting higher goals for themselves (e.g. Bong, 2001; Prat-Sala & Redford, 2010; Yashima, Zenuk-Nishide & Shimizu, 2004), as well as using more strategies and using strategies more effectively (Bandura, 1993; Pajares, 2003; Zimmerman, Bandura & Martinez-Pons, 1992). Furthermore, McCarthy, Meier and Rinderer (1985), as well as Prat-Sala and Redford (2010) found that while more self-efficacious students were more likely to use deep approaches, those with less confidence were more likely to use surface-level approaches. This can have a double impact on less self-efficacious learners because they choose inadequate approaches which lead to poor performance and this poor performance can serve to make them even less confident in their ability.

Decreased self-efficacy may lead to learners changing their courses, majors (Daly & Shamo, 1978; Gungle & Taylor, 1989) or even future careers (Hyland, F. 1998). As Bandura

(1993) states, the more self-efficacious a learner is, the more different career paths he or she considers. Pajares and Johnson (1994: 327) argue that “Students who lack confidence in the skills they possess are not likely to engage in tasks where those skills are required; they will more quickly give up in the face of difficulty”. Both Eysenck (1979) and Scharzer (1986) stated that learners with less self-efficacy may be distracted from the task at hand, focussing on their perceived lack of ability, the possibility of failure and the potential opinions of others about their work, thus increasing the cognitive load and the amount of effort required to complete the task. Therefore, even when low self-efficacy does not impair performance, the greater effort required by these learners can be seen as a disadvantage. Others have stated that lower levels of self-efficacy can lead to students being preoccupied (Martinez, Kock & Cass, 2011), being absent from class (Saito & Samimy, 1996), not submitting homework (Martinez, Kock & Cass, 2011) and even cheating (Finn & Frone, 2004). According to Bandura (1986), along with high levels of confidence go a belief in one’s ability to overcome difficulties and this may be crucial for maintaining momentum in the learning process.

4.2.2 The effects of feedback on self-efficacy

In the literature on feedback in L2 writing, teacher feedback has been stated to decrease student confidence and if this is true then this could potentially also stifle efforts to increase writing ability. Truscott (1996) not only stated that teacher feedback was unnecessary, but also claimed that it actually harmed students. More specifically, one of the ways in which he claimed that teacher feedback could be harmful was by decreasing student confidence. Some other writers have suggested that the more written teacher feedback a student gets, the less confident they feel (e.g. Cleary, 1990; Ferris, 2002; Krashen, 1982). It has also been suggested that too much constructive teacher feedback leads to students feeling less confident (e.g. Andrade & Evans, 2013; Bitchener & Ferris, 2012; Ferris, 2002). In interviews, 90.5% of

the students in Zacharias' (2007) study admitted that their feelings depended on the amount of teacher feedback they received on their drafts. If they received too much feedback, students reported feeling discouraged from writing, whereas if they received little feedback they reported feeling motivated.

On the other hand, many social cognitive theorists have argued that student self-efficacy beliefs develop through increased exposure to feedback, which constitutes verbal persuasions (e.g. Graham, MacArthur, Schwartz & Page-Voth, 1992; Pajares, 2003; Schunk & Swartz, 1993). However, Schunk (1985) points out that the effects of verbal persuasions may not last long, as performance has a stronger effect on self-efficacy. Nevertheless, verbal persuasions are important to temper the effect of performance outcomes.

Another important source of efficacy information, according to Schunk (1983) is social comparison between oneself and one's peers. Bandura (1981) claims that peer models may have a more positive effect on self-efficacy than teacher modelling. Zhang (1995) reviewed the literature on feedback in L1 classrooms and concluded that when compared to peer feedback, teacher feedback lacks affective appeal. Several writers in the field of L2 writing have also suggested that peer feedback has a more positive impact on learners' confidence than teacher feedback (e.g. Chaudron, 1984; Ferris & Hedgcock, 2005; Leki, 1990). Students who participate in peer feedback may be advantaged in terms of self-efficacy because they see their peers' writing and thus are able to form more realistic expectations for their own writing (Ferris, 2003; Gungl & Taylor, 1989; Martinez, Kock & Cass, 2011). While Mittan (1989) states that learners can increase their confidence and that of their peers by sharing their difficulties and problems in writing during peer feedback, he goes on to say that when students see that their peers also have difficulty writing, they gain confidence in their own abilities. Additionally, Schunk (1983) states that when learners evaluate their own writing negatively in

comparison to their peers this can motivate them and lead to increased effort. Finally, Tsui and Ng (2000) propose that the agency created by peer feedback gives students increased confidence in their writing.

4.2.3 Praise and Self-Efficacy

A great number of second language writing experts have suggested that teachers should include praise as a part of teacher feedback (e.g. Ferris, 2003; Grabe & Kaplan, 1996; Hyland, K. 1996). Ferris and Hedgcock (2005), as well as Andrade and Evans (2013), suggest the ‘sandwich’ approach to feedback, which involves writing praise first, followed by constructive feedback, and then a final note of praise. Indeed, Ferris and Hedgcock (2005) go as far as to suggest that as much as half of all feedback given by the teacher should consist of praise. Some in the field of second language writing have suggested that students who receive more praise feel more confident in their writing ability (e.g. Ferris, 1995; Hyland, 1996; Hyland, F. 1998). Others have stated that praise increases learner motivation (e.g. Goldstein, 2004; Hyland, F. 1998; Hyland, K. 1996), which may be a result of increased self-efficacy.

However, Bandura (1986) states that it is usually easier to weaken self-efficacy through negative evaluations than it is to strengthen it through praise. Furthermore, receiving praise from a teacher is not always a positive experience. Cohen (1987) found that too much praise could lead to decreased enjoyment of writing. Likewise, Cleary (1990) found that some students came to resent the effort they spent fulfilling teachers’ expectations and subsequently lost intrinsic motivation for writing. She goes on to say that teachers' feedback, whether it is praise or advice, destroys intrinsic motivation in writing. Similarly, “Some teachers believe that providing too much praise, especially at the early stages of writing, can make students complacent and discourage revision (Hyland, K. 1996: 187).” Furthermore, although two

learners in F. Hyland's (1998) case study research supported the idea that praise leads to increased confidence, one learner found the praise insincere, not helpful and even condescending. Grabe and Kaplan (1996) also caution against giving false praise. Indeed, Hyland and Hyland (2001) claim that the 'sandwich' approach is all too obvious to students. They also found that comments of praise were more general than constructive comments and tended to be formulaic.

Some studies have been conducted to evaluate the amount of praise given by teachers and by peers. Beason (1993) found that while 35% of peer comments constituted praise, only 15% of teachers' did. Ferris, Pezone, Tade and Tinti (1997) found that 17.8% of marginal comments and 31.5% of end comments given by teachers consisted of praise. Hyland and Hyland (2001) analysed the feedback given by two teachers and found that 44% of their feedback was praise. Dragga (1986, cited in Hyland & Hyland, 2001) found that only 6% of all comments given by teachers constituted praise. It appears that few teachers are close to 50% of their feedback being praise as suggested by Ferris and Hedgcock (2005). However, it also appears that the proportion of praise given by different teachers varies widely. Hyland and Hyland (2001) suggest that, because teachers tend to give more constructive feedback and less praise, whereas peers give more praise and less constructive feedback, both teacher and peer feedback should be used, allowing learners to get sufficient quantities of both constructive feedback and praise.

4.2.4 Comparisons of the effect of peer and teacher feedback on self-efficacy

Two previous studies have investigated the effect of both teacher and peer feedback on learners' confidence in their writing ability. Fox (1980) compared two groups of students over a period of one semester. One of the groups received feedback from the teacher, while the

other received peer feedback. He found that overall, the peer feedback group increased in confidence over the semester when compared to the teacher feedback group. However, not only the feedback, but also the assessment was different between the two groups. The teacher feedback group received grades from the teacher, whereas the peer feedback group self-evaluated their writing. It may have been this different assessment, rather than the feedback itself, that increased learners' confidence.

Ruegg (2010) looked at which factors during the writing process predicted higher confidence in the final draft for Japanese university students. It was found that the only significant predictor of high confidence was the amount of peer feedback, showing that the more learners participated in peer feedback, the higher confidence they had in the finished product. However, in Ruegg's (2010) study all of the learners had received both teacher and peer feedback on every assignment. In the present study, in order to compare the unique effects of peer and teacher feedback on self-efficacy, one group of students received peer feedback alone while the other group received only teacher feedback. Furthermore, the learners in Ruegg's (2010) study were from a number of different classes with a different instructor teaching each class. Different instructors are likely to have different effects on student self-efficacy as a result of their classroom manner. On the other hand, in the present study only one instructor taught all four classes thus eliminating the instructor as a variable. Ferris (2006) asserts that "for experimental research designs" (p. 93) it is better to use only one teacher to ensure more consistency. In this way, effects of different treatments can be measured without the confounding effect of the instructor variable.

Although the connection between self-efficacy and other aspects of motivation has been empirically corroborated, less is known about the ways in which verbal persuasions and social comparisons affect self-efficacy beliefs. Pajares (1997) called for more empirical evidence of

sensible intervention strategies and practical ways to change self-efficacy beliefs. The purpose of the present study is to determine whether the source of feedback on writing has a significant effect on students' writing self-efficacy. The research questions are:

1. Are there any significant differences between the changes in self-efficacy of students who receive systematic teacher feedback on every preliminary draft of every essay over the period of a year compared to those of students who give and receive peer feedback on every preliminary draft of every essay over the same period?
2. What differences are there, if any, in the amount of constructive feedback and the amount of praise offered by the teacher and by peers that may account for differences in self-efficacy between the two groups?

4.3 Research Method

4.3.1 Context

The present study was carried out in a foreign language university in Eastern Japan. All students in the study were in their second year of study in the English department. In the first year programme, they had learnt to write paragraphs in the first semester and five-paragraph essays in the second semester. During the second year writing programme, in which this study took place, students were required to write different types of essays in the first semester and to focus on one extended research paper in the second semester. The instructor involved in this study required students to write a critical essay, a compare and contrast essay and a cause and effect essay in the first semester. In the second semester, the research paper was divided into five separate sections, with each section being drafted, revised and the final draft submitted for grading before work was commenced on the subsequent section. The second year writing programme was particularly challenging for students because of a large increase in the amount

of writing required between the first year and the second year of study.

4.3.2 Pedagogical Treatment

In total, the students were expected to complete two preliminary drafts and one final draft of each of the eight assignments over the academic year. The study was carried out in four intact second year classes. Two of the classes received systematic teacher feedback on every preliminary draft over the one year period. The other two classes gave and received systematic peer feedback on every preliminary draft of every essay over the same period. All final drafts were collected by the instructor for grading.

In the interest of encouraging learner autonomy, all writers wrote their own four questions, which they gave to the peer or teacher reader to respond to. At the beginning of the academic year, all learners were given a list of example questions. The list included roughly equal numbers of questions relating to content, organization and lexicogrammar, because these three aspects were also equally weighted in the course grading. An example of a question relating to content is “Does the introduction arouse your interest?” An example of a question relating to organization is “Is the order of ideas logical?” An example of a question relating to lexicogrammar is “Are verb tenses used correctly?” All students were also encouraged to use their own questions rather than just selecting from the list. In fact, at the beginning of the second semester the instructor did not distribute the example questions again, rather encouraging students to think of their own questions, since they already had one semester of experience with the feedback system. The peer or teacher reader was required to answer each of the writer’s questions by checking a box (either ‘yes’ or ‘no’). In addition to this, if the answer was ‘no’ the reader was required to mark the draft where they thought the problems lay. For example; if the question was ‘Is my vocabulary varied?’ and the answer was ‘No’, the

reader should then mark words which had been used repeatedly in the draft. The teacher used a colour coding system, with four different colours, one for each question asked by the writer. Peer readers were also instructed to use this colour system; however, they were inconsistent in doing so.

In addition to the four questions written by the writer, every feedback form included one additional question ‘Give one piece of constructive feedback’. The purpose of this additional question was twofold. Firstly, if the writer had chosen their questions poorly and no constructive feedback resulted from the first four questions, this gave the writer the opportunity to receive some constructive feedback for every draft. Therefore, they would have something to revise between every draft. Secondly, it was an opportunity for the reader to point out any problems that they felt were particularly apparent in the writing and thus it contributed to the writer's understanding of their weaknesses in writing and gave them ideas of possible questions to ask in subsequent feedback cycles. ‘Systematic feedback’ in this article, refers to feedback given using this system and does not include one-off questions asked by learners to individual peers or the teacher.

One 90 minute training session was conducted with both groups at the beginning of the academic year. The peer feedback groups received training on how to provide effective peer feedback, while the teacher feedback groups received training on how to understand and use teacher feedback when revising. At the beginning of the second semester, a forty five minute refresher training session was conducted, in which learners in the peer feedback group discussed what they remembered about how to provide effective feedback and those in the teacher feedback group discussed what they remembered about how to understand and use teacher feedback.

It seems that the grades a student receives on their assignments are likely to have a strong influence on their self-efficacy. In order to ensure that the grades students received for their writing assignments over the academic year did not influence changes in self-efficacy, first each assignment was graded by the instructor and the average grade for all students in the four classes was calculated. Subsequently, the grades for each class were scaled so that the average for each class was the same as the average for all classes combined. These scaled grades were the only ones reported to the students. This scaling meant that the average grade for each class was the same. Therefore, differences in students' self-efficacy deriving purely from the grades they received for their essays would not result in any difference between classes or treatment groups.

4.3.3 Participants

There were 67 participants in this study. Of those, 30 were in the peer feedback group and 37 were in the teacher feedback group. The university where the study took place has several writing support services available to students to receive extra writing support outside of class time; a face-to-face writing centre, an online writing centre and a peer writing forum. Both the face-to-face and the online writing centres offer 15 minute sessions for which learners choose one specific area of focus. Any student in one of the peer feedback classes who had visited the face-to face or online writing centre eight times or more over the one year period was excluded from the study as it was considered that they had received too much teacher feedback to be considered a member of the peer feedback group. Visiting the face-to-face or using the online writing centre eight times meant that the learner could ask eight questions, which was the equivalent of receiving systematic feedback (as practiced in this study) twice. The learners in the study received feedback 16 times over the one year period, meaning that they could ask 64 questions over the one year period. Therefore, even if a

student had used one of the writing centres seven times, asking one question each time, they would have only received roughly 10% as much teacher feedback as peer feedback. Two students who had used the writing centres eight times or more were found by the researcher and excluded from the study. Likewise, it was decided that any student in a teacher feedback group who had used the peer writing forum eight times or more over the academic year would be excluded, although in this study none of the participants fell into this category.

Students at the university are streamed into three ability tiers based on their overall English proficiency level. English proficiency is measured by students reading and listening scores on the TOEFL as well as in house speaking and writing tests. The speaking test consists of a 10 minute group discussion, while the writing test consists of a 30 minute timed essay. All four of these classes were in the low ability stream and students were randomly allocated to classes within the stream. The participants in this study represented four out of five intact classes of students in the low proficiency stream. Their language proficiency was judged to range from pre-intermediate to upper-intermediate level and their TOEFL scores ranged from 370 to 500. An independent samples *t*-test showed that there was no significant difference between the in-house writing test scores of the two teacher feedback classes and the two peer feedback classes at the beginning of the academic year; $t(60) = -0.929, p = 0.356$.

4.3.4 Analysis of Self-efficacy

According to Bandura (2006), questions designed to target the concept of self-efficacy must ask about factors that actually determine quality performance in the domain in focus. In addition to this, people should be asked to evaluate their capabilities as of now; questions should never be phrased in the past or future tenses. Pajares (1996) stated that the ideal level of specificity for questions designed to target the concept of self-efficacy depends on the

relative complexity of the performance criteria with which it will be compared. Pajares (2003) suggests that there should be as much correspondence as possible between the self-efficacy assessments and the performance assessments. In this study, four questions were used to target the construct of writing self-efficacy: 1) Now, how good do you think you are at organizing the ideas in your essays? 2) Now, how good do you think you are at giving support for each main idea in your essays? 3) Now, how good do you think your grammar skills are when writing essays? 4) Overall, how good do you think you are at writing essays in English now? These four questions are congruent with the performance criteria with which students were assessed. The writing section of the English proficiency test taken by all students at the end of the academic year is rated using four analytic rating scales: organisation, grammar, vocabulary and content. The content scale encompasses logical connection between ideas, support and development. In addition to this, in this class student writing was graded for lexicogrammar, organisation and content. Overall, if the student was capable of writing essays which were grammatically correct, well supported and well organized they would have a good chance of success both in terms of classroom assessment and institution-wide assessment of writing. Students were already aware of the rating scales used for the English proficiency test and were introduced to the grading criteria used for class writing assignments during the process of writing their first essay for the year.

Two questionnaires were used as the instruments in this study. A pre-treatment questionnaire was administered at the beginning of the academic year, before the students had started writing. The pre-treatment questionnaire consisted of 12 questions. The first six questions asked about students' past experiences in terms of writing instruction and feedback. Questions seven and eight asked about students' preferences relating to feedback, while the final four questions asked about students' confidence in their writing ability. The pre-treatment

questionnaire was based on the one used by Ruegg and Koyama (2010) and can be found in Appendix A.

A post-treatment questionnaire was administered at the end of the academic year, after all writing had been completed and all graded essays had been returned to the students. The post-treatment questionnaire consisted of 11 questions. The first four questions were the same as the final four questions in the pre-treatment questionnaire. Questions five and six asked students for their perceptions of the amount of peer and teacher feedback they had received. The following four questions asked for their perceptions of the usefulness of the feedback received. The final question asked students how much of the feedback they had understood. The post-treatment questionnaire was based on the one used by Ruegg and Koyama (2010). For this study, two different versions of this questionnaire were necessary because of the different feedback treatments received by the students. The post-treatment questionnaire completed by the peer feedback group can be found in Appendix B. The post-treatment questionnaire completed by the teacher feedback group can be found in Appendix C.

All questionnaires had been translated into Japanese and previously trialled on Japanese university students. Students were told that they could consult the instructor if they had any questions; however, all items seemed to be clear and easily comprehensible.

All questionnaire items used Likert scales with six choices from which to select. The choices were labelled alphabetically. Prior to statistical analysis, all alphabetical labels were converted into numeric responses. Principal Components Analysis rests on the assumptions that the sampling size is adequate, that there is a linear relationship between all variables, that correlations exist between all variables, and that there are no outliers. In order to test these assumptions, the Kaiser-Meyer-Olkin measure of sampling adequacy was employed,

correlations between variables were investigated by examining the correlation matrix and by running Bartlett's Test of sphericity. Descriptive statistics were calculated and used to determine whether there were any outliers. Cronbach's alpha reliability estimates were also calculated. Following this, principal components analysis with Varimax rotation was performed on both the pre-treatment and post-treatment questionnaires. Changes in writing self-efficacy were calculated by deducting each student's self-efficacy level at the beginning of the year from their self-efficacy level at the end of the year for the same skill. Thus, if they became more self-efficacious the value would be expressed as a positive; however, if they became less self-efficacious over the one year period the change would be expressed as a negative value. Finally, one-way analysis of variance (ANOVA) (2-tailed) was employed to find statistically significant differences in changes in writing self-efficacy over the one year period between the teacher feedback group and the peer feedback group. ANOVA rests on the assumptions of equal variance and normal distribution, so these assumptions were tested before the parametric test was employed.

