Peer Feedback in the Academic English Classroom: A Pilot Study

David ALLEN & Amy MILLS

Abstract

Past research has shown that peer feedback in a second language (L2) context is beneficial to students. However, there is relatively information about the type of feedback that learners give and to what degree this is incorporated as revisions. The present study focuses primarily on two aspects of the revision process: the suggested revisions made by peers, and the revisions made by the writer after receiving comments from their peer. We examine feedback in terms of type (surface vs. meaning), level (word, sentence, paragraph), and whether the feedback concerns general and register-related issues. An analysis was conducted of research articles written by first-year Japanese university students. The suggestions made tended to focus on non-meaning-related issues at the word and sentence levels. The revisions made following peer feedback contained a greater proportion of those related to meaning, but also were mainly at the word and sentence levels. Both suggestions and revisions sufficiently addressed register-related issues. Finally, revised texts were significantly more complex than initial drafts, showing that through the process of peer feedback learners can and do improve the complexity of their writing.

Introduction

A fundamental part of the writing process is the revision of texts. When revising a text, academic writers often gain feedback from others in order to improve the quality of the text; for instance, doctoral students gain feedback from their professors and aca-

demics receive feedback from their peers. *Peer feedback* (also referred to as *peer response* or *peer review*) is also used in the second language (L2) classroom, where students learn writing skills through interaction with other students (their peers) as part of the writing process (see Liu & Hansen, 2002, for an overview).

A considerable amount of research has been conducted on peer feedback in the L2 context. L2 writing studies have investigated a number of important questions, such as whether peer feedback leads to more and better revisions than other forms of feedback such as teacher feedback and self-feedback or *self-revision* (Chaudron, 1983; Connor & Asenavage, 1994; Paulus, 1999; Suzuki, 2008) and whether training and appropriate structuring of the peer feedback task can improve the amount and quality of feedback provided by peers (Berg, 1999; Min, 2005, 2006; McGroarty & Zhu, 1997; Zhu, 1995). Overall, while it is still undetermined if peer feedback generally results in more revisions than other forms of feedback, training has generally been shown to improve the number and quality of revisions.

A recent study by Lundstrom & Baker (2009) investigated whether peer feedback leads to gains in L2 writing ability for the giver and receiver of feedback. Peers were divided into two groups: one that only gave feedback and another that only received feedback on their writing. Over a semester course, they found that those giving feedback increased their writing ability significantly more than those receiving feedback as measured by pre- and post-tests. This indicates that through the process of giving feedback learners notice and acquire writing skills. In sum, this study suggests that peer feedback is beneficial, but potentially more so for the giver than the receiver.

Other research has shown that peer feedback also provides an opportunity for developing other communicative and cognitive skills such as negotiation of meaning and meta-cognitive strategies (Suzuki, 2008; Min, 2005; Villamil & De Guerrero, 1996). Moreover, through the process of peer feedback, learners are allowed the opportunity to improve their social skills and experience the collaborative nature of writing from a peer perspective (as opposed to taking 'expert' advice from the teacher or supervisor). Overall, while there are uncertainties regarding the effectiveness of peer feedback as opposed to other kinds of feedback, with appropriate training, peer feedback is likely to lead to

improved language ability, text quality, development of social, cognitive and meta-cognitive skills, and an improved understanding of the writing process.

Although research points to the benefits of utilising peer feedback in the language classroom, there is still a lot of work to be done in determining how students in a foreign language context give and receive feedback. In the present research we focus primarily on two aspects of the revision process: the *suggested revisions* made by peers, that, is the comments and suggestions made by the peer when reviewing a peers' paper; and the *revisions* made by the writer after receiving comments from their peer. By virtue of analysing both suggested revisions and the actual revisions we can also investigate the number and type of suggested revisions that are acted upon by the writer, in other words, the number of *incorporated suggestions*.

A distinction can be made between surface-level revisions, which do not alter the meaning of the text (such as revising subject-verb agreement) and meaning-related revisions, which modify the meaning of the text (such as adding new information to a text). Based on previous research, it is still unclear whether peer feedback tends to lead to more surface-level revisions or meaning-related revisions. In one study, Paulus (1999) found that peer and teacher feedback leads to increased meaning-related revisions when compared to self-revisions, but overall, revisions tended to be at the surface level. Other studies have also suggested that reviewers in the L2 context tend to focus on surface-level errors (McGroarty & Zhu, 2002) but with training may improve their ability to make meaning-related suggested revisions (Berg, 1999). More weight has often been given to meaning-related revisions as 'better' revisions (Zamel, 1985), though it is also true that poor grammar and mechanics reflect badly on the writer and can lead to severe criticism from readers/assessors (especially if the target audience is an academic one; Hyland & Hyland, 2006).