4.3.5 Analysis of feedback

In order to determine whether there were any differences between the amount of constructive feedback or the amount of praise offered by the teacher and by peers, all preliminary drafts written by students over the one year period were collected. Because some students had failed to write or receive feedback on some of the drafts some data had to be excluded. It was decided to include only assignments for which both preliminary drafts and corresponding feedback checklists were present. After organizing the data in this way, there were an uneven number of assignments in the data set from the two groups. In order to include an even number of assignment samples from each group it was decided to randomly select ten assignment samples (each consisting of two preliminary drafts and their corresponding

feedback checklists) from the teacher feedback group and ten from the peer feedback group for each assignment. As the students completed eight assignments over the academic year, this resulted in a total of 80 assignment samples from the teacher feedback group (or 160 preliminary drafts) and 80 from the peer feedback group.

Constructive feedback was determined as any mark or comment that was intended to give advice on where or how to improve the writing. An example of constructive feedback is: “Please write your opinion”. On the other hand, praise was determined to be any comment that pointed out a strength of the writing. An example of praise is: “It is really easy to read and understand.” Some other comments given by peers and the teacher were neither constructive feedback nor praise, but rather comments of encouragement and these were excluded from the analysis.

The researcher (who was also the instructor) coded the feedback in terms of the number of different pieces of constructive feedback given and the number of instances of praise given on each preliminary draft. Silva and Kei Matsuda (2008) have noted that feedback is often coded according to its intention, but coded by someone other than the instructor who gave the feedback. This can cause problems because the coder may misinterpret the intention of the instructor. They conclude that it is preferable for instructors to be the ones who code their own feedback. After a period of one month, the researcher randomly selected and recoded 20% of the data (62 drafts) in order to determine intra-rater reliability. Intra-rater reliability of 0.996 was found using Pearson correlation. The feedback data was analysed in terms of equality of variance and normality of distribution, in order to ascertain that the assumptions for parametric testing were met. The data did not meet the assumptions of equality of variance and normality of distribution and therefore a nonparametric Mann-Whitney U test was employed in order to ascertain significant difference

between the amounts of constructive feedback and praise received by the teacher feedback group and the peer feedback group.

4.4 Results

4.4.1 Analysis of Self-efficacy

The Kaiser-Meyer-Olkin test of sampling adequacy for the pre-treatment questionnaire, resulted in a score of 0.560. Although the sample size is not good, Field (2013) suggests that values above 0.5 are minimally acceptable. In addition to this, the diagonal anti-image correlation values ranged from 0.510 to 0.684 and were therefore above the minimally acceptable level of 0.50 (Field, 2013). Furthermore, it was found that all four of the questions discussed in this chapter correlated at least 0.30 with at least one of the other questions, suggesting that it is reasonable to employ Principal Component Analysis. Bartlett's Test of sphericity had a highly significant result of <0.001 . Moreover, there were no significant outliers in the data. It was thus determined that the data was appropriate for use with Principal Components Analysis.

Principal components analysis of the pre-treatment questionnaire was conducted and extracted two factors. Factor one was identified as a background component. It comprised the first eight questions in the questionnaire. Factor two constituted writing self-efficacy and was made up of questions nine to twelve. Factor one accounted for 0.24, which indicates that 24% of the variance is accounted for by this factor. Factor two accounted for 0.17, indicating that 17% of variance is accounted for by this factor. The Cronbach's alpha reliability estimate for all questions in the pre-treatment questionnaire was 0.653.

The Kaiser-Meyer-Olkin test of sampling adequacy for the post-treatment questionnaire, resulted in a score of 0.778. This indicates that the sample size is good (Field,

2013). In addition to this, the diagonal anti-image correlation values ranged from 0.599 to 0.738 and were therefore above the minimally acceptable level of 0.50 (Field, 2013).

Furthermore, it was found that all four of the questions discussed in this chapter correlated at least 0.30 with at least one of the other questions, suggesting that it is reasonable to employ Principal Component Analysis. Bartlett's Test of sphericity had a highly significant result of <0.001 . Moreover, there were no significant outliers in the data. It was thus determined that the data was appropriate for use with Principal Components Analysis.

Principal components analysis of the post-treatment questionnaire also extracted two factors. Factor one was identified as feedback and factor two was identified as writing self-efficacy. Factor one consisted of questions five to eleven and accounted for 32% of variance, while factor two included questions one to four and accounted for 22% of total variance. The Cronbach's alpha reliability estimate for all questions in the post-treatment questionnaire was 0.696.

Descriptive statistics for the changes in self-efficacy experienced by students in the two groups can be seen in Table 1. In relation to changes in confidence in the ability to organise writing, the scores for the peer feedback group ranged from -2 to 2, whereas the teacher feedback group's scores ranged from -1 to 4. In relation to changes in confidence in the ability to support ideas in writing, the scores for the peer feedback group ranged from -2 to 2, whereas the teacher feedback group's scores ranged from -1 to 3. Concerning changes in confidence in the ability to use correct grammar when writing; the scores of the peer feedback group ranged from -2 to 1, whereas those of the teacher feedback group ranged from -2 to 3. Concerning changes in overall writing self-efficacy, the scores of the peer feedback group ranged from -2 to 2, whereas those of the teacher feedback group ranged from -1 to 4. The data were found to meet the assumption of equality of variance, skewness values ranged from

-0.107 to 0.572 and kurtosis measures ranged from 0.471 to 1.931. These values indicate that the data meets the assumptions of equality of variance and normality of distribution and it is therefore appropriate to employ parametric statistics.

Table 1. *Descriptive Statistics for changes in self-efficacy*

Confidence	Peer feedback group ($n=30$)		Teacher feedback group ($n=37$)	
	Mean	SD	Mean	SD
Organisation	0.167	0.834	0.797	1.115
Content	0.267	1.112	0.811	1.102
Grammar	-0.133	0.681	0.486	1.121
Overall	0.333	0.994	1.108	1.048

One-way ANOVA found that there was no significant difference (at the 0.05 level) between the confidence levels of the two groups at the beginning of the academic year. The differences in confidence in organization skills ($p = 0.876$), content ($p = 0.499$), grammar ($p = 0.078$) and total confidence ($p = 0.701$) were all insignificant.

One-way ANOVA found that the difference between the two groups' changes in confidence in organisation skills; $F(1) = 6.599, p = 0.013$, confidence in their ability to support their ideas in writing; $F(1) = 4.009, p = 0.049$, and confidence in their grammatical ability in writing; $F(1) = 7.047, p = 0.010$ were all significant at the 0.05 level. Finally, the difference between the two groups in terms of changes in overall writing self-efficacy was highly significant; $F(1) = 9.473, p = 0.003$. This shows that the teacher feedback group increased in confidence in their writing skills significantly more than the peer feedback group. In fact, the mean change in confidence in grammatical ability in writing for the peer feedback group was negative, indicating that, on average, they felt less confident at the end of the academic year than they had at the beginning of the academic year.

4.4.2 Analysis of Feedback

Descriptive statistics for the number of constructive feedback points and the number of instances of praise given by the teacher and by peers to each draft can be seen in Table 2.

Every draft in the teacher feedback group received at least one constructive feedback point, while the minimum value for constructive feedback in the peer feedback group was 0. Also, while the maximum amount of constructive feedback in the teacher feedback group was 59 points, the maximum in the peer feedback group was only 24 points. The number of instances of praise given by the teacher to each draft ranged from 0 to 1, while in the peer feedback group the range was wider; 0 to 4.

Table 2. *Descriptive Statistics for feedback*

Type	<u>Peer feedback group</u>		<u>Teacher feedback group</u>	
	Mean	SD	Mean	SD
Constructive	4.08	3.818	10.66	8.299
Praise	1.10	0.966	0.04	0.205

$n = 160$

The results of the Mann-Whitney U test show that in terms of the amount of constructive feedback given by the teacher and the peers, there was a significant difference ($p = <0.001$). On average, the teacher gave 2.5 times as much constructive feedback as peers did. There was also a statistically highly significant difference between the number of instances of praise given to each draft in the peer feedback group and in the teacher feedback group ($p = <0.001$). Indeed, peers gave, on average, more than 25 times as much praise as the teacher did.

4.5 Discussion

Each draft in the teacher feedback group received anywhere from one to 59 constructive feedback points and the average was almost 11 per draft. Students also received

feedback on 16 individual drafts over the academic year, so they would have received 171 constructive feedback points, on average, over a period of 10 months. In contrast to this, each draft in the peer feedback group received anywhere from 0 to 24 constructive feedback points, with an average of four per draft. Over the entire academic year, the average learner would have received 65 constructive feedback points. Looking at these figures, it is difficult to get a feel for whether the amounts would have seemed like too much to students. This raises the question: How much is too much?

The amount of constructive feedback given by the teacher was significantly more than that given by peers; nevertheless, the teacher feedback group still exhibited an increase in self-efficacy when compared to the peer feedback group. The finding that the students in the teacher feedback group increased in self-efficacy significantly more than the peer feedback group despite receiving considerably more constructive feedback is in contrast to the assertions of many experts on English language writing feedback who have maintained that excessive constructive advice may decrease learner confidence (e.g. Andrade & Evans, 2013; Bitchener & Ferris, 2012; Hyland & Hyland, 2001). However, this finding corroborates the idea from social cognitive theory that students' self-efficacy increases through increased exposure to feedback (e.g. Graham, MacArthur, Schwartz & Page-Voth, 1992; Pajares, 2003; Schunk & Swartz, 1993).

Only just over 4% of drafts in the teacher feedback group received praise. This is contrary to the advice of many second language writing specialists who have claimed that it is important to incorporate praise and constructive feedback side by side (e.g. Ferris & Hedgcock, 2005; Grabe & Kaplan, 1996; Hyland, K. 1996). However, despite this marked lack of praise received by students in the teacher feedback group, learner self-efficacy increased over the one year period.

4.6 Conclusion

It is clear that students who received systematic teacher feedback consistently throughout the year increased in self-efficacy over the academic year more than those who gave and received systematic peer feedback consistently throughout the year. Moreover, their confidence in all four aspects of writing investigated (organisation, support, grammar and overall writing ability) increased significantly. This is in contrast to the findings of Ruegg (2010), who found that peer feedback was the only variable that predicted increased confidence. However, in Ruegg's (2010) study all the participants received both peer and teacher feedback throughout the course. It seems that peer feedback does increase writing self-efficacy. Nevertheless, if students receive peer feedback alone it may be less beneficial, and may even be detrimental, to their self-efficacy when compared to those who receive teacher feedback alone. Like Ruegg (2010), Fox (1980) also found that the peer feedback group increased in confidence more than the teacher feedback group, although it is unclear whether this change in confidence can be attributed to the different assessment practices between the two groups in his study.

The findings of this study show that teacher feedback alone has the potential to increase learner confidence more than peer feedback alone. However, it does not necessarily follow that increased teacher feedback will always increase learner confidence as many other variables affect confidence. Examples of some such variables may be: the frequency and method of giving feedback, the instructor's attitude towards feedback and teaching style, the way in which the feedback is presented to the learners, the relationships between teacher and students and even the classroom atmosphere. It is recommended that future research of this nature examine different feedback variables in order to find out more about the extent to which these variables affect confidence.

While it is unclear how much is too much with regards to written teacher feedback, the idea from second language writing literature that too much teacher feedback leads to decreased self-efficacy (e.g. Andrade & Evans, 2013; Bitchener & Ferris, 2012; Hyland & Hyland, 2001) has so far not been corroborated by research and is perhaps not always true. Furthermore, the suggestion that praise should be given on every draft (Ferris, 1995; Ferris & Hedgcock, 2005) merits more empirical investigation. It may be more important, as implied by Hyland and Hyland (2006: 223), that our comments are “consistent, clear, helpful and constructive”.

Although many in the field of L2 writing have stated that increased teacher feedback results in decreased confidence (e.g. Cleary, 1990; Ferris, 2002; Truscott, 1996), this position has not been corroborated through research. The current study appears to refute this assertion. Indeed, however discouraged students may feel upon receiving constructive feedback from a teacher on their drafts, receiving increased teacher feedback appears to actually increase students' confidence in their L2 writing ability over the long term.

References

- Aida, Y. (1994). Examination of Horowitz, Horowitz and Cope's construct of foreign language anxiety: The case of students of Japanese. *The modern language journal*, 78(2), 155-168.
- Andrade, M. & Evans, N. (2013). *Principles and practices for response in Second Language Writing: Developing self-regulated learners*. New York: Routledge.
- Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Englewood Cliffs: Prentice Hall.
- Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In L. Pervin (Ed.). *Goal concepts in personality and social psychology* (pp. 19-38). Hillsdale: Erlbaum.

- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. Retrieved from:
<http://www.emory.edu/EDUCATION/mfp/effguide.html>
- Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the teaching of English*, 27(4), 395-422.
- Bitchener, J. & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Bong, M. (2001). Role of self-efficacy and task value in predicting college students' course performance and future enrolment intentions. *Contemporary Educational Psychology*, 26, 553-570.
- Bown, J. & White, C. (2010). A social and cognitive approach to SLA. *International Review of Applied Linguistics*, 48, 331-353.
- Cleary, L. (1990). The fragile inclination to write: Praise and criticism in the classroom. *English journal*, 79(2), 22-28.
- Chaudron, C. (1984). Evaluating Writing: Effects of feedback on revision. *RELC Journal*, 15(2), 1-14.
- Cohen, A. (1987). Student processing of feedback on their compositions. In A. Wenden & J. Rubin (Eds.) *Learner strategies in language learning*. Englewood Cliffs, N.J.: Prentice-Hall.
- Daly, J. (1978). Writing apprehension and writing competency. *Journal of educational research*, 72, 10-14.
- Daly, J. & Shamo, W. (1978). Academic decisions as a function of writing apprehension. *Research in the teaching of English*, 12, 119-126.

- Eysenck, M. (1979). Anxiety, learning and memory: A reconceptualisation. *Journal of research in personality*, 13, 363-385.
- Ferris, D. (1995). Student reactions to teacher response in Multiple-draft composition classrooms. *TESOL Quarterly*, 29(1), 33-53.
- Ferris, D. (2002). *Treatment of Error in Second Language Student writing*. Ann Arbor: University of Michigan Press.
- Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah: Lawrence Erlbaum Associates.
- Ferris, D. (2006). Does error feedback help student writers? In K. Hyland & F. Hyland (Eds.). *Feedback in Second Language Writing: Contexts and Issues* (pp. 81-104). Cambridge: Cambridge University Press.
- Ferris, D. & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process and practice*. Mahwah: Lawrence Erlbaum Associates.
- Ferris, D., Liu, H., Sinha, A. & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22, 307-329.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.
- Finn, K. & Frone, M. (2004). Academic performance and cheating: Moderating role of school identification and self-efficacy. *Journal of Educational Research*, 97, 115-122.
- Fox, R. (1980). Treatment of writing apprehension and its effects on composition. *Research in the teaching of English*, 14(1), 39-49.
- Gardner, R. (1997). Individual differences and second language learning. In R. Tucker & D. Corson (Eds.). *Encyclopaedia of Language and education: Second Language Education*, (pp. 33-42). Dordrecht: Kluwer Academic Publishers.
- Gardner, R., Tremblay, P. & Masgoret, A. (1997). Towards a full model of Second Language

- Learning: An empirical investigation. *Modern Language Journal*, 81(3), 344-362.
- Goldstein, L. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, 13(1), 63-80.
- Grabe, W. & Kaplan, R. (1996). *Theory and practice of writing*. Harlow: Pearson Education.
- Graham, S. (2007). Learner strategies and self-efficacy: making the connection. *Language Learning Journal*, 35(1), 81-93.
- Graham, S., MacArthur, C., Schwartz, S., & Page-Voth, V. (1992). Improving the compositions of students with learning disabilities using a strategy involving product and process goal setting. *Exceptional Children*, 58(4), 322-335.
- Gungle, B. & Taylor, V. (1989). Writing apprehension and second language writers. In D. Johnson & D. Roen (Eds.). *Richness in writing: Empowering ESL students* (pp. 235-148). New York: Longman.
- Hyland, F. (1998). The impact of Teacher Written Feedback on Individual Learners. *Journal of Second Language Writing*, 7, 255-286.
- Hyland, F. & Hyland, K. (2001) Sugaring the Pill: Praise and criticism in written feedback. *Journal of Second Language Writing* 10, 185-212
- Hyland, K. (1996). *Second language writing*. New York: Cambridge.
- Hyland, K. & Hyland, F. (2006). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In Hyland, K. & Hyland, F. (Eds.). *Feedback in Second Language Writing: Contexts and issues* (pp. 206-224). New York: Cambridge.
- Krashen, S. (1982). *Principles and practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL*

Journal, 3, 5-19.

- Martinez, C., Kock, N. & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students' writing anxiety and writing self-efficacy. *Journal of Adolescent and Adult Literacy*, 54(5), 351-360.
- McCarthy, P., Meier, S. & Rinderer, R. (1985). Self-efficacy and writing: A different view of self-evaluation. *College Composition and Communication*, 36(4), 465-471.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In Johnson, D. & Roen, D. (Eds.). *Richness in writing: Empowering ESL students* (pp 207-219). New York: Longman.
- Pajares, F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*. Paper presented at the Annual meeting of the American Educational Research Association, New York.
- Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. Pintrich (Eds.). *Advances in motivation and achievement* (pp 1-49). Greenwich: JAI Press.
- Pajares, F. (2003). Self-efficacy beliefs, motivation and achievement in writing: A review of the literature. *Reading and Writing Quarterly*, 19, 139-158.
- Pajares, F. & Johnson, M. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy and apprehension. *Research in the teaching of English*, 28(3), 313-331.
- Prat-Sala, M. & Redford, P. (2010). The interplay between motivation, self-efficacy, and approaches to study. *British Journal of Educational Psychology*, 80, 283-305.
- Robbins, S., Lauver, K., Le, H., Davis, D. & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261-288.

- Ruegg, R. (2010). Aspects affecting confidence in writing. *Studies in Linguistics and Language Teaching*, 21, 123-135.
- Ruegg, R. & Koyama, D. (2010). Confidence in writing: The effect of feedback. *Studies in Linguistics and Language Teaching*, 21, 137-166.
- Saito, Y. & Samimy, K. (1996). Foreign language anxiety and language performance: A study of learner anxiety in beginning, intermediate and advanced-level college students of English. *Foreign Language Annals*, 29(2), 239-249.
- Scharzer, R. (1986). Self-related cognition in anxiety and motivation: An introduction. In R. Scharzer (Ed.), *Self-related cognition in anxiety and motivation* (pp. 1-17). Hillsdale: Lawrence Erlbaum.
- Schunk, D. (1983). Developing children's self-efficacy and skills: The roles of social comparative information and goal setting. *Contemporary Educational Psychology*, 8, 76-86.
- Schunk, D. (1985). Self-efficacy and classroom learning. *Psychology in the schools*, 22, 208-223.
- Schunk, D. (1991). Self-efficacy and academic motivation. *Educational psychologist*, 26(3&4), 207-231.
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progressive feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18, 337-354.
- Silva, T. & Kei Matsuda, P. (Eds.). (2008). *On Second Language Writing*. Mahwah: Routledge.
- Truscott, J. (1996). The case against grammar correction. *Language Learning* 46(2), 327-369.
- Tsui, A. & Ng, M. (2000). Do Secondary L2 Writers Benefit from Peer Comments? *Journal of Second Language Writing* 9, 147-171.

- Zacharias, N. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal*, 38(1), 38-52.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209-222.
- Zimmerman, B., Bandura, A. & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663-676.

Appendix 4.A

Pre-treatment questionnaire

1. Did you ever learn to write essays in Japanese?

A: Yes

B: No (Go to question 4)

2. When you learnt to write essays in Japanese, how often did you receive peer feedback?

A: Never

B: Once or twice

C: Often

D: Every essay

3. When you learnt to write essays in Japanese, how often did you receive teacher feedback?

A: Never

B: Once or twice

C: Often

D: Every essay

4. Did you ever learn to write essays in English?

A: Yes

B: No (Go to question 7)

5. When you learnt to write essays in English how often did you receive peer feedback?

A: Never

B: Once or twice

C: Often

D: Every essay

6. When you learnt to write essays in English, how often did you receive teacher feedback?

A: Never

B: Once or twice

C: Often

D: Every essay

7. How often would you like to practice peer feedback during this Advanced Writing course?

A: Never

B: Once or twice

C: Often

D: Every essay

8. How often would you like to receive teacher feedback during this Advanced Writing course?

A: Never

B: Once or twice

C: Often

D: Every essay

9. How good do you think you are at organizing the ideas in your essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

10. How good do you think you are at giving support for each main idea in your essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

11. How good do you think your grammar skills are when writing essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

12. Overall, how good do you think you are at writing essays in English?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

Appendix 4.B

Post-treatment questionnaire (Peer feedback group)

1. Now, how good do you think you are at organizing the ideas in your essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

2. Now, how good do you think you are at giving support for each main idea in your essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

3. Now, how good do you think your grammar skills are when writing essays?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

4. Overall, how good do you think you are at writing essays in English now?

A: Terrible

B: Poor

C: Not so good

D: Good

E: Very good

F: Excellent

5. How was the amount of peer feedback you did?

A: Very insufficient

B: A little insufficient

C: Just right

D: A little too much

E: Far too much

6. How was the amount of teacher feedback you received?

- A: Very insufficient
- B: A little insufficient
- C: Just right
- D: A little too much
- E: Far too much

7. How helpful was peer feedback for improving your grammar?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

8. How helpful was peer feedback for improving your organization of ideas in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

9. How helpful was peer feedback for improving the support you gave for each main idea in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

10. Overall, how helpful was peer feedback?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

11. How much of the peer feedback did you understand?

- A: None
- B: A little
- C: Some
- D: Most
- E: All

Appendix 4.C

Post-treatment questionnaire (Teacher feedback group)

1. Now, how good do you think you are at organizing the ideas in your essays?
A: Terrible
B: Poor
C: Not so good
D: Good
E: Very good
F: Excellent

2. Now, how good do you think you are at giving support for each main idea in your essays?
A: Terrible
B: Poor
C: Not so good
D: Good
E: Very good
F: Excellent

3. Now, how good do you think your grammar skills are when writing essays?
A: Terrible
B: Poor
C: Not so good
D: Good
E: Very good
F: Excellent

4. Overall, how good do you think you are at writing essays in English now?
A: Terrible
B: Poor
C: Not so good
D: Good
E: Very good
F: Excellent

5. How was the amount of peer feedback you did?
A: Very insufficient
B: A little insufficient
C: Just right
D: A little too much
E: Far too much

6. How was the amount of teacher feedback you received?

A: Very insufficient

B: A little insufficient

C: Just right

D: A little too much

E: Far too much

7. How helpful was teacher feedback for improving your grammar?

A: Not helpful at all

B: A little helpful

C: Very helpful

8. How helpful was teacher feedback for improving your organization of ideas in your essays?

A: Not helpful at all

B: A little helpful

C: Very helpful

9. How helpful was teacher feedback for improving the support you gave for each main idea in your essays?

A: Not helpful at all

B: A little helpful

C: Very helpful

10. Overall, how helpful was teacher feedback?

A: Not helpful at all

B: A little helpful

C: Very helpful

11. How much of the teacher feedback did you understand?

A: None

B: A little

C: Some

D: Most

E: All

CHAPTER FIVE

Study Four

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

Abstract

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 in terms of both quantity and quality. 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 only 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.

5.1 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 cited in Hirose, 2008) looked at the perceptions of lower-level Japanese students to 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 see it as 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. However, F. 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, F. 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, K. 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, K. 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 just 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, K. 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?

5.2 Research methods

5.2.1 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 ITP 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 ITP 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 performance 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 (2 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 a little 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, were 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.

5.2.2 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 (only 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. 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. A student who offered no constructive feedback, but rather praise alone, would receive a score of 1 out of 5. A student who offered only one or a few minor instances of feedback, such as correcting surface-level grammatical

errors, would receive a score of 2 out of 5. A student who offered a reasonable amount of feedback, most of which was accurate, would receive a score of 3 out of 5. A student who gave quite a lot of feedback, most of which was accurate, would receive a score of 4 out of 5. Only 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. When discrepancies occurred, for example between the quantity and quality of the peer feedback, the instructor balanced the two aspects to arrive at a fair grade for the task. A separate score was given for each of the two peer review sessions and the scores were then averaged, to give the final peer review grade assigned to the student for the assignment. 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. 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. These three or four separate grades 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.

5.2.3 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. Because all of the feedback relates to articles, this would be counted as one constructive point (i.e. one point is covered by the feedback overall). On the other hand, another student might give less feedback, quantity wise, but address a larger number of different issues, showing more breadth. This feedback would be counted as more constructive points than the previous example (i.e. more than one point is covered by the feedback as a whole). The number of marks on each draft related specifically to the act of marking up the peer's 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 number of comments included multiple comments on the same issue, whereas the number of constructive points covered only the number of different points made by the peer reviewer, ignoring repetition of the same advice. 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 scale with three data points 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. There is no halfway point between being able to be linked to a revision or not, and therefore there was no score of 0.5 for this variable.

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. Silva and Kei Matsuda (2008) have noted that feedback is often coded according to its intention, but coded by someone other than the instructor who gave the feedback. This can

cause problems because the coder may misinterpret the intention of the instructor. They conclude that it is preferable for instructors to be the ones who code their own feedback. For this reason, the data was coded by the researcher, who was also the instructor. 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 (Appendix A and B respectively).

T-tests are a parametric test to determine whether the means of groups are equal (Huizingh, 2007). Two assumptions need to be met when using t-tests: 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 only 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.

5.3 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 only 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.

It can also be seen that in the only 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 only product-assessed group. The proportionally higher standard deviations show that the quality and quantity of feedback given in the only 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; ($p = 0.005$), the total number of words written in the two groups; ($p = 0.003$), and the average length of comments made; ($p = 0.006$), while the number of different points made by the readers in the two groups; ($p = 0.094$), the number of marks made on drafts; ($p = 0.812$), the accuracy level of the comments and marks; ($p = 0.720$) and the proportion of specific

Variable	Semester One				Semester Two				Whole year			
	Only product-		Feedback-		Only product-		Feedback-		Only product-		Feedback-	
	assessed		assessed		assessed		assessed		assessed		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

Table 1. Descriptive Statistics for all drafts
n = 80

comments made by the two groups; ($p = 0.173$) 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; ($p = < 0.001$), the number of marks made on drafts in the two groups; ($p = 0.001$), the number of comments written in the two groups; ($p = < 0.001$), the total number of words written in the two groups; ($p = 0.001$) and the average length of comments in the two groups; ($p = < 0.001$). However, no significant difference was found (at the 0.05 level) between the accuracy level of the comments and marks; ($p = 0.318$), nor the proportion of specific comments; ($p = 0.075$) 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; ($p = < 0.001$), the number of marks made on drafts in the two groups; ($p = 0.007$), the number of comments written in the two groups; ($p = < 0.001$), the total number of words written in the two groups; ($p = < 0.001$), the average length of comments; ($p = < 0.001$), and the proportion of specific comments made in the two groups; ($p = 0.026$), while the accuracy level of the comments and marks in the two groups was not significantly different at the 0.05 level; ($p = 0.319$). Table 2 gives an overview of all significant and non-significant results.

Looking at the table, it seems clear that the number of comments and length of comments were increased from the outset by the assessment of feedback. The breadth of feedback and number of marks on drafts increased more slowly when feedback was assessed. While the specificity of comments did increase in the feedback-assessed group, it increased at a slower

rate.

Table 2. *Overview of Mann-Whitney U Test results*

Measure	Time period	Differences between feedback-assessed and only product-assessed groups	Group favoured
No. of points	Semester One	Not significant	
	Semester Two	Significant ($p \leq 0.001$)	Feedback-assessed
	Whole year	Significant ($p \leq 0.001$)	Feedback-assessed
No. of marks	Semester One	Not significant	
	Semester Two	Significant ($p = 0.001$)	Feedback-assessed
	Whole year	Significant ($p \leq 0.01$)	Feedback-assessed
No. of comments	Semester One	Significant ($p = 0.005$)	Feedback-assessed
	Semester Two	Significant ($p \leq 0.001$)	Feedback-assessed
	Whole year	Significant ($p \leq 0.001$)	Feedback-assessed
Total words	Semester One	Significant ($p \leq 0.005$)	Feedback-assessed
	Semester Two	Significant ($p = 0.001$)	Feedback-assessed
	Whole year	Significant ($p \leq 0.001$)	Feedback-assessed
Length of comments	Semester One	Significant ($p \leq 0.01$)	Feedback-assessed
	Semester Two	Significant ($p \leq 0.001$)	Feedback-assessed
	Whole year	Significant ($p \leq 0.001$)	Feedback-assessed
Specificity of comments	Semester One	Not significant	
	Semester Two	Not significant	Feedback-assessed
	Whole year	Significant ($p \leq 0.03$)	
Accuracy of comments	Semester One	Not significant	
	Semester Two	Not significant	
	Whole year	Not significant	

5.4 Discussion and conclusion

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 only product-assessed group and they also wrote longer comments and a significantly higher proportion of specific comments than the only 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 only 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 only product-assessed group, wrote more words overall than the only product-assessed group and wrote on average longer comments than did the only 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 only 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 only 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 were used, this difference may have reached the level of significance. The precept that practice makes perfect may be at work here, with the accuracy increasing as a direct result of more practice at giving peer feedback. 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 reviewers 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 generalizable 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 with encouragement 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. Furthermore, a more prescriptive peer feedback format and more detailed peer feedback training may increase the quality of peer feedback. Therefore, it would be worthwhile to conduct a study similar to this one with a more controlled peer feedback format and after learners have been trained to provide peer feedback in a particular manner. 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 and M. Swanson (Eds.). *JALT2007 Conference Proceedings*. Tokyo: JALT.
- 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*. In D. Belcher & J. Liu (Eds.) Michigan Series on Teaching Multicultural Writers.
- 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 D. Johnson & D. Roen (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 6th June 2013 from: <http://www.tesl-ej.org/wordpress/issues/volume2/ej08/ej08a2/?wscr=>
- 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.
- Silva, T. & Kei Matsuda, P. (Eds.). (2008). *On Second Language Writing*. Mahwah:

Routledge.

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.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209-222.

Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication*, 12(4), 492-528.

Appendix 5.A

Unhelpful feedback sheet

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

Feedback Checklist

	<u>Yes</u>	<u>No</u>
1. <u>Do you understand everything clearly?</u>	■	<input type="checkbox"/>
2. <u>Is their enough information?</u>	■	<input type="checkbox"/>
3. <u>Are prepositions used correctly?</u>	■	<input type="checkbox"/>
4. <u>Does every sentence have a subject, verb and object?</u>	■	<input type="checkbox"/>
5. Please give one piece of constructive feedback.		
<u>I can understand everything clearly. It's easy to understand and has a lot of information. I wanna go to Bangkok!!</u>		

Helpful feedback sheet

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

Feedback Checklist

	<u>Yes</u>	<u>No</u>
1. <u>Do you understand everything clearly?</u>	■	□
2. <u>Is the organisation of ideas clear?</u>	■	□
3. <u>Are the verb tenses correct?</u>	□	■
4. <u>Are all words spelled correctly?</u>	■	□
5. Please give one piece of constructive feedback.		
<u>It's interesting result. Especially Q. 2. I like this topic!! But....you have some careless mistakes. Please check these!!</u>		

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

CHAPTER SIX

Study Five

The effect of assessment of process after receiving teacher feedback

Abstract

This quasi-experimental study investigated the effect of assessing both process and product compared to assessing written products alone. Two groups of students received teacher feedback over a one year period. One group was assessed on their revisions in addition to the quality of final drafts while a second group was assessed on the quality of final drafts alone. The uptake of feedback, in the form of revision attempts, and the success of those attempts were compared between the two groups using a Mann-Whitney U Test. In addition, improvement in writing performance was measured and compared between the two groups using one-way ANOVA. It was found that the group assessed on process had significantly higher uptake of feedback than the group assessed on product alone. However, the increased uptake consisted of unsuccessful revision attempts while the number of successful revisions was almost identical between the groups. It was also found that the group assessed on final product alone improved significantly more than the group assessed on process in terms of essay content. It is concluded that the assessment of the use of teacher feedback in revised versions of the same essay may not be necessary in some contexts and suggestions are given for increasing the quality of learning achieved by teacher feedback.

6.1 Introduction

Over several decades, a gradual change has been occurring in the teaching of writing in English. The introduction of the process approach in first language writing classrooms

(Flower & Hayes, 1981) was soon adopted by second/foreign language writing instructors (Keh, 1990; Raimes, 1993; Zhang, 1995). This was followed by a subsequent shift in assessment practices. Standardised assessment of writing moved from indirect assessments of writing, such as grammar tests and gap-fill exercises to the direct assessment of writing samples, usually a timed essay (Vaughan, 1991). Similarly, classroom assessment also changed from the assessment of essays produced by learners to alternative assessment forms such as portfolio assessment, which allows for the comparison of various different kinds of written products as well as the evaluation of the drafting process used by learners (Hamp-Lyons, 1991; Romova & Andrew, 2011). Although in some contexts a product-oriented approach still exists, in most first language and second/foreign language contexts the process approach to writing is now firmly entrenched and the direct assessment of writing is widely used around the world, the change in classroom assessment practices is still taking place, with many assessing only the final draft of student texts, even in process approach classrooms. Currently, many learners are assessed on only the final draft of the writing they produce, sometimes even in classrooms in which portfolio assessment is used (Gearhart & Wolf, 1997), while others are assessed on the process they use in the production of the writing in addition to the quality of the final draft.

It has been suggested that assigning a grade for the writing process as well as the quality of the final draft encourages learners to pay more attention to the feedback they receive and put more effort into revision (Ferris, 2002; Goldstein, 2006). However, it has not been ascertained through research whether the assessment of process changes student revision behaviours and outcomes or not.

6.2 Review of previous studies

6.2.1 Literature on classroom assessment

Literature from the field of educational measurement states clearly that classroom grades should represent student achievement alone. Measurement experts recommend that non-achievement factors such as effort and improvement should not be taken into consideration when assigning classroom grades (Brookhardt, 2004; Linn & Miller, 2005; Randall & Engelhard, 2010; Zhang & Burry-Stock, 2003). According to Zhang and Burry-Stock (2003), factors like effort, attitude and motivation should not be included in grades because they are subjective in nature and therefore difficult to measure. Tierney, Simon and Charland (2011) add that considering effort in grades could cause problems if it is not applied consistently, and that students may vary considerably in how they display effort.

Yet a number of researchers in the field have found that teachers invariably do take non-achievement factors such as effort, participation, attitude, improvement and motivation into consideration when grading (Green, Johnson, Kim & Pope, 2007; Griswold, 1993; Hills, 1991; Hunter, Mayenga & Gambell, 2006; Stiggins, Frisbie & Griswold, 1989; Tierney, Simon & Charland, 2011). Furthermore, this practice is not incidental; rather teachers seem to believe that effort should be considered during grading. Moreover, results of research conducted by Tierney, Simon and Charland (2011) suggest that effort is considered in grades more than other non-achievement factors, such as attitude, motivation or participation. Indeed, 85% of the teachers in Green, Johnson, Kim and Pope's (2007) study considered it was ethical to consider effort when grading, and half of the teachers in Resh's (2009) study considered it fair to employ different grading systems for weak and strong students, putting more stress on effort when grading weak students.

In practice, one third of the teachers in Tierney, Simon and Charland's (2011) study, and almost two thirds of the teachers in Hunter, Mayenga and Gambell's (2006) study, consider effort in their grading. In their study, which employed factor analysis, Hunter, Mayenga and Gambell (2006) found that the first factor incorporated effort, attendance, improvement and participation and this factor accounted for nearly 28% of total variance in grades. Teachers in Resh's (2009) study, on average, thought that 25.5% of grades should be allocated on the basis of effort, while students, on average, thought that 33.2% of grades should be awarded for effort. In practice, it seems that success that comes after a great deal of effort is rewarded more than that which results from less effort (Griswold, 1993; Hunter, Mayenga & Gambell, 2006; Randall & Engelhard, 2010).

Brookhart (1991) suggested that the reason teachers depend on non-achievement factors, such as effort, is to avoid the negative consequences which may occur if students' grades were based on achievement alone. That is to say, if effort were not taken into account, students would be more likely to receive lower, or even failing, grades. The vast majority of teachers in Griswold's (1993) study assessed both effort and achievement and their reason for doing so was to encourage future effort. Similarly, teachers in Tierney, Simon and Charland's (2011) study were concerned about possible long term consequences of ignoring non-achievement factors. Bonesronning (2004) claims that effort is the variable which students are most easily able to control; therefore, when a grading system rewards effort, students are likely to increase their effort expenditure and consequently achieve more.

Usually, assessment is made up of two parts, classroom grades and standardized assessments (Zhang & Burry-Stock, 2003). Although it is clear that standardized assessments are more reliable than classroom grades, Griswold (1993) also claims that teachers do not even try to apply the concept of validity to their grading. Although they may state that the grade

represents a certain construct, such as achievement, they are willing to systematically change grades using different criteria, such as attitude, effort or improvement.

6.2.2 The assessment of process in L2 writing

According to Gullickson (1985), grading practices should be designed to assess not only the quality of students' final products, but also their abilities and skills and as a way to communicate and apply the expectations of the teacher and the curriculum. Assessment tools used in the classroom should align with the objectives of the curriculum as different assessment practices will develop different types of skills or knowledge (Hunter, Mayenga & Gambell, 2006). While the process approach to writing is now firmly entrenched in second/foreign language writing courses, it is often used for instruction but not for assessment. That is, students are asked to focus on revision and editing, class time is devoted to training and practice in these skills and yet students' writing performance is assessed on the final draft alone. If the teacher and the curriculum place importance on revision skills, then this should be communicated to students through the assessment criteria.

Although the benefits of the process approach and peer and teacher feedback have been debated vigorously, there has been relatively little discussion of different options for assessing writing skills (Walker & Perez-Riu, 2008). In addition, few empirical research studies have been published investigating the assessment of process writing (Xiaoxiao & Yan, 2010). Even if instructors explain the importance of revision, if revision is done outside of the classroom it is not certain how much effort students will put into it. It was mentioned by Goldstein (2006) that although the instructor in her study believed revision to be important, some students did not put much effort into the revision process because they were busy and the revision was not assessed by the instructor. Goldstein implies that if the revision had been

graded, this would have emphasised its importance to students and encouraged them to put more effort in. Ferris (2002) also recommends assigning a grade for revision students carry out between drafts after receiving teacher feedback. It seems that students will increase the amount of effort they invest in an activity in proportion to the marginal returns on that effort (Linn & Miller, 2005).

6.2.3 Investigation of uptake of feedback

In the multitude of studies on both oral and written feedback in language learning, uptake has been defined in a number of different ways. Lyster and Ranta (1997) refer to two types of uptake in their study 'repair' and 'needs repair'. Repair constitutes successful revision, whereas 'needs repair' simply indicates that there was some kind of response as opposed to the learner ignoring the feedback altogether. These same categories are used by Loewen (2004) who calls them 'successful uptake' and 'unsuccessful uptake'. Sheen (2004) on the other hand, did not distinguish between the two types of uptake and simply investigated whether or not there was any uptake. These three studies on oral feedback highlight the idea that whether or not there is an attempt to repair/revise is the important point, for it is attempts which signal hypothesis testing and hypothesis testing demonstrates possible development of the learner's interlanguage.

There has been a great deal of debate about the value of studies investigating uptake in order to evaluate the effectiveness of various feedback practices. While many have been quick to point out that uptake has not been shown to indicate acquisition (Long, 2007; Ohta, 2000; Mackay & Philp, 1998; Truscott, 2007), others believe that uptake may be an indication that one step in the process of acquisition has taken place (Ellis, Basturkmen & Loewen, 2001; Ferris, 2004; 2006; Loewen, 2004; Lyster & Ranta, 1997; Sachs & Polio, 2007; Swain, 1995;

Williams, 2001). Meanwhile, most in the field of L2 writing see uptake as highly relevant in a context in which the development of revision skills is important (Chandler, 2003; Chaudron, 1984; Ferris, 2010).

Although assessment of improvement is not recommended by measurement specialists (Brookhardt, 2004; Linn & Miller, 2005; Randall & Engelhard, 2010; Zhang & Burry-Stock, 2003), anecdotal evidence suggests that many L2 writing teachers incorporate improvement between drafts as a grading criterion. Although a number of studies in both Second Language Acquisition (SLA) (e.g., Bitchener, 2008; Bitchener & Knoch, 2009; 2010; Sheen, Wright & Moldova, 2009) and Second Language (L2) writing (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001; Ferris, 2006) have examined differences in uptake in order to determine which types of feedback are more effective, to my knowledge no-one has evaluated the uptake of feedback by students who are assessed for improvement compared to those who are not in order to determine the effect of assessment of improvement between drafts on student revisions. Furthermore, as far as I am aware there has been no attempt to ascertain any long-term pedagogical affects the assessment of improvement between drafts might have.

The purpose of the present study is to determine whether the assessment of process in addition to final product leads to any significant differences in the number of (successful) revision attempts made by students between drafts or long-term improvement in writing performance. The research questions for the present study are:

1. Does the assessment of process in addition to product lead to more revision attempts than the assessment of product alone?
2. Does the assessment of process in addition to product lead to more successful revisions than

the assessment of product alone?

3. Does the assessment of process in addition to product affect the amount of improvement in writing performance achieved by learners over a one year period compared to the assessment of product alone?

6.3 Method

6.3.1 Methodology

This research involved a quasi-experimental design with a grading intervention, to explore the effects of grading improvement between drafts in addition to the quality of final written product. In order to determine whether the grading of improvement between drafts has any effect on the amount of attention learners pay to teacher feedback, an independent group design was used, with the two groups being graded in different ways for a period of one year. Although the data was collected longitudinally, the total number of participants was small. The learners in this study all had the same cultural and educational background and were studying at the same institution. Therefore, it is unclear how generalizable these results may be to different institutions or different cultural and educational backgrounds. It is suggested that further research be conducted with a larger number of participants. Similarly, conducting a similar study in a different context would certainly be a worthwhile pursuit as it may offer additional dialogue to this discussion.

In order to ensure that the study did no harm to the students, the grades for each assignment were averaged for each of the two classes. The difference between the grades in the two classes was divided by two, and then this figure was added to the grades of every student in the class with lower average grades and subtracted from the grade of every student in the class with higher average grades. This ensured that the average grade reported to

students in each of the two classes was the same.

6.3.2 Participants and data

The participants were all consenting members of two entire second-year classes in the English department at a private university of foreign languages in Eastern Japan. Both of the classes were in the lowest of three proficiency streams. Students had been placed in the classes randomly, based on their TOEFL ITP scores as well as their scores on the writing and speaking sections of the in-house English proficiency test. Their TOEFL ITP scores ranged from 370 to 500. Class 1 had 24 students, of whom 15 gave their consent to participate in the study. Class 2 had 30 students, of whom 24 gave their consent to participate in the study. Each student was required to write and submit two preliminary drafts and one final draft of each assignment over a period of one academic year. In total eight assignments were completed during the one year period. However, because some students failed to write or submit all three drafts for every assignment, only those assignments for which complete information was available were included in the data pool. From this data pool, five essays were randomly selected from each class for each assignment. This resulted in two preliminary drafts each of ten examples of each assignment being included in this study, a total of 80 essays (160 drafts). This random selection process resulted in all 15 of the 15 consenting students from class 1 and 21 of the 24 consenting students from class 2 being included in the study. The participants from class 1 included 12 females (80%) and 3 males (20%), while the participants from class 2 included 14 females (67%) and 7 males (33%). All of the participants were Japanese nationals and they ranged in age from 19 to 21 years of age at the time of the study.

6.3.3 Pedagogical treatment

This research involved a quasi-experimental design with independent groups being graded in

different ways for a period of one year. The two classes involved in this study had the same writing instructor who was also the researcher. The assignments written by the students over the academic year were a critical essay, a compare and contrast essay, a cause and effect essay and a long research paper which was written, submitted and graded in sections before being compiled and resubmitted as a single long paper. The classes were taught in the same way throughout the year with the exception of the grading criteria used to assess each assignment. Both classes received systematic teacher feedback exclusively on every preliminary draft of every essay over the one year period, meaning that if a student attended class and completed their homework without fail they would have received teacher feedback 16 times over the academic year.

In the interest of encouraging learner autonomy, all writers wrote their own four questions, which they gave to the teacher to respond to. At the beginning of the academic year, all learners were given a list of example questions. The list included roughly equal numbers of questions relating to content, organization and lexicogrammar, because these three aspects were also equally weighted in the course grading. An example of a question relating to content is “Does the introduction arouse your interest?” An example of a question relating to organization is “Is the order of ideas logical?” An example of a question relating to lexicogrammar is “Are verb tenses used correctly?” All students were also encouraged to use their own questions rather than just selecting from the list. In fact, at the beginning of the second semester the instructor did not distribute the example questions again, rather encouraging students to think of their own questions, since they already had one semester of experience with the feedback system. The teacher answered each of the writer’s questions by checking a box (either ‘yes’ or ‘no’). In addition to this, if the answer was ‘no’ the teacher marked the draft where they thought the problems lay. For example; if the question was ‘Is

my vocabulary varied?’ and the answer was ‘No’, the teacher then marked words which had been used repeatedly in the draft. The teacher used a colour coding system, with four different colours, one for each question asked by the writer. In addition to the four questions written by the writer, every feedback form included one additional question: ‘Give one piece of constructive feedback’. This was a chance for the teacher to comment on one salient problem encountered in each draft. ‘Systematic feedback’ in this article, refers to feedback given using this system and does not include one-off questions asked by learners to individual peers or the teacher.

One 90 minute training session was conducted at the beginning of the academic year. Students received training on how to understand and use teacher feedback when revising. At the beginning of the second semester, a forty five minute refresher training session was conducted, in which learners discussed what they remembered about how to understand and use teacher feedback.

Class 1 was assessed for their revision (defined as the overall quality and quantity of revisions made between the first and the final draft) as well as the organisation, grammar and vocabulary, and content of their final draft. All four scores were equally weighted so that the revision score constituted 25% of the total grade for each essay. Class 2 was assessed on the final three criteria alone, with the three scores being equally weighted. The grading criteria were explained in detail to students in both classes at the beginning of the academic year and the scores for each essay were received within 2 weeks of having submitted it. This also means that in every case students received their grades for one essay before the subsequent one was due to be submitted. Although the feedback on the final drafts consisted of the scores for each criterion and the total grade alone, students had a chance to ask questions about their scores and grades and often did. If the students asked why they were given a certain score the

instructor explained in detail the reasons for the score, using examples from the essay in question. Overall, it seems that the students were clearly aware of how their essays would be assessed.

6.3.4 Uptake analysis

Each feedback point from each sample preliminary draft was numbered and then categorised in terms of the resulting revision in the subsequent draft. Firstly, the feedback point was categorized as to whether or not it resulted in uptake: ‘attempt’ or ‘no attempt’. Following this, all revision attempts were further categorized into the following three categories: ‘misunderstood’, ‘incorrect revision’ or ‘corrected’. The first category refers to situations in which the learner made altogether the wrong kind of change in response to feedback. An example of a feedback point that was misunderstood by a student follows.

Original sentence (in an article summary paragraph): *“Few days, a gay couple was arrested with discriminating law in Malawi.”*

Feedback: *“You should not write your own personal opinion in this paragraph.”*

Revised sentence: *“Few days, a gay couple was arrested with one-sided law in Malawi.”*

Understanding that there was some problem with the word ‘discriminating’, it was replaced with a different word, although it is clear that the student did not understand the intention of the feedback. It had been explained very clearly that the first paragraph of this critical essay should constitute an objective summary of the article they had read. It was considered that the student should have known this as it had been mentioned several times in class.

The second category refers to when the learner made the right kind of change but still failed to correct the problem. An example of a feedback point that was incorrectly revised by a student follows.

Original sentence: “After graduate from school, almost people work.”

Feedback: “You have several verb tense errors (underlined).”

Revised sentence: “After graduated from school, almost people works.”

Although the verb tense has been changed in this revision, it is still incorrect and therefore this was coded as ‘incorrect revision’.

The third category refers to cases where the learner revised the text in a way that was both consistent with the intention of the feedback point and corrected the problem previously apparent. An example of a feedback point that was corrected by a student follows:

Original sentence (in an essay comparing the US and the UK): “Also, it is thought that American English is more popular than British English, while British English is more refined.”

Feedback: “You need to support this sentence with examples or details.”

Revised sentence: “Moreover, according to Tan (1997), British people are very conservative in their language, and even in their lifestyle.”

The meaning of this sentence is very different from the original; however, it is supported and was integrated coherently into the revised paragraph. It was therefore coded as ‘corrected’.

As is commented by Silva and Kei Matsuda (2008), studies on uptake often code teachers’ feedback based on its intention although the teacher who gave the feedback is not

directly involved in the research. This can lead to problems in validity. Silva and Kei Matsuda argue that if we are to code feedback for teacher intention, the teachers themselves need to code their comments. In the case of this study, the teacher (who was also the researcher) coded all the feedback and revisions. In order to assure that the coding was done in a consistent way, 20% of the data was recoded by the researcher after a period of two months. The Pearson correlation intra-rater reliability was >0.999 .

Each draft of an essay was treated as one case and the results of all feedback points were calculated as proportions of all the feedback provided on the draft. This ensures that one sample essay does not have more influence on the data than others by virtue of having received substantially more feedback points as would be the case if each feedback point were to be included as a separate case in the data. The proportions ranged from 0 to 1. If, for example, 50% of the feedback led to no revision attempt, then the resulting proportion would be 0.5, whereas if 100% of the feedback was ignored then the proportion would be 1. Ideally, a parametric test such as a one-way ANOVA would be conducted to determine significant difference between the two groups. However, the use of ANOVA for statistical analysis rests on the assumptions of standard distribution and equal variance of the data (Huizingh, 2007). Because the assumption of equal variance was not met for all variables, a non-parametric test was required (Huizingh, 2007). Therefore, the data was analysed using a Mann-Whitney U Test in order to ascertain whether there was any significant difference between the amount of uptake observed in the two groups and whether there was any significant difference in terms of the relative success of that uptake.

6.3.5 Analysis of improvement in writing performance

All students are required to take the writing section of the in-house general English

proficiency test at the beginning and end of each academic year. The test consists of a single argumentative essay written in thirty minutes, in response to a prompt which is created each year by the in-house assessment research group. It is double-rated by around 30 trained raters and the scores are then scaled using many-facet Rasch modelling, to adjust for the relative strictness or leniency of each rater on each rating scale. For example; if a rater is strict when rating the content of essays and lenient when rating grammar, that raters' content ratings would be scaled up, while their grammar ratings would be scaled down to make them close to the mean. The results are used for placing students into proficiency streams as well as to monitor progress in different areas of language proficiency. The students in this study had taken the proficiency test at the end of the previous academic year (approximately two months before the treatment period began) and took it again at the end of the academic year in question (five days after the treatment period ended).

In order to ascertain that the groups were of equal writing proficiency at the beginning of the academic year, the in-house writing test scores from the end of the previous academic year were compared between the two groups using one-way ANOVA. For the purpose of measuring improvement in writing performance over the one year treatment period, gain scores were calculated by taking each student's score for each rating scale at the beginning of the academic year and deducting from it their score on the same scale in the subsequent test administration. A total gain score was also calculated for each student. Subsequently the writing proficiency test gain scores were compared between the group that was assessed on both revision and product throughout the academic year and the group that were assessed on product alone. The assumptions of standard deviation and equal variance were met and one-way ANOVA was used to determine whether there was any statistically significant difference between the gain scores of the two groups.

6.4 Results

The descriptive statistics for the uptake of feedback can be seen in Table 1. ‘Assessed’ refers to the group who were assessed for improvement between drafts in addition to the quality of the final draft, while ‘unassessed’ refers to those who were assessed on the quality of the final draft alone. By combining the results of the assessed and unassessed groups, we can see that overall, 83.7% of the feedback points given on preliminary drafts led to uptake, while the other 16.3% were ignored by the learners. Of the revision attempts, overall just over 1% were misunderstood by learners, while just over 18% were revised with revisions that failed to resolve the apparent problem (incorrect revisions) and 64.5% were revised with the apparent problem resolved (corrected). Table 1. *Descriptive Statistics for results of feedback*

Result	<u>Only product-assessed group</u>		<u>Process-assessed group</u>	
	Mean	SD	Mean	SD
No attempt	0.1842	0.2195	0.1215	0.1734
Attempt	0.8047	0.2182	0.8693	0.1751
Misunderstood	0.0103	0.0615	0.0123	0.0444
Incorrect	0.1473	0.1665	0.2146	0.1783
Correct	0.6471	0.2641	0.6424	0.2355

n =80

RQ1: Does the assessment of process in addition to product lead to more revision attempts than the assessment of product alone?

RQ2: Does the assessment of process in addition to product lead to more successful revisions than the assessment of product alone?

The Mann-Whitney U Test analysis found that there was a significant difference (at the 0.05 level) between the uptake attempts made by students in the two groups, (0.050). The

process-assessed group made significantly more uptake attempts than the only product-assessed group. In terms of the results of those revision attempts, the number of incorrect revision attempts was significantly different (at the 0.05 level) between the two groups (0.011), with the process-assessed group making significantly more incorrect revisions, whereas the difference between the number of misunderstood revisions (0.135) and the number of correct revisions (0.725) were not statistically significant.

Table 2. *Descriptive Statistics for writing pre-test scores*

Type	<u>Only product-assessed group (n =18)</u>		<u>Process-assessed group (n =10)</u>	
	Mean	SD	Mean	SD
Organisation	2.6905	0.6617	3.0071	0.6006
Vocabulary	2.6476	0.5879	2.6714	0.4462
Content	2.6571	0.5240	2.8000	0.4756
Grammar	2.7190	0.3932	2.8500	0.2902
Total	10.6905	1.6778	11.3214	1.2033

The descriptive statistics for the writing proficiency test scores at the end of the previous academic year can be seen in Table 2. Levene's test of equality of variance was employed and found that the data met the assumption of equal variance. Skewness and kurtosis values were also measured and fell between -0.370 and 0.603. These values indicate that the data is sufficiently close to normal distribution to conduct parametric statistics. The one-way ANOVA found no significant difference (at the 0.05 level) between any of the writing test sub scores: organization [F (1, 26) = 2.067, $p = 0.160$], vocabulary [F (1, 26) = 0.017, $p = 0.898$], grammar [F (1, 26) = 1.135, $p = 0.294$], content [F (1, 26) = 0.671, $p = 0.419$], total [F (1, 26) = 1.469, $p = 0.234$].

RQ3: Does the assessment of process in addition to product affect the amount of improvement

in writing performance achieved by learners over a one year period compared to the assessment of product alone?

Descriptive statistics for the writing proficiency test gain scores can be seen in Table 3. Results of the one-way ANOVA found no significant difference (at the 0.05 level) between improvement in organization [$F(1, 26) = 1.839, p = 0.187$], vocabulary [$F(1, 26) = 1.186, p = 0.286$], grammar [$F(1, 26) = 1.852, p = 0.185$], or total gain scores [$F(1, 26) = 3.546,$

Table 3. *Descriptive Statistics for writing test gain scores*

Type	<u>Only product-assessed group (n =18)</u>		<u>Process-assessed group (n =10)</u>	
	Mean	SD	Mean	SD
Organisation	0.6742	0.5484	0.2925	0.9501
Vocabulary	0.6644	0.5939	0.4195	0.5224
Content	0.6519	0.5277	0.1845	0.5451
Grammar	0.6561	0.7050	0.3065	0.5354
Total	2.6467	1.8262	1.2030	2.1487

$p = 0.071$]. On the other hand, there was found to be a statistically significant difference (at the 0.05 level) between the improvement in content [$F(1, 26) = 4.930, p = 0.035$], between the two groups. The only product-assessed group improved significantly more than the process-assessed group.

6.5 Discussion and conclusion

There is a great deal of uncertainty amongst teachers of writing about how best to assess a writing course. While the process approach is commonly adopted in the instruction of writing, process is not always assessed even in courses in which it is emphasised, with many still assigning grades for the final product alone. On the other hand, measurement experts maintain

that measures such as ‘improvement’ or ‘effort’ should not be incorporated into classroom grades because they are too subjective. Meanwhile, incredibly little research has been conducted to determine the effects of different assessment treatments on learners. This study was designed in an attempt to explore whether the assessment of process (in this case, improvement between drafts) in addition to product has any significant effect on learners’ writing processes (in this case, revisions) or long-term improvement in writing performance.

The sample size of this study was very small. Therefore, it is suggested that more research in this area be conducted with a sample size of at least 30 students in each group. If a larger sample size is used, it is possible that more significant differences will be found between the two groups. However, it seems apparent from the results of the analysis of uptake is that the group who were assessed on their revisions attempted to make more use of the teacher feedback they received than the group who were unassessed. However, the number of revisions that were successfully navigated by both groups was almost identical, the difference being seen in unsuccessful revision attempts. Indeed, if we look more closely we can see that the assessed group made almost 7% more revision attempts than the unassessed group. They also made almost 7% more unsuccessful revisions.

According to Loewen (2004), uptake constitutes one form of what Swain (1995) calls ‘pushed output’. Pushed output occurs when learners are forced to modify their output. It appears that in the revision process, the learners who were assessed on their revision in addition to the final product were pushed more than those who were assessed on the final product alone. It would seem that this increased push to modify their output would have a positive effect, leading to more language development in the long term. However, the results of the analysis of variance between groups in terms of writing test scores shows that on average the students in the unassessed group improved more than those in the assessed group

in terms of every writing skill assessed, although only the difference in the improvement in logical consistency, support and development of content was statistically significant.

The finding that the unassessed group improved significantly more than the assessed group is an indication that revisions made after learners receive teacher feedback on their writing do not benefit learners in terms of more successful revisions or improvement in writing performance over the long term. Indeed, the assessment of revision may have had the negative effect of focusing students' attention on the finer details of writing, such as surface level issues, at the expense of the overall quality of the writing, such as the support and development of ideas. When faced with a great deal of feedback from the teacher, the group who were assessed on their improvement between drafts seem to have been pushed to attempt to revise based on every feedback point received. If they received more feedback on grammar and less on content, then this may similarly push them to spend more time revising grammar and less time on content revision.

However, clearly feedback on content is more substantive than that on grammar. Whereas feedback on grammar can require changing just one or two letters within a word, even one instance of feedback on content can necessitate largely rewriting the entire essay. Students who were assessed on feedback between drafts may have therefore placed importance on addressing as many feedback points as possible in the time they had available, which would prioritize feedback on grammar over feedback on content, as a large number of grammatical feedback points can often be revised in the same time it takes to revise just one content issue. On the other hand, the group that was assessed on only the final draft of their writing, learners may have focussed on what they perceived to be the most important feedback points in order to make the essay the best it could be, and ignored surface-level issues if they were not sure how to revise them or if they did not have enough time to address all feedback.

It is worth noting that assessing improvement between drafts in addition to final written product, as would be expected, did take longer for the instructor than assessing the final written product alone. However, the additional time required to assess the improvement between drafts was not a great deal of time. It took a few minutes per essay to assess the improvement between drafts, and usually less than five minutes. Therefore, the additional time required for the whole academic year was around 30 minutes per student. This would certainly be a worthwhile use of instructor time if it could be demonstrated that the assessment of improvement between drafts resulted in increased uptake or improvement in writing performance.

It is gratifying to see that learners in the unassessed group still used the feedback received from the teacher to the extent that they could. This shows that they recognized the importance of the revision process and therefore the actual assessment of revision may not be necessary. However, the learners in this study were majoring in English and had a high level of intrinsic motivation which may have pushed them to revise. Learners with less intrinsic motivation may require additional extrinsic motivation, in the form of assessment of revision, in order to push them to make more effort. Therefore, the assessment of process in addition to final written product may have more effect in a context in which learners have less intrinsic motivation.

The learners in this study were required to revise their writing and submit all three drafts to the instructor for grading. This requirement to submit all three drafts may have been enough to push the learners in the unassessed group to make effort to address the teacher feedback. Although they did not receive a grade for the improvement between drafts, and this was communicated to the students, they may have assumed that the instructor would compare the three drafts in arriving at the final grade and therefore, made effort in their revisions. It

would be interesting to consider a situation in which learners in the unassessed group were not even required to submit their preliminary drafts, whereas the assessed group submitted theirs and were assessed on their improvement between drafts in addition to the final written product. However, this may not be feasible as the preliminary drafts are required by the researcher in order to evaluate (successful) uptake of feedback.

Furthermore, there may be different educational practices that can lead to more effective pushed output. For example, students could be required to keep an error log in which they record all the feedback they receive. Different variations of this could be employed such as asking students to correct their errors, to simply keep a tally in order to understand the kinds of errors they are making, or they could be asked to actually use this information to inform their self-directed learning. These kinds of activities require students to engage more deeply with the feedback they receive and it is easy to imagine that these kinds of activities may have more effect in terms of long-term improvement in writing performance, even if not in terms of (successful) uptake of feedback.

References

- Ashwell, T. (2000). Patterns of Teacher Response to Student Writing in a Multi-Draft Composition Classroom: Is Content Feedback Followed by Form Feedback the Best Method? *Journal of Second Language Writing* 9, 227-257.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118.
- Bitchener, J. & Knoch, U. (2009). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31(2), 193-214.
- Bitchener, J. & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19, 207-217.

- Bonesronning, H. (2004). Do the teachers' grading practices affect student achievement?
Education Economics, 12 (2), 151-167.
- Brookhardt, S. (2004). *Grading*. Upper Saddle River, NJ: Pearson Education.
- Brookhart, S. (1991). Letter: Grading practices and validity. *Educational measurement: Issues and practice*, 10(1), 35-36.
- 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.
- Chaudron, C. (1984). Evaluating Writing: Effects of feedback on revision. *RELC Journal*, 15, 1-14.
- Ellis, R., Basturkmen, H. & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51(2), 281-318.
- Fathman, A. & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.) *Second Language Writing: Research Insights for the Classroom*. (pp. 178-190). Cambridge: Cambridge University Press.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: The University of Michigan Press.
- Ferris, D. (2004). The "Grammar Correction" debate in L2 writing: Where are we and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing* 13, 49-62.
- Ferris, D. (2006). Does error feedback help student writers? In K. Hyland & F. Hyland (Eds.). *Feedback in Second Language Writing: Contexts and Issues* (pp. 81-104). Cambridge: Cambridge University Press.
- 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. & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing* 10, 161-184.
- Flower, L. & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387.
- Gearhart, M. & Wolf, S. (1997). Issues in portfolio assessment: Assessing writing processes from their products. *Educational Assessment*, 4(4), 265-296.
- Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher and student variables. In Hyland, K. & Hyland, F. (Eds.). *Feedback in Second Language Writing: Contexts and issues* (pp 185-205). New York: Cambridge University Press.
- Green, S., Johnson, R., Kim, D. & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and teacher education*, 23, 999-1011.
- Griswold, P. (1993). Beliefs and influences about grading elicited from student performance sketches. *Educational Assessment*, 1(4), 311-328.
- Gullickson, A. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research*, 79(2), 96-100.
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.). *Assessing Second Language Writing in Academic Contexts* (pp 241-276). Norwood: Ablex Publishing Corporation.
- Hills, J. (1991). Apathy concerning grading and testing. *Phi Delta Kappan*, 72, 540-545.
- Huizingh, E. (2007). *Applied statistics using SPSS*. London: Sage.
- Hunter, D., Mayenga, C. & Gambell, T. (2006). Classroom assessment tools and uses:

- Canadian English teachers' practices for writing. *Assessing Writing*, 11, 42-65.
- Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294-304.
- Linn, R. & Miller, M. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Loewen, S. (2004). Uptake in incidental focus on form in meaning focused ESL lessons. *Language Learning*, 54(1), 153-188.
- Long, M. (2007). *Problems in SLA*. Mahwah: Lawrence Erlbaum.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66.
- Mackay, A. & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses and red herrings. *Modern Language Journal*, 82(3), 338-356.
- Ohta, A. (2000). Re-thinking recasts: a learner-centred examination of corrective feedback in the Japanese language classroom. In J. Hall & L. Verplaeste (Eds.) *The construction of second and foreign language learning through classroom interaction* (pp. 47-71). Mahwah: Erlbaum.
- Raimes, A. (1993). Out of the woods: Emerging traditions in the teaching of writing. In S. Silberstein (Ed.) *State of the art TESOL essays* (pp. 237-260). Alexandria: TESOL.
- Randall, J. & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and teacher education*, 26, 1372-1380.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education*, 12, 315-325.
- Romova, Z. & Andrew, M. (2011). Teaching and assessing academic writing via the portfolio: Benefits for learners of English as an additional language. *Assessing Writing*, 16, 111-

- Sachs, R. & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29, 67-100.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across educational settings. *Language Teaching Research*, 8(3), 263-300.
- Sheen, Y., Wright, D., & Moldova, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System* 37(4), 556-569.
- Silva, T. & Kei Matsuda, P. (Eds.). (2008). *On Second Language Writing*. Mahwah: Routledge.
- Stiggins, R., Frisbie, R. & Griswold, P. (1989). Inside high school grading practices: Building a research agenda. *Educational measurement: Issues and practice*, 8(2), 5-14.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.) *Principles and practices in the study of language* (pp. 125-144). Oxford: Oxford University Press.
- Tierney, R., Simon, M. & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards based grading. *The Educational Forum*, 75, 210-227.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16, 255-272.
- Vaughan, C. (1991). Holistic assessment: What goes on in the raters' mind? In L. Hamp-Lyons (Ed.) *Assessing Second Language Writing in Academic Contexts* (pp 111-125). Norwood: Ablex Publishing Corporation.
- Walker, R. & Perez-Riu, C. (2008). Coherence in the assessment of writing skills. *ELT Journal*, 62(1), 18-28.

- Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, 29, 325-340.
- Xiaoxiao, L. & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33(1), 24-40.
- Xiaoxiao, L. & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33(1), 24-40.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209-222.
- Zhang, Z. & Burry-Stock, J. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.

Appendix 6.A

Feedback form

Writer's name: _____

Reader's name: _____

Editing Checklist

	<u>Yes</u>	<u>No</u>
1. _____	<input type="checkbox"/>	<input type="checkbox"/>
2. _____	<input type="checkbox"/>	<input type="checkbox"/>
3. _____	<input type="checkbox"/>	<input type="checkbox"/>
4. _____	<input type="checkbox"/>	<input type="checkbox"/>
5. Give one piece of constructive feedback.		

Appendix 6.B

Writing test essay rating scales

	Organisation Think about: <ul style="list-style-type: none"> ▪ Coherence among ideas/sentences ▪ Overall structure 	Lexis Think about: <ul style="list-style-type: none"> ▪ Range ▪ Variety ▪ Preciseness 	Content Think about: <ul style="list-style-type: none"> ▪ Relevancy to the main idea ▪ Strength of support ▪ Convincingness 	Grammar Think about: <ul style="list-style-type: none"> ▪ Range ▪ Clarity ▪ Accuracy
1-1.5	No coherence or organisation, unconnected sentences which communicate little.	Demonstrates minimal word knowledge. Vocabulary is too limited to effectively complete task.	Information provides little or no support to the main idea. Ideas are all pretty much irrelevant. It's hard to find a main argument or if there is one it's hardly supported at all.	Many grammatical inaccuracies make message difficult to understand and/or very simple sentences with some errors.
2-2.5	Some attempts to organise information but with little connection between ideas and sentences apparent.	Uses a limited variety of vocabulary with little control. Has enough vocabulary to complete task in a very limited manner.	A little support for main idea, but mostly the information is not relevant or provides very weak support. The main argument is not convincing.	Inadequate range of grammar used to make ideas clear and/or much grammar is repetitive or inaccurate.
3-3.5	Obvious attempts to organise information though sometimes the lack of coherence between ideas and/or sentences creates ambiguity.	Uses an adequate variety of vocabulary with moderate control. Has enough vocabulary to complete task but not very effectively.	Most information is relevant but the support is quite weak. The main argument is a little convincing.	An adequate range of grammar is used, but some inaccuracies impede the clarity of the meaning of sentences.
4-4.5	The writing displays an organisational structure which enables the message to be followed. Ideas and sentences are connected but sometimes the lack of coherence might create ambiguity.	Uses a wide variety of vocabulary but there are some inaccuracies in word choice and/or formation. Has enough vocabulary to complete the task somewhat effectively.	Most or all information is relevant to the main argument. The support is quite strong. The main argument is somewhat convincing.	An adequate range of grammar is used and grammatical errors rarely impede the meaning of sentences. A fair number of errors which do not impede meaning are present.
5	The writing displays a coherent organisational structure which enables the message to be followed effortlessly at the sentence, paragraph and essay levels.	Uses a wide variety of vocabulary, quite accurately and with good control. Has enough vocabulary to complete task effectively.	All information is relevant and well-supported. The main argument is very convincing.	A wide range of grammar is used. Meaning is not negatively impacted by the grammar. A few errors which do not impede meaning may be present.

CHAPTER SEVEN

Study Six

The Influence of Assessment of Classroom Writing on Feedback Processes and Product vs. on Product Alone

Abstract

Although second language writing classes mainly use a process approach, anecdotal evidence suggests that assessment of writing often still focuses on the written product alone. This assessment practice continues despite specialists having recommended that both process and product be assessed. This study compares second-year university students in Japan who were assessed on feedback processes and product with others assessed on product alone in terms of perceptions of the feedback received. Perceptions were determined through a post-treatment questionnaire. Neither the assessment of the use of teacher feedback in revisions nor the assessment of the quality and quantity of peer feedback was found to have a clear benefit in terms of students' perceptions of the feedback received. This finding suggests the need for further research to confirm whether the assessment of both process and product is worth the considerable time investment required.

7.1 Introduction

Of particular importance within the process approach to writing instruction are drafting, feedback, and the use of that feedback in revision. However, while the emphasis of instruction is often on the writing process, my personal communication with numerous instructors suggests that a large number of writing instructors still use a strict product approach when assessing the learning that has taken place. Furthermore, it is extremely difficult to find any

literature discussing ways of assessing the processes used by learners during the writing process. This is despite the fact that many books published on the instruction of writing to second language learners, recommend that the process as well as the written product should be assessed (e.g. Ferris, 2002; 2003; Ferris & Hedgcock, 2005; 2014; Goldstein, 2006; Hyland, 1996; Leki, 1990; Mittan, 1989).

Depending on the educational context in question, it may be desirable for grades to reflect achievement alone or it may be preferable for them to reflect both the amount of effort made and achievement. If the educational context calls for the assessment of both effort and achievement through grading and if the process approach to writing is being used, there are several ways in which a writing course grade can be made to reflect both effort and outcomes. Student learning can be assessed through the quality of the final draft or through scores from timed writing tests. Effort can be assessed through the completion of each step of the writing process (i.e. whether the student wrote an outline, wrote the required number of drafts, carried out peer review, etc.) However, as suggested by Ferris (2002; 2003), Ferris and Hedgcock (2005; 2014), Hyland (1996) and Mittan (1989), assessing whether each part of the writing process was completed merely encourages students to complete each step of the process; it does not necessarily encourage them to consider the teacher or peer feedback carefully or to take it seriously. (Much less does it measure achievement.) On the other hand, assessing feedback processes with a number grade may communicate the importance of the feedback process to students, thereby encouraging them to take the revision process more seriously.

The purpose of the current study is to determine whether assessing feedback processes in addition to the final written product is beneficial for learners in terms of perceptions of the feedback they received over the treatment period of one academic year (10 months). In the context of teacher feedback, whether or not learners pay attention to feedback when revising is

important (Chandler, 2003; Chaudron, 1977; Fazio, 2001; Leki, 1990). Teacher feedback which receives no learner attention is unlikely to be beneficial for learners. For this reason, learners' use of teacher feedback in their revisions was selected as the feedback process to assess in the teacher feedback group in the present study of the effects of different ways of assessing writing. In the context of peer feedback, on the other hand, it has been stated that it is the giving rather than the receiving of the feedback that benefits learners most (Lundstrom and Baker, 2009). Clearly, feedback from peers needs to be carefully evaluated rather than simply being adopted by writers as teacher feedback usually is (Rollinson, 2005; Tsui & Ng, 2000). For these reasons, the quality and quantity of feedback provided to peers was selected as the feedback process to assess in the peer feedback group.

This study examined the influence of the assessment of process in terms of learner perceptions of the feedback received, and three different learner perceptions were focused on: the amount of feedback, the helpfulness of the feedback, and the comprehensibility of the feedback. These represent the three main potential problems of feedback highlighted in the literature: that students receive too much feedback (Evans, Hartshorn, McCollum & Wolfersberger, 2010; Hairston, 1986; Lee, 1997); that the feedback they receive is not helpful (Chiang, 2004; Cohen, 1987; Hyland, 2000; Straub, 1997; Zamel, 1985); and that learners do not understand the feedback they receive (Cohen & Cavalcanti, 1990; Ferris, 1995; Hyland & Hyland, 2001; Zacharias, 2007). The research questions addressed in this study are:

- (1) When receiving systematic teacher feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on both their use of the teacher's feedback and the final draft of their writing have different perceptions of the feedback they have received than will those who are assessed on the final draft alone?

- (2) When receiving systematic peer feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on quality and the quantity of feedback they give in addition to the final draft of their writing have different perceptions of the feedback they have received than will those who are assessed on the final draft alone?

7.2 Background

Over the preceding decades and in particular with the publication of the Flower and Hayes (1981) model of process writing, there was a shift of focus within the field of second language writing. Previously, emphasis was invariably placed on the end product whereas later emphasis has often been on the process of creating a text. Of particular importance within the process approach to writing instruction are drafting, feedback, and the use of that feedback in revision. According to Zhang (1995: 209), “As a recursive model, the process approach focuses on how to revise in response to feedback from the reader.” The emergence of the process approach to writing in second language contexts has also seen an increase in the types of feedback used in writing classrooms.

As mentioned by Hamp-Lyons and Condon (2000), prewriting, peer and teacher feedback, and revision are of crucial importance in process-based approaches to writing. However, while the emphasis of instruction within the process approach to writing is on the process, my personal communication with numerous instructors suggests that a large number of writing instructors still use a strict product approach when assessing the learning that has taken place. Although literature on classroom assessment of writing within the process approach has been published, a great deal of this literature describes ways of assessing the final drafts of essays (e.g. Archibald, 2011; Ferris & Hedgcock, 2005). On the other hand, it is extremely difficult to find any literature discussing ways of assessing the processes used by learners during the

writing process.

Even within the literature on portfolio assessment, which offers a rich opportunity for instructors to evaluate each step of the writing process, almost all of the literature mentions solely assessment of the quality of the final draft of writing, such as in the areas of language use, organization, and content (Conrad, 2001; Ferris & Hedgcock, 2005; Freedman, 1993; Gearhart & Wolf, 1997; Nystrand, Cohen & Dowling, 1993; Romova & Andrew, 2011). Hamp-Lyons and Condon (1993) expected instructors in their study to assess the extent to which students had improved their writing through the drafting process. However, they found that most instructors did not require students to include preliminary drafts in their portfolios; and some instructors actually penalized students if the final draft of their multiple draft writing was better than their timed writing, believing that the timed writing reflected their true ability.

While there is little or no literature discussing the direct assessment of the process involved in a learner's writing (Xiaoxiao & Yan, 2010), some literature does attempt to give credit for progress in writing over the duration of a course. The assessment of progress may be another way for grades to reflect both effort and performance in a writing course, as it could be assumed that those who make more effort will progress more. Ferris and Hedgcock (2005: 337), for example, state that "...teachers can make room in their grading policies for fair assessment of both products and progress...by weighting assignments given later in the term more heavily than those given in the beginning." While assessment of progress in addition to product should encourage learners to make effort in the writing process, it seems that it would be more appropriate for learners who are already highly autonomous and know which strategies to use and how to use those strategies to the best effect. On the other hand, assessing certain processes that learners go through in the production of final drafts can be seen as scaffolding in that it demonstrates the importance of those processes within the practice of

writing. Even less autonomous learners may spend time to develop revision strategies when they realize how important the revision process is. In contrast, if these feedback processes are not assessed, less autonomous learners may not see the importance of these processes to the development of their writing skills.

Although it is difficult to find literature discussing the assessment of process in the process writing classroom, many books published on the instruction of writing for second language learners recommend that the process should be assessed. It has been suggested that the students' use of teacher feedback in the revision of their texts should be evaluated. Ferris (2002) states that assigning a grade for the utilization of teacher feedback during the revision process encourages students to attend to accuracy in their writing. Goldstein (2006) conducted a case study of two students and found that one of them often removed portions of the text that had received teacher feedback rather than revise them and also simply ignored a lot of the teacher feedback provided. Goldstein suggested that one reason for this was that the instructor graded only the final draft and so the learner felt that she would not get credit for time spent revising. The corollary of this is that if writing teachers assess the extent to which the student has paid attention to the teacher feedback received, the learner may invest more in the revision process.

Quite a number of writers also indicate that students may go through the motions of the peer review process without really making an effort to give good feedback if that process is not evaluated by the teacher. Mittan (1989), Leki (1990), Hyland (1996), Ferris (2003), and Ferris and Hedgcock (2005) all recommend the evaluation of feedback given by a learner to his/her peer. The assessment of peer feedback through a number of points or a percentage of the grade being given for the feedback emphasizes the importance of peer feedback to learners (Ferris, 2003). Ferris and Hedgcock (2005) suggest making students accountable for the

feedback they give to peers by providing grades on quality and substance of the peer feedback given. It is suggested that this helps students to take the peer review process seriously. Mittan (1989) and Hyland (1996) also suggest that grades be awarded for the feedback given by a learner, while Leki (1990) recommends that teachers read all of the feedback given by each student and then provide their own feedback on the students' feedback.

While Ferris and Hedgcock (2005) in their chapter on peer response recommend holding students accountable for their peer response by assigning a grade, in the chapter on classroom writing assessment the discussion focuses on the assessment of “student-generated texts” using a variety of different scoring methods, with only a few mentions of assessment of progress (the most substantial of which is quoted above) and no discussion of how instructors might integrate scores assigned for peer response into their classroom grades. The revised edition (Ferris & Hedgcock, 2014), which was published after this article went to press does discuss portfolio assessment in addition to the assessment of “student-generated texts”. In relation to portfolio assessment, it discusses giving students opportunities to reflect on their writing and to continue to revise their writing after it has been submitted. However, it does not explain whether instructors should assess process in addition to product and if so, how they could go about doing so.

7.3 The Current Study

7.3.1 Context and participants

The participants in the current study were second-year students majoring in English at a university of foreign studies in Japan. The university has a strict English-only policy in place, so all classes and communications between teachers and students are conducted in English. This policy is popular with students as it increases chances to improve their language

proficiency. The policy is one of the reasons students give for having chosen the university. However, in an EFL environment it does pose a certain level of challenge to some learners, particularly those at lower proficiency levels. The participants involved in the present study were in four classes consisting of 26 or 27 students each; in total there were 105 students in the four classes. Of the 105 students in the four classes, 34 did not give their consent to participate in the study, so the total pool of participants was 71.

In their first year at the university, students had focused on writing paragraphs for one semester and writing basic five-paragraph essays for one semester. In the second year, the curriculum involved writing essays in different rhetorical modes in the first semester and a research paper in the second semester. The students were in the bottom ability tier at the university based on their Test of English as a Foreign Language (TOEFL) Institutional Testing Programme (ITP) scores as well as their scores on the writing and speaking sections of an in-house proficiency test. Within the bottom tier students were placed into classes randomly. Therefore, it is reasonable to believe that the ability of the students in the four classes was roughly equal at the beginning of the academic year. Their ability can be described as ranging from pre-intermediate to upper-intermediate level. Their TOEFL ITP scores ranged from 370 to 500.

7.3.2 Feedback Treatment

Two of the classes received only teacher feedback on all preliminary drafts for one academic year (16 drafts in total), while the other two gave and received peer feedback exclusively on all preliminary drafts over the same period. In order to encourage learner autonomy, learners in this study thought of their own questions for feedback. They were given a blank feedback checklist and instructed to think of and write four questions that they would like their teacher

or peer reader to answer. The type of questions to ask was left completely up to the writers themselves. For example, all of the questions could relate to rhetorical issues or they could all relate to surface errors, depending on the writer's needs and preferences. The teacher and peer readers marked problems found in each draft related to the questions asked. In addition to this, every feedback checklist had one prompt at the bottom "Give one piece of constructive feedback." This gave the teacher or peer reader some freedom to provide feedback on one issue they noticed while reading the draft.

Apart from the feedback given on the feedback form, learners were free to ask questions to each other or to the instructor. However, the questions had to be specific. The instructor did not offer comprehensive oral feedback nor any feedback in response to requests such as "Please check my grammar". On the other hand, students would receive oral feedback from the teacher if they asked a specific question such as "In this sentence is it better to use the past simple tense or the present perfect tense?", or "Can you understand what I want to say in this sentence?"

Students in the peer feedback group took part in a lesson on the purposes of peer feedback and how to give useful feedback, while those in the teacher feedback group were instructed on how to understand the teacher feedback and use it in their revisions. During the lesson on peer feedback, learners discussed the meaning of the word *constructive* in detail and practiced giving constructive feedback. They were told that, in answering the prompt, any problem they noticed in the draft which was not covered by the questions asked by the writer would be a suitable basis from which to develop a piece of constructive feedback. In practice, some peer readers did not give any additional constructive feedback and some gave several additional pieces of constructive feedback. The instructor tried to choose the most important point, but occasionally gave two additional pieces of constructive feedback.

7.3.3 Assessment

In order to determine the effect of assessment of feedback processes and product as opposed to the assessment of product alone, one teacher feedback class group was assessed on their use of teacher feedback in their revisions and the final draft of their writing while the other was assessed on the final draft of their writing alone. Students in the process assessed teacher feedback class group received scores out of a possible 20 points which included scores out of a possible 5 points each for: revision process, grammar and vocabulary, organization, and content. The score for revision process was determined by the proportion of teacher feedback points attended to during revision and the overall quality of the changes made.

Similarly, one of the peer feedback class groups was assessed on the feedback they gave to their partner and the final draft of their writing while the other was assessed on the final draft of their writing alone. Students in the process assessed peer feedback class group received scores out of a possible 20 points which included scores out of a possible 5 each for: peer feedback, grammar and vocabulary, organization, and content. The scores for peer feedback were determined by the number of comments given, the amount of detail in the comments, and the overall quality of the comments.

The two class groups (one teacher feedback class group and one peer feedback class group) who were assessed on their final draft alone received scores out of a possible 15 points which included scores for: grammar and vocabulary, organization, and content. As is common with grading of student writing, the grades given were subjective, reflecting the teacher's overall judgment of the writing in terms of each category.

Previous studies (e.g. Ruegg, Fritz & Holland, 2011) have found that grammar and vocabulary are difficult to distinguish and studies in the field of corpus linguistics suggest that

lexis and grammar are fundamentally inseparable (e.g. Hoey, 2005; Römer, 2009). For these reasons, grammar and vocabulary were assessed in combination, using a single analytic rating scale.

7.3.4 Instrument

In order to determine the effect of the different assessment methods on students' perceptions of the feedback they had received, a post-treatment questionnaire was used. The post-treatment questionnaire was completed at the end of the academic year, after all assignments had been graded and returned to students. The questionnaire was written in Japanese and was based on the one used by Ruegg and Koyama (2011).

Only students who gave consent to be included in the study and who were present on the day of the post-treatment questionnaire were included in this second part of the study. Of the 71 participants in the study, 4 were absent from class on the day of the post-treatment questionnaire. After removing students who missed the post-treatment questionnaire, 22 students remained in the only product-assessed teacher feedback group and 15 remained in the process-assessed teacher feedback group, while 15 remained in both the only product-assessed and the process-assessed peer feedback groups.

The post-treatment questionnaire contained 11 questions in two sections. The Cronbach's alpha reliability of the post-treatment questionnaire as a whole was 0.696 (Ruegg, 2014). Principal components analysis extracted two factors for this questionnaire. Questions 5 –11 comprise Factor 1, which was identified as *perceptions* and accounted for 32% of total variance. Questions 1 – 4 comprise Factor 2, which was identified as *self-efficacy* and accounted for 17% of variance. The focus of this study is students' perceptions and therefore this study focuses on the seven questions in Section Two of the post-treatment questionnaire,

which can be seen in Appendix A and Appendix B. The questions asked about either teacher feedback or peer feedback depending on which of the treatment groups the participant was in.

Principal Components Analysis rests on the assumptions that the sampling size is adequate, that there is a linear relationship between all variables, that correlations exist between all variables and that there are no outliers. In order to test these assumptions, the Kaiser-Meyer-Olkin measure of sampling adequacy was used, correlations between variables were investigated by examining the correlation matrix and by running Bartlett's Test of sphericity. In addition, descriptive statistics were calculated and used to determine whether there were any outliers.

7.3.5 Data Analysis

The questionnaire data was ordinal data rather than interval data and therefore, a non-parametric MANOVA was employed in order to determine whether the two different assessment methods had any significant effect on students' perceptions of the feedback they had received.

7.3.6 Suitability of the questionnaire data for Principal Component Analysis

The Kaiser-Meyer-Olkin Test of sampling adequacy for the questionnaire resulted in a score of 0.822. According to Field (2013), this is a good score and indicates that the sample size is more than adequate. The diagonal anti-image values ranged from 0.592 to 0.905 and were therefore above the minimally acceptable value of 0.5 (Field, 2013). It was found that all seven questions discussed in this article correlated at least 0.3 with at least three of the other questions, suggesting that it is reasonable to employ Principal Components Analysis.

Bartlett's test of Sphericity had a highly significant result of <0.001 . Moreover, there were no significant outliers in the data. It was therefore determined that the data was suitable for the

employment of Principal Components Analysis.

7.4 Results

7.4. 1 Teacher Feedback Groups

Descriptive statistics for questions 5–11 of the post-treatment questionnaire for the teacher feedback groups can be seen in Table 1.

RQ1: When receiving systematic teacher feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on both their use of the teacher's feedback and the final draft of their writing have different perceptions of the feedback they have received than will those who are assessed on the final draft alone?

The non-parametric MANOVA revealed no significant difference (at the 0.05 level) between the process-assessed and the only product-assessed teacher feedback groups in terms of their perceptions of the amount of teacher feedback received; Pillai's Trace = 0.329, $F(7, 29) = 2.027$, $p = 0.086$. This indicates that, in terms of teacher feedback, the group who were assessed on their use of teacher feedback did not differ significantly from the group who were assessed on their final product alone in their perceptions of the amount of teacher feedback they received, the amount of peer feedback received, how helpful the feedback was or how much of the teacher feedback they understood. However, it is interesting to note that the only product-assessed teacher feedback group perceived the teacher feedback to be slightly more useful in terms of all four writing skills than the process-assessed teacher feedback group did.

A post-hoc analysis of power, using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), showed that in order to identify even a large effect in terms of differences in perceptions of the feedback between the two groups, a total sample size of 64 would be

Table 1. Descriptive Statistics for Perceptions of Feedback in the Teacher Feedback Groups

Question	<u>Only product-assessed (n =22)</u>		<u>Process-assessed (n =15)</u>	
	Mean	SD	Mean	SD
Amount of teacher feedback	2.77	0.528	3.13	0.516
Amount of peer feedback	1.86	0.834	2.27	0.458
Usefulness for grammar	2.91	0.294	2.80	0.561
Usefulness for organization	2.86	0.351	2.80	0.561
Usefulness for support	2.77	0.429	2.73	0.594
Usefulness overall	2.91	0.294	2.80	0.414
Comprehension	4.09	0.750	3.67	1.234

required. Although these differences were not statistically significant, it is likely that the lack of significant results was due to the small sample size rather than no effect being present.

Therefore, although no significant difference was found, these slight trends in the data suggest the need for further investigations with a larger number of participants.

7.4.2 Peer Feedback Groups

Descriptive statistics for questions 5–11 of the post-treatment questionnaire for the peer feedback groups can be seen in Table 2.

RQ2: When receiving systematic peer feedback on every preliminary draft of every essay for a period of one academic year, will students who are assessed on quality and the quantity of feedback they give in addition to the final draft of their writing have different perceptions of the feedback they have received than will those who are assessed on the final draft alone?

The nonparametric MANOVA revealed no significant difference (at the 0.05 level) between the process-assessed and the only product-assessed peer feedback groups in terms of

Table 2. Descriptive Statistics for Perceptions of Feedback in the Peer Feedback Groups

Question	<u>Only product-assessed (<i>n</i> =15)</u>		<u>Process-assessed (<i>n</i> =15)</u>	
	Mean	SD	Mean	SD
Amount of teacher feedback	2.67	0.488	3.00	0.756
Amount of peer feedback	2.13	0.640	2.00	0.655
Usefulness for grammar	2.20	0.414	2.47	0.516
Usefulness for organization	2.00	0.535	2.20	0.561
Usefulness for support	2.00	0.756	2.07	0.594
Usefulness overall	2.13	0.516	2.53	0.516
Comprehension	3.60	0.986	3.53	0.834

their perceptions of the amount of peer feedback received; Pillai's Trace = 0.245, $F(7, 22) = 1.021$, $p = 0.444$. This indicates that, in terms of peer feedback, the group who were assessed on the peer feedback they gave did not differ significantly from those who were assessed on product alone in their perceptions of the amount of peer or teacher feedback they received, the usefulness of the feedback received or the amount of peer feedback they understood.

However, in contrast to the teacher feedback groups, in the peer feedback groups the process-assessed peer feedback group found the feedback to be slightly more useful for all four skills than the only product-assessed peer feedback group.

Once again, while these differences were not statistically significant, this lack of significant findings may have resulted from the small number of participants in this study. Therefore, further investigations with a larger number of participants are suggested in order to clarify this discrepancy.

7.5 Discussion and Conclusion

Although the assessment of students' use of teacher feedback in revisions has been recommended in books on second language writing (e.g. Ferris, 2002; Goldstein, 2006), the assessment of students' use of teacher feedback in their revisions increased the grading time needed by the teacher in this study. Previous studies (Chandler, 2003; Chaudron, 1977; Fazio, 2001; Leki, 1990) have shown that whether or not students pay attention to teacher feedback when revising their writing is important. Therefore, if it can be shown that the assessment of the use of teacher feedback encourages learners to pay more attention to the feedback, then it would be a worthwhile use of teacher time. Clearly, students are likely to pay more attention to feedback if they consider the feedback to be useful and comprehensible. However, in comparing students' perceptions of the teacher feedback, no significant differences were found between the process-assessed and only product-assessed groups. Therefore, it may not be worthwhile to spend the time needed to assess the use of teacher feedback. Unless future research can find benefits of assessing the use of teacher feedback in revisions, it may be better to offer teacher feedback to support writers in the development of writing skills but to leave them to use that feedback as they please, rather than encouraging them to use the feedback through assessing their use of it.

The assessment of peer feedback has been recommended widely in second language writing books (Ferris, 2003; Ferris & Hedgcock, 2005; Hyland, 1996; Leki, 1990; Mittan, 1989) based on the premise that the assessment of peer feedback leads to students taking the process more seriously and trying in earnest to give the best feedback they can. However, previously this premise has not been tested by research. If it could be found that the assessment of peer feedback did encourage students to take the process more seriously, then it would be a worthwhile use of teacher time and effort to assess the feedback given during peer

feedback sessions. However, while assessment indicates the importance of revision, and therefore pushes learners to be actively involved in the feedback process, it does not show them how to improve their feedback processes. It appears that some learners may require detailed training in addition to encouragement and practice in order to improve their revision skills. Therefore, rather than giving an overall assessment of the quality and quantity of feedback given, it may be better to offer detailed peer review training and then assess students on the extent to which they followed the advice given during training.

Intuitively, it would seem better if each learner could receive detailed qualitative feedback about their strengths and weaknesses in terms of feedback processes, as suggested by Leki (1990) in regard to peer feedback, rather than a number grade. At the same time, the assessment of peer feedback is very time-consuming. Therefore, it seems likely that many writing teachers do not have enough spare time and energy to provide the kind of qualitative feedback suggested by Leki (1990). In an authentic classroom context a number grade may be all the feedback time allows. Providing a number grade is a much less time-intensive solution and, in this study, even that was found to be quite demanding on teacher resources. Students needed to keep all drafts and the peer feedback they receive to submit for grading along with the final draft. Moreover, the teacher needed to not only read all the feedback and the preliminary drafts, but also evaluate both the quantity and quality of the comments given. Clearly, teachers need to consider whether the benefits of assessing peer feedback outweigh the extra time and effort necessary to assess it. This study did not find the assessment of peer feedback in the form of number grades to lead to any differences in terms of perceptions of the amount of feedback, the usefulness of that feedback or the amount of feedback learners understood and this suggests that it may not be a worthwhile use of resources. Therefore, one potential solution may be to ask students to give peer feedback and to support them in the

process by giving qualitative oral feedback on the peer feedback they give during the peer feedback sessions, but not to collect nor read the actual feedback given by peers.

This was a small study, with only 15 to 22 students in each group and students from only one cultural and educational context at similar proficiency levels. While this study did not find any significant differences between the perceptions of students who were assessed on both process and product when compared to those who were assessed on final written product alone, the small number of students in each group may have led to no significant differences when in fact differences were present. Clearly, a larger scale study is required to provide evidence concerning the value of the assessment of feedback processes in process writing classes. However, until such empirical evidence is available, it may be better to offer teacher feedback and provide support for peer feedback but to assess only the final written product.

Learners from different cultural and educational contexts may also react differently to different methods of assessment, and thus similar studies in different cultural and educational contexts would certainly be fruitful. Additionally, students at different proficiency levels may be affected differently by the assessment of feedback processes. Specifically, teacher feedback would be more comprehensible to those at higher proficiency levels, and students at higher proficiency levels may be able to give better peer feedback, leading to more positive perceptions of that feedback. A study comparing students at different proficiency levels could therefore prove valuable.

The purpose of formative assessment is for learners to improve their learning. However, we cannot simply assume that the assessment practices we implement will lead to the outcomes we aim for. Therefore, it is necessary to empirically test the effects of our assessment practices and reconsider them if they are not resulting in the desired outcomes. In

this case, it was hoped that the assessment of feedback processes would illustrate the importance of those processes to students. It was hoped that this would lead to more positive perceptions of feedback and more serious engagement in the feedback processes, which in turn would improve learning outcomes. This study did not find any benefit in the assessment of peer feedback or the assessment of use of teacher feedback in revision over the assessment of final product alone in terms of learner perceptions. Until further studies are conducted and the benefits of the assessment of process in addition to final written product are ascertained, the results of this study suggest the possibility that it may not be worthwhile for teachers to spend the time required to assess the use of teacher feedback or the feedback given by peers in the process-writing classroom. Since many teachers are investing a great deal of time and effort into the assessment of process, further research on a larger scale is urgently needed.

References

- Archibald, A. (2011). *Writing in a second language*. Retrieved 16th September 2011 from the Centre for Languages, Linguistics and Area Studies website:
<http://www.llas.ac.uk/resources/gpg/2175>.
- 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. [http://dx.doi.org/10.1016/S1060-3743\(03\)00038-9](http://dx.doi.org/10.1016/S1060-3743(03)00038-9).
- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning* 27, 29–46. <http://dx.doi.org/10.1111/j.1467-1770.1977.tb00290.x>.
- Chiang, K. (2004). An investigation into students' preferences for and responses to teacher feedback and its implications for writing teachers. Retrieved 16 September 2011 from <http://edb.org.hk/HKTC/download/journal/j3/10.pdf>.

- Cohen, A. (1987). Student processing of feedback on their compositions. In A. Wendin and J. Rubin (eds.) *Learner Strategies in Language Learning* 57–69. Englewood Cliffs: Prentice Hall.
- Cohen, A. & Cavalcanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (ed.) *Second Language Writing* 155–177. Cambridge: Cambridge University Press.
- Conrad, C. (2001). Portfolio assessment: The influences of the assessment criteria and rating process on holistic scores. Retrieved 16 September 2011 from http://www.carla.umn.edu/resources/working-papers/documents/WP20_WritingPortfolio.pdf.
- Evans, N., Hartshorn, J., McCollum, R. & Wolfersberger, M. (2010). Contextualising corrective feedback in second language writing pedagogy. *Language Teaching Research* 14, 445–463. <http://dx.doi.org/10.1177/1362168810375367>.
- Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing* 10, 235-249. [http://dx.doi.org/10.1016/S1060-3743\(01\)00042-X](http://dx.doi.org/10.1016/S1060-3743(01)00042-X).
- Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly* 29, 33–53. <http://www.jstor.org/stable/3587804>.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: The University of Michigan Press.
- Ferris, D. (2003). *Response to Student Writing: Implications for Second Language Students*.

Mahwah: Lawrence Erlbaum Associates.

Ferris, D & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, Process, Practice*.

Mahwah: Lawrence Erlbaum Associates.

Ferris, D. & Hedgcock, J. (2014). *Teaching L2 composition: Purpose, process and practice (3rd Edition.)*. Oxon: Routledge.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences*, 1(1), 27-38.

Flower, L. & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication* 32, 365–387. <http://www.jstor.org/stable/356600>.

Freedman, S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment* 1, 27–52.

http://dx.doi.org/10.1207/s15326977ea0101_3

Gearhart, M. & Wolf, S. (1997). Issues in portfolio assessment: Assessing writing processes from their products. *Educational Assessment* 4, 265–296.

http://dx.doi.org/10.1207/s15326977ea0404_1.

Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher and student variables. In K. Hyland and F. Hyland (eds.) *Feedback in Second Language Writing: Contexts and issues* 185–205. New York: Cambridge University Press.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College composition and communication* 33, 76–88.

<http://www.jstor.org/stable/357846>.

Hamp-Lyons, L. & Condon, W. (1993). Questioning assumptions about portfolio based

- assessment. *College composition and communication* 44, 176–190.
- <http://www.jstor.org/stable/358837>.
- Hamp-Lyons, L. & Condon, W. (2000). *Assessing the Portfolio: Principles for Practice, Theory and Research*. Cresskill, New Jersey: Hampton Press.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- 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, 33 – 54. <http://dx.doi.org/10.1177/136216880000400103>.
- Hyland, F. & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing* 10, 185 – 212. [http://dx.doi.org/10.1016/S1060-3743\(01\)00038-8](http://dx.doi.org/10.1016/S1060-3743(01)00038-8).
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System* 25, 465–477. [http://dx.doi.org/10.1016/S0346-251X\(97\)00045-6](http://dx.doi.org/10.1016/S0346-251X(97)00045-6).
- Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (ed.) *Second Language Writing: Research Insights for the Classroom* 57–68. Cambridge: Cambridge University 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, 30–43. <http://dx.doi.org/10.1016/j.jslw.2008.06.002>.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. Johnson, and D. Roen (eds.) *Richness in Writing: Empowering ESL Students* 207–219. New York: Longman.
- Nystrand, M., Cohen, A. & Dowling, N. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment* 1, 53–70.

http://dx.doi.org/10.1207/s15326977ea0101_4.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal* 59, 23–30.

<http://dx.doi.org/10.1093/elt/cci003>.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives.

Annual Review of Cognitive Linguistics, 7(1), 141-163.

Romova, Z. & Andrew, M. (2011). Teaching and assessing academic writing via the portfolio:

Benefits for learners of English as an additional language. *Assessing Writing* 16, 111–122.

<http://dx.doi.org/10.1016/j.asw.2011.02.005>.

Ruegg, R. (2014). The effect of peer and teacher feedback on changes in EFL students’

writing self-efficacy. *The Language Learning Journal*:

DOI:10.1080/09571736.2014.958190

Ruegg, R., Fritz, E. & Holland, J. (2011). Rater sensitivity to qualities of lexis in writing.

TESOL Quarterly 45(1), 63-80.

Ruegg, R. & Koyama, D. (2011). Confidence in writing: The effect of feedback. *Studies in*

Linguistics and Language Teaching 21, 137–166. Retrieved 31 March 2011 from

http://ci.nii.ac.jp/els/110008439899.pdf?id=ART0009681216&type=pdf&lang=en&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1382590439&cp=.

Straub, R. (1997). Students’ reactions to teacher comments: An exploratory study. *Research in*

the teaching of English 31, 91–119. <http://www.jstor.org/stable/40171265>.

Tsui, A. & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of*

Second Language Writing 9, 147 – 171. [http://dx.doi.org/10.1016/S1060-3743\(00\)00022-9](http://dx.doi.org/10.1016/S1060-3743(00)00022-9).

Xiaoxiao, L. & Yan, L. (2010). A case study of dynamic assessment in EFL process writing.

Chinese Journal of Applied Linguistics 33, 24–40. Retrieved 16 September 2011 from

<http://www.celea.org.cn/teic/89/10042202.pdf>.

Zacharias, N. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal* 38, 38–52. <http://dx.doi.org/10.1177/0033688206076157>.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly* 19, 79–101.
<http://www.jstor.org/stable/3586773>.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing* 4, 209–222. [http://dx.doi.org/10.1016/1060-3743\(95\)90010-1](http://dx.doi.org/10.1016/1060-3743(95)90010-1).

Appendix 7.A

Questions Concerning Students' Perceptions of the Feedback They Received (Teacher Feedback Group)

5. How was the amount of teacher feedback you received?

- A: Very insufficient
- B: A little insufficient
- C: Just right
- D: A little too much
- E: Far too much

6. How was the amount of peer feedback you did?

- A: Very insufficient
- B: A little insufficient
- C: Just right
- D: A little too much
- E: Far too much

7. How helpful was teacher feedback for improving your grammar?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

8. How helpful was teacher feedback for improving your organization of ideas in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

9. How helpful was teacher feedback for improving the support you gave for each main idea in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

10. Overall, how helpful was teacher feedback?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

11. How much of the teacher feedback did you understand?

- A: None
 - B: A little
 - C: Some
 - D: Most
 - E: All
-

Appendix 7.B

Questions Concerning Students' Perceptions of the Feedback They Received (Peer Feedback Group)

5. How was the amount of peer feedback you did?

- A: Very insufficient
- B: A little insufficient
- C: Just right
- D: A little too much
- E: Far too much

6. How was the amount of teacher feedback you received?

- A: Very insufficient
- B: A little insufficient
- C: Just right
- D: A little too much
- E: Far too much

7. How helpful was peer feedback for improving your grammar?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

8. How helpful was peer feedback for improving your organization of ideas in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

9. How helpful was peer feedback for improving the support you gave for each main idea in your essays?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

10. Overall, how helpful was peer feedback?

- A: Not helpful at all
- B: A little helpful
- C: Very helpful

11. How much of the peer feedback did you understand?

- A: None
 - B: A little
 - C: Some
 - D: Most
 - E: All
-

CHAPTER EIGHT

Overall conclusion

In this chapter, the findings of each of the six studies will be summarized. An overview of the contribution made by the dissertation will then be given. Finally, implications the results of the research have for practitioners in the field will be discussed, followed by limitations of the research and avenues for further investigation.

8.1 Summary of findings

8.1.1 Source of feedback

The findings of Study One suggest that teacher feedback offers an advantage over peer feedback in terms of the development of grammatical performance when writing in English. It is clear from Study One that this improved grammatical performance in writing did not result from the teacher giving more feedback on surface-level issues than peers, because it was found that there was no significant difference between the proportion of surface-level grammatical feedback given by the teacher and the peers. However, the teacher did give significantly more meaning-level feedback and it may have been the higher proportion of meaning-level feedback that promoted improved grammatical performance in writing. It may be difficult for speakers of the same language to give meaning-level feedback to each other, as sharing the same first language means that they are likely to understand each other to a greater extent, sharing the same L1 interference.

The findings of Study Two suggest that Japanese learners pay more attention to teacher feedback than peer feedback, even when they receive feedback from the same source longitudinally. It appears that learners believe they should pay attention to teacher feedback,

even when they do not understand that feedback. On the other hand, the learners were equally reluctant to accept peer feedback and use it in their revisions unless they knew or could easily ascertain that it was good advice. This reluctance to critically engage with peer feedback supports the findings of Lundstrom and Baker (2009), that giving peer feedback is more beneficial than receiving it in order to develop critical thinking skills.

The results of Study Three suggest that teacher feedback has a positive effect on learner self-efficacy when compared to peer feedback. Furthermore, while the amount of constructive feedback and praise may play a role in determining changes in learner self-efficacy, the results of this research suggest that the source of feedback has a stronger influence on learner self-efficacy than the relative amounts of constructive feedback and praise.

Overall, the results of these three studies suggest that Japanese learners pay more attention to teacher feedback than to peer feedback, even when receiving feedback from the same source longitudinally. In addition to this, teacher feedback also helps students in terms of improved grammatical performance in their writing and increased learner self-efficacy.

8.1.2 Assessment of feedback processes

The results of Study Four demonstrate that the assessment of the feedback given by peers results in better peer feedback in terms of both quantity and quality. The students in the feedback-assessed peer feedback group gave more individual feedback points, wrote more and longer comments and gave a significantly higher proportion of specific comments than the only product-assessed group. Furthermore, while the difference in the accuracy of the feedback was not statistically significant, in the feedback-assessed group accuracy increased over the one year period while in the only product-assessed group it did not. The most

important effect of assessing the peer feedback was that every draft in the feedback-assessed group received constructive feedback, while 12.5% of drafts in the only product-assessed group received no feedback at all. If learners receive no constructive feedback on a draft, they will probably be left unclear as to what revisions to make and may therefore not make any revisions at all. This would seem to defeat the purpose of employing the process approach.

The results of Study Five indicate that the assessment of the use of teacher feedback does push learners to modify their output, as the process-assessed group made significantly more attempts at uptake than the only product-assessed group, while the number of successful revisions was the same between the two groups. The results of Study Five also show that the only product-assessed group improved in writing performance significantly more than the process-assessed group.

There was also no significant difference found in the learners' perceptions of the feedback they had received, regardless of the feedback treatment they were involved in. However, it is quite probable that the number of participants was too small to detect any significant difference between the groups.

Overall, the results of these three studies show that while the assessment of peer feedback did appear to help learners, in the sense that they received more and more specific feedback than those assessed only on product, the assessment of the use of teacher feedback did not appear to promote learning. However, the learners in this study were majoring in English and had a high level of intrinsic motivation. Learners with a lower level of intrinsic motivation are more likely to benefit from such assessment of process. It is therefore recommended that similar research be conducted with learners who are taking compulsory English language classes but majoring in a different subject, as this type of students are likely

to have less intrinsic motivation and may be affected more by the assessment of writing process.

8.2 Contribution of the research

This section will revisit the main gaps in the literature which prompted the research reported in this dissertation. Following this, it will explain the extent to which the research reported in this dissertation has filled these gaps and thus contributed to the field of L2 writing.

8.2.1 Source of feedback

As reported in the introductory chapter of this dissertation, while a number of studies have compared peer and teacher feedback (Caulk, 1994; Hedgcock & Lefkowitz, 1992; Nakanishi & Akahori, 2005 cited in Hirose, 2008; Yang, Badger & Yu, 2006) there appears to have been only one previous study comparing the effect of peer and teacher feedback longitudinally (Fox, 1980). Fox (1980) compared a teacher feedback group and a peer feedback group at the end of one semester of having received feedback from only one source and found no significant difference in the writing performance of the two groups. However, since Fox (1980) did not conduct a pre-test at the beginning of the semester, it is impossible to determine whether the groups' writing performances would have differed at the beginning of the semester. Because we do not know this, we also have no way of knowing whether the gains experienced by both groups were indeed the same.

In chapter two of this dissertation, a peer feedback group and a teacher feedback group were compared longitudinally using a pre-test, post-test research design. The results of the study are more robust than the results of Fox (1980) because a pre-test and a post-test allowed gains to be measured and compared between the two groups rather than simply comparing writing performance at the end of the treatment period. Accordingly, the findings of the pre-

post study presented in Chapter 2 can contribute to our knowledge of “source of feedback”. When Japanese university students receive regular feedback over a long period, teacher feedback can help them to improve their grammatical performance in writing. On the other hand, in terms of improvement in content, vocabulary and organization skills, either peer feedback or teacher feedback can be equally effective.

8.2.2 Focus of feedback

A majority of previous studies on feedback in L2 writing have focused on feedback on grammar as a form of implicit grammar instruction. However, in the writing classroom it is clear that learners need to learn more than just grammar and therefore the feedback given by writing instructors needs to cover a range of issues. Indeed, in a meta-analysis of research, Biber, Nekrasova and Horn (2011) found that a combined focus on both language form and content led to greater gains in grammar than either a focus on form or focus on content alone. This indicates that even if a learner’s goal is limited to learning the English language, giving feedback on both grammar and content would still be the most effective approach.

Study one verified the finding of the meta-analysis by Biber, Nakrasova and Horn (2011) in that the teacher gave significantly more feedback on meaning-level issues and content yet the learners in the teacher feedback group significantly outperformed those in the peer feedback group in terms of grammatical performance in writing. This happened despite the fact that the proportion of feedback focused on surface-level grammatical errors was not significantly different between the two groups. As such, instructors should pay attention to the findings of the meta-analysis and always offer feedback on grammar in conjunction with feedback on content.

Moreover, a number of writers have recommended that teachers tailor their feedback to each learner's needs (Bitchener & Knoch, 2010b; Ferris, 2002; 2003; Ferris & Hedgcock, 2005; Goldstein, 2004; Hyland, F. 1996, Storch, 2010). On the other hand, while there are a number of ways one could tailor feedback to learners' needs, there is little published about how one would go about doing this. Since the practice has been recommended in the literature, it is recommended that future research investigates and shares various different ways of tailoring feedback to the needs of individuals in the classroom.

This dissertation offers one way of tailoring feedback to individual learners' needs which is efficient and effective even for large classrooms. In the research described in this dissertation, the instructor used a blank feedback checklist which students populated with their own questions. The teacher or peer reviewer then offered feedback based on the questions posed by the writer. Having students write specific questions for feedback does not take up much classroom time, and once students understand the process and have experienced it once or twice, they can prepare the feedback checklist outside of the classroom. Furthermore, providing feedback on questions posed by the writers does not take any additional time than providing feedback using a generic feedback form. Therefore, this method may also be effective for other instructors who would like to tailor their feedback to individual learners but do not have any efficient method of doing so.

8.2.3 Length of feedback treatment

Although a large number of previous studies have demonstrated that teacher feedback is effective for improving grammatical performance in writing, a vast majority of these studies have been one shot studies involving students receiving feedback just once and then being tested in terms of the grammatical form/s they received feedback on (Bitchener, 2008;

Bitchener & Knoch, 2008; 2009; 2010a; 2010b; Fathman & Whalley, 1990; Sheen, 2007; Sheen, Wright & Moldowa, 2009; Truscott & Hsu, 2008; Van Beuningen, De Jong & Kuiken, 2012). However, as stated by Xu (2009), the feedback given on particular form/s in this kind of research may communicate to learners to pay attention to those particular form/s. This heightened attention to the form/s may be enough to increase performance in terms of those particular form/s. This is particularly true if the researcher conducts the research in the classroom him or herself, because seeing the same person in the classroom may remind learners of the particular form/s they received feedback on. Furthermore, since language development happens gradually over a long period of time, it appears that repeated feedback over a long period of time would be more likely to lead to improved performance because it acts as a form of spaced repetition and provides opportunities for hypothesis testing in between.

However, "...longitudinal research on student improvement as a result of teacher feedback has been virtually nonexistent" (Ferris & Hedgcock, 2005: 187) and recently several writers have stressed the need for more longitudinal research (Hyland, K. 2010; Van Beuningen, 2010; Storch, 2010). The studies reported in this dissertation contribute to the field in terms of having added to this virtually nonexistent pool of previous longitudinal research on feedback in writing. From the various findings of the studies reported in this dissertation, we have a better understanding of the effects of feedback over the long term. Truscott (2007: 349) stated that teacher feedback is unlikely to have any effect on students' grammatical performance in writing in the long term. However, this idea had previously not been tested through research. The findings of the research reported in this dissertation appear to refute Truscott's assertion.

8.2.4 Uptake of feedback

While a number of studies have compared uptake of peer and teacher feedback after learners have received both peer and teacher feedback on the same draft (Connor & Asenavage, 1994; Paulus, 1999; Tsui & Ng, 2000), the results of these studies were not surprising since it is natural for a learner to pay more attention to their teacher than a peer when faced with feedback from both sources on a single draft. However, there does not appear to have been any previous research investigating the uptake of peer and teacher feedback when learners have received feedback from only one source longitudinally. The results of such longitudinal research were expected to be different from previous studies, since learners were deprived of feedback from other sources.

Chapter three of this dissertation illustrated that learners do pay more attention to teacher feedback than peer feedback. Moreover, they appear to do so whether they understand the feedback or not. This can be seen by the fact that a significantly higher proportion of the revisions undertaken by students after receiving teacher feedback were misunderstood or unsuccessful revisions, whereas a significantly higher proportion revisions undertaken by students after receiving peer feedback were successful. On the one hand, feedback which is misunderstood or leads to an unsuccessful revision may be considered by many to have been ineffective. On the other hand, feedback which leads to successful revisions may have only pointed out mistakes that students could have noticed themselves (as mentioned by Sengupta, 1998), rather than errors which indicate the current state of their interlanguage. Swain (1995) has claimed that pushed output is necessary for language acquisition and it is unclear whether the former type of feedback actually leads to pushed output. This suggests that teacher feedback may be more beneficial than peer feedback because it represents a stronger push to improve their subsequent output. As far as I am aware, no-one has compared the uptake of

peer and teacher feedback longitudinally before and therefore this study has raised some interesting avenues of future investigation.

8.2.5 Confidence

Prior to the research reported in this dissertation, many scholars in the field of L2 writing had observed that increased teacher feedback leads to decreased learner confidence (Cleary, 1990; Cohen, 1987; Ferris, 2002; Krashen, 1982; Semke, 1982; Truscott, 1996) or that too much teacher feedback leads to students feeling less confident (Andrade & Evans, 2013; Bitchener & Ferris, 2012; Cleary, 1990; Ferris, 2002; Hyland, F. 1998, Hyland & Hyland, 2001; Krashen, 1982; Semke, 1982), while others have suggested that peer feedback has a more positive influence on learner confidence than teacher feedback (Chaudron, 1984; Ferris, 2003; Ferris & Hedgcock, 2005; Gungle & Taylor, 1989; Leki, 1990; Martinez, Kock & Cass, 2011; Mittan, 1989; Tsui & Ng, 2000; Zhang, 1995). However, this has been merely suggested by the authors and these views had not been corroborated with empirical evidence.

The research reported in chapter four of this dissertation appears to have been the first empirical research investigating this issue. The findings suggest that these views may not be well founded in all contexts. Although students may feel discouraged upon receiving feedback from a teacher, it appears that over the long term teacher feedback may increase writers' confidence in their writing ability. Likewise, many social cognitive theorists have argued that student self-efficacy beliefs develop through increased exposure to feedback (e.g. Graham, MacArthur, Schwartz & Page-Voth, 1992; Pajares, 2003; Schunk & Swartz, 1993). This appears to be true even when teachers give very little praise and give significantly more constructive feedback than peers do.

8.2.6 Assessment of feedback process

A great deal of the previous literature on classroom assessment of writing within the process approach describes ways of assessing the final drafts of essays (e.g. Archibald, 2011; Ferris & Hedgcock, 2005). However, it is extremely difficult to find previous literature discussing ways of assessing the feedback processes used by learners during the writing process. This is despite the fact that many previously published books do recommend that processes in addition to product should be assessed and despite the fact that many instructors do assess feedback processes in addition to final written product.

The research reported in this dissertation investigated the assessment of process in addition to product in three different ways: the assessment of improvement between drafts after learners had received teacher feedback, the assessment of quality and quantity of peer feedback provided by peers and learners' perceptions of the feedback they had received after having received teacher feedback (process-assessed or only product-assessed) and after having received peer feedback (process-assessed or only product-assessed). By investigating these three areas, the research intended to open the door to further research in these three separate avenues of investigation. The L2 writing field has established that the process approach is beneficial in the acquisition of writing skills and strategies (e.g. Ferris, 2003; Hyland & Hyland, 2006; Liu & Hansen, 2002). However, the ways in which these skills and strategies should be assessed is still far from sure.

The findings of the research in this dissertation suggest that improvement between drafts should not be assessed, rather learners should be offered teacher feedback and then left to decide what they would like to do with it. The research reported in chapter five of this dissertation found that the assessment of peer feedback did lead to more feedback and better

quality feedback being provided by peers. However, the students in the group whose peer feedback was assessed still did not perform significantly better than those in the group whose final written product alone was assessed. It therefore remains unclear whether peer feedback is pedagogically effective or whether its merits lie in the social aspect alone.

Overall, the research reported in this dissertation has made several contributions to the field of L2 writing. The findings show that systematic teacher feedback, provided longitudinally can have positive effects in terms of grammatical performance in subsequent writing and learner self-efficacy. In addition, teacher feedback, when offered repeatedly may be superior to peer feedback in terms of pushing students to test various hypotheses and thus contribute to their knowledge of how the language works. On the other hand, while peer feedback has been shown to be less effective for subsequent grammatical performance, it was shown to be as effective as teacher feedback for issues related to organization of ideas and academic style. Moreover, it is more beneficial from the social aspect. Research reported in this dissertation has found that, while the assessment of the use teacher feedback does not improve performance for students with high motivation, the assessment of peer feedback is effective in increasing both the quality and the quantity of feedback provided by peers during peer review.

8.3 Implications

Clearly, there is a need for continued discussion of ways to increase student acceptance of peers as a legitimate source of feedback. Certainly, peer feedback should not be abandoned as it provides many other benefits such as increased interaction, improved classroom culture and critical thinking skills (in the giving, even if not in the receiving). However, ultimately if

learners are reluctant to accept advice from their peers it is apparent that the teacher is a necessary source of feedback.

The fact that learners were selective in their use of peer feedback could be said to be an advantage of peer feedback in that it encourages critical evaluation of a learner's own writing. However, the fact that learners in the teacher feedback group used more of the feedback in their revisions gives evidence of hypothesis testing, which is a form of 'pushed output' and may result in more language acquisition. This conclusion is supported by the fact that the teacher feedback group improved in their grammatical performance significantly more than the peer feedback group. It may have been the higher level of hypothesis testing that led to gains in grammatical performance in writing, despite the fact that many of the revision attempts were misunderstandings of the feedback or unsuccessful revisions.

As with the teacher as a source of feedback, when students were assessed on their use of teacher feedback in addition to the final written product they made more uptake attempts than when they were assessed on the final written product alone. The increased hypothesis testing involved in these uptake attempts may increase acquisition. On the other hand, the uptake attempts represent an increase in unsuccessful, rather than successful, revision attempts, which may be an indicator that increased hypothesis testing through pushed output does not increase language acquisition. However, we know that interlanguage development is a long and slow process (Gass, 2003) and it may be that only one year is not long enough to see such change or that receiving feedback only 16 times does not allow enough hypothesis testing for such change to manifest.

It has been suggested that feedback should be tailored to the needs of each individual learner. While some have suggested this only for feedback on form, as shown by Hyland

(2000), the extent to which students actually want feedback on form also varies a great deal. The blank feedback checklist utilized in this research is one possible method of tailoring feedback to each learner that does not require a great deal of extra effort on the part of the instructor. This method not only allows learners to choose what kind of feedback to receive on each draft, but also fosters the development of learner autonomy as it encourages learners to reflect on their strengths and weaknesses in writing, and (at least in the case of teacher feedback) receive confirmation of those strengths and weaknesses. As they reflect on each draft they have written, they have the benefit of previous feedback to inform their choice of questions for the subsequent feedback. This kind of repeated reflection and developing greater understanding of one's strengths and weaknesses is one of the cornerstones of the development of learner autonomy.

Ultimately, until more longitudinal research is conducted, it is suggested that the teacher should provide feedback on meaning-level issues and content, while peers provide feedback on organization and academic style.

8.4 Limitations

As with any research, the results of this research are bound by the context in which it was conducted. Firstly, it was conducted on intermediate proficiency learners and secondly, on Japanese learners. Intermediate proficiency learners are the least likely to benefit from peer feedback because their limited language skills prevent them from entering into deep discussions about their writing and that of their peers. Furthermore, due to the fact that the learners are of intermediate English proficiency, the feedback they give may not be as accurate and they may not be able to give as much feedback as higher proficiency learners.

Japanese learners may benefit more from teacher feedback and less from peer feedback because of the hierarchical nature of the society. The strong hierarchical nature of the society may have influenced the results of the first three studies, on the issue of source of feedback as well as Study Five. Assessment of the use of teacher feedback may have made little difference because the learners believe strongly that they should use the feedback provided by the teacher. On the other hand, many of the claims about Japanese learners are outdated and certainly many young Japanese students these days do not hold the same beliefs as the older generation.

It is usually considered ideal to have at least 30 students in any one group when groups of students are compared statistically. In the studies reported in this dissertation, the number of students in each of the different groups ranged from just 13 to 37 and only 3 of the 14 groups contained 30 or more students. However, there were 106 students in the four classes in which the research was conducted. The fact that 25 students did not give consent to participate in the study and attrition both contributed to the small number of students ultimately included in each group. However, it is also difficult to avoid these problems in longitudinal research conducted in a naturalistic setting. In the case of this research, increasing the sample size would have involved including groups of students with other instructors and this in turn would add the instructor as a moderating variable.

Another limitation of this research was the use of scores assigned by human raters using analytic rating scales as the only measure of writing performance. Although the scores were scaled using many Facet Rasch modelling, which adjusts the scores to ameliorate the relative strictness or leniency of each rater on each rating scale, human raters may not be sensitive to subtle differences between essays and the measurement tool may therefore not be sharp enough to find such subtle differences. On the other hand, as stated by Truscott (2007),

it is important for pre-tests and post-tests to be conducted outside of the context of the research. Otherwise, students may attempt to reflect what they see as the important points within that context, rather than simply trying to write the best essay they can (Xu, 2009). These competing problems; the advantage of using a test which is external to the research context and the necessity of using a sharp measurement tool offer a conundrum to researchers as tests which are used within the educational context in question may not represent the sharpest measurement tools available. Furthermore, using an institutional test which the students are required to take by the university also precludes the possibility of conducting a delayed post-test.

Other institutional factors also limited the way in which the research was conducted to some extent. For example, the curriculum at the university is fixed to some degree, stating which kinds of writing assignments should be completed within each semester and how much writing is appropriate for learners to produce within each semester.

8.5 Suggestions for further research

More research needs to be conducted in authentic classroom settings and over longer periods of time in order to determine the effect of our usual teaching practices. Research on feedback in writing should not only consider outcomes in terms of uptake of feedback or improvement in writing performance, but also look at aspects of the feedback itself alongside those outcomes, to shed more light on the nature of and the reasons behind the outcomes.

A follow-up study, comparing peer and teacher feedback longitudinally in an ESL context in which the learners have different first languages would help to clarify whether these learners were hampered by coming from the same linguistic background. It is possible that

systematic peer feedback provided by ESL learners in multilingual classrooms over a whole year could result in gains similar to those found in the teacher feedback group in this study.

While the results of this research refute the claim that peer feedback has a more positive effect on learner confidence than teacher feedback, the effects of the amounts of constructive feedback and praise on learner confidence still have not been corroborated in the field of L2 writing and warrant further research. In addition to this, the issues of how much constructive feedback is too much and how much praise is too much have also not been investigated comprehensively. This would be another worthwhile avenue for future research.

Process writing has been widely adopted around the world and the assessment of writing is a thriving research field. However, I had difficulty finding any research at all on the assessment of writing processes. If we are to place importance on the writing process and not only the product, it is essential that we demonstrate this to learners, not only by saying that process is important or asking them to actively engage in the writing process, but also through our assessment methods, which are arguably one of the most powerful messages we send to our learners. Furthermore, a number of writers have already suggested this (Ferris, 2002; Ferris, 2003; Ferris & Hedgcock, 2005; Goldstein, 2006; Hyland, 1996; Leki, 1990; Mittan, 1989), yet it is difficult to find any information about how we should go about doing this. Because our assessment methods send powerful messages to our learners, it is time that our methods of assessment in the writing classroom (classroom assessment, rather than testing) are taken up with more fervor as an area of investigation.

References

Andrade, M., & Evans, N. (2013). *Principles and practices for response in second language*

- writing: *Developing self-regulated learners*. New York: Routledge.
- Archibald, A. (2011). Writing in a second language. Retrieved 16th September 2011 from the Centre for Languages, Linguistics and Area Studies website:
<http://www.llas.ac.uk/resources/gpg/2175>
- Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *TOEFL iBT Research Report 14*.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102-118.
- Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal, 12*, 409-431.
- Bitchener, J., & Knoch, U. (2009). The value of a focussed approach to written corrective feedback. *ELT Journal, 63*, 204-211.
- Bitchener, J., & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*, 207-217.
- Bitchener, J., & Knoch, U. (2010b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 193-214.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly, 28*, 181-188.
- Chaudron, C. (1984). Evaluating Writing: Effects of feedback on revision. *RELC Journal, 15*, 1-14.
- Cleary, L. (1990). The fragile inclination to write: Praise and criticism in the classroom.

English Journal, 79, 22-28.

- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: The University of Michigan Press.
- Ferris, D. (2003). *Response to Student Writing*. Mahwah NJ: Lawrence Erlbaum Associates.
- Ferris, D., & Hedgcock, J. (2005). *Teaching ESL composition*. Mahwah: Lawrence Erlbaum Associates.
- Fox, R. (1980). Treatment of writing apprehension and its effects on composition. *Research in the teaching of English*, 14, 39-49.
- Gass, S. (2003). Input and interaction. In C. Doughty & M. Long (Eds.). *The Handbook of Second Language Acquisition* (pp. 224-255). Oxford: Blackwell.
- Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher and student variables. In K. Hyland & F. Hyland (Eds.). *Feedback in Second Language Writing: Contexts and issues* (pp 185-205). New York: Cambridge University Press.
- Graham, S., MacArthur, C., Schwartz, S., & Page-Voth, V. (1992). Improving the compositions of students with learning disabilities using a strategy involving product and process goal setting. *Exceptional Children*, 58(4), 322-335.
- Gungle, B., & Taylor, V. (1989). Writing apprehension and second language writers. In D. Johnson and D. Roen (Eds.). *Richness in writing: Empowering ESL students* (pp. 235-148). New York:

Longman.

- Hedgcock, J., & Lefkowitz, N. (1992). Collective oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing* 1, 255-276.
- Hirose, K. (2008). Peer Feedback in L2 English Writing Instruction. In K. Bradford-Watts, T. Muller and M. Swanson (Eds.). *JALT2007 Conference Proceedings* (pp. 543-552). Tokyo: JALT.
- Hyland, K. (1996). *Second language writing*. New York: Cambridge University Press.
- Hyland, F. (1998). The impact of Teacher Written Feedback on Individual Learners. *Journal of Second Language Writing*, 7, 255-286.
- Hyland, F. (2000). ESL writers and feedback: giving more autonomy to students. *Language Teaching Research* 4(1), 33-54.
- Hyland, F. (2010). Future directions in feedback on second language writing: Overview and research agenda. *International Journal of English Studies*, 10, 171-182.
- Hyland, F., & Hyland, K. (2001) Sugaring the Pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10, 185-212.
- Hyland, K. & Hyland, F. (2006). *Feedback in Second Language Writing: Contexts and issues*. New York: Cambridge.
- Krashen, S. (1982). *Principles and practice in Second Language Acquisition*. Oxford: Pergamon Press.
- 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*. In D. Belcher & J. Liu (Eds.) Michigan Series on Teaching Multicultural Writers. 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, 30-43.
- Martinez, C., Kock, N., & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students' writing anxiety and writing self-efficacy. *Journal of Adolescent and Adult Literacy*, 54, 351-360.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. Johnson and D. Roen (Eds.). *Richness in writing: Empowering ESL students* (pp 207-219). New York: Longman.
- Pajares, F. (2003). Self-efficacy beliefs, motivation and achievement in writing: A review of the literature. *Reading and Writing Quarterly*, 19, 139-158.
- Paulus, T. (1999). The Effect of Peer and Teacher Feedback on Student Writing. *Journal of Second Language Writing*, 8, 265-289.
- Semke, H. (1982). Effects of the Red Pen. *Foreign Language Annals*, 17, 195-202.
- Sengupta, S. (1998). Peer evaluation: 'I am not the teacher'. *ELT Journal*, 52(1), 19-26.
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progressive feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18, 337-354.
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10, 29-46.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.) *Principles and practices in the study of language* (pp. 125-144).
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning* 46, 327-369.

- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255-272.
- Truscott, J., & Hsu, A. (2008). Error correction, revision and learning. *Journal of Second Language Writing, 17*, 292-305.
- Tsui, A., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*, 147-171.
- Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights and future directions. *International Journal of English Studies, 10*, 1-27.
- Van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*, 1-41.
- Xu, C. (2009). Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing, 18*, 270-275.
- Yang, M., Badger, R. & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179-200.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 4*, 209-222.

Appendix 8.A

Dear Rachael,

This letter is to confirm that the following ethics application/s cited below received final approval from the Macquarie University Human Research Ethics Committee:

Chief Investigator: Associate Professor Mehdi Riaz

Ref: 5201000118

Date Approved: 25 March 2010

Title: "Feedback in EFL writing"

Please do not hesitate to contact me if you have any questions.

Yours sincerely,

Dr Karolyn White

Director, Research Ethics & Integrity