Another distinction has been made between revisions that are at the paragraph, sentence and word levels (e.g., Suzuki, 2008). These different types of revisions reflect the writers' engagement with the text at multiple levels: the discourse level (i.e., rhetorical structure, inter- and intra-paragraph cohesion and coherence), the sentence level (i.e., modifications of syntax,

phrase usage, agreement), and the word level (i.e., word choice, spelling). The revisions that may be more highly regarded are those at the discourse (paragraph) level, as these usually require a more critical understanding of the coherence of an argument across paragraphs. However, writers often need to revise at all of these levels in order to improve text quality; this is particularly important when writing in a second language, where grammatical and word choice errors are more prominent than when writing in the first language. Suzuki (2008) found that revisions following peer feedback and self revision both tended to be those related to word- and sentence-level revisions, while few revisions were made at the paragraph level. However, the distinction between word, sentence and paragraph levels does not discriminate between meaning-related and non-meaning-related revisions; therefore it is difficult to compare Suzuki's results with previous research such as Paulus (1999).

In the present study, we investigate both the type (surface vs. meaning) and the level (word, sentence, paragraph) of revision that learners make. Moreover, we focus not only on the revisions but also the suggestions that peer reviewers make during the feedback sessions, because the latter have received little attention in the research. By coding both suggested revisions and revisions, we will be able to directly assess what type and level of written feedback learners most often provide, and whether this leads to actual revisions. Moreover, the issue of what type of feedback learners provide is in itself an important issue as recent research suggests that it is the process of giving feedback that leads to gains in writing ability (Lundstrom & Baker, 2009).

We also make an additional distinction between revisions that are related to register and those that are not. Register is important in the context of genre approaches to academic writing (Hyland & Hyland, 2006), which include the use of appropriate language features as one of the criteria for assessing students' writing. In many countries, such as Japan, prior to a university education learners of second languages are unlikely to have had much (if any) experience with learning register variations associated with different discourse communities and genres of writing. An interesting question, therefore, is whether peer feedback helps writers to improve the register-appropriateness of their own writing following peer feedback. In other words, can peer

feedback lead to more register-appropriate texts if this is one of the elements attended to in the revision process?

An additional related question concerns whether texts become more *complex* as a result of revision. Complexity can be defined in terms of multiple measures that have been used to compare texts at different levels of reading difficulty, such as the number of words per sentence, the number of words per text or the number of word types used. Such measures have been used for measuring text complexity of reading material and have been shown to predict differences in text types such as simplified texts at different levels (Crossley, Allen & McNamara, 2011a; Crossley, Allen & McNamara, 2011b). The question is important for two reasons: firstly, second language learner writing (at least at the intermediate level or below) is likely to be simpler than published academic writing (the target), when measured by basic indices related to text complexity. However, when revising a text, second language writers will work on *inter alia* improving simple sentences by linking them together (i.e., using words such as *although, whereas,* or *but*) and adding more information to support their arguments. These revisions may lead to learner texts that are more complex. If this is the case, the texts will also become more appropriate for the target genre (in the present case, published research articles), because such academic genres tend to be more complex than less formal written genres. For instance, academic writing is known for its use of complex noun phrases, which are found less often in informal styles of writing, such as fiction (Biber, Johansson, Leech, Conrad & Finegan, 1999). Because undergraduate learners are not accustomed to writing in academic style, their use of such complex sentence structures, which in turn affect simple measures such as sentence length, may be limited. However, through the process of revision it is possible that learner writing becomes more complex. Thus, in the present research we consider this issue by utilising some simple measures to assess whether peer feedback results in more complex texts.

In this paper we present a pilot study that is part of a larger research project investigating peer feedback in the second language writing classroom. We address the issues outlined above by proposing the following research questions:

1. What is the distribution of types of revisions that learners make following peer feedback (including both general and register-related revisions)?
2. What is the distribution of types of suggested revisions that learners make during peer feedback (including general and register-related revisions)? Moreover, what percentage of the reviewers' suggested comments are incorporated in the second drafts as actual revisions?
3. Does the revision process result in more complex texts? In other words, do the initial and revised drafts differ on measures of text complexity?

Study Design

Teaching and Learning Context

The present study is designed in the context of an integrated English course for first-year science-route students at a high-level Japanese university. The course aims are to introduce students to academic writing in the sciences and to foster creativity in designing simple experiments for testing hypotheses. Students are introduced to the genre of written science research papers and to register-specific features of language, first by comparing them with other genres and registers. Then, through a drafting process learners practice suggesting and making revisions related to content, language, register and formatting in order to improve their own and others' writing.

Peer Feedback Training and Procedure

Training in peer revision was provided in the form of an instructional DVD (Middleton, Allen & Shibata, 2009) that promotes the value of the activity and provides examples of peer feedback in practice. Also, the following awareness-raising activities were completed in classes prior to the revision sessions: a pair of exercises that involve analysing texts for potential issues in register and formatting and an exercise in which students critique each others' experimental designs in terms of fitness for purpose, method and practicality. In addition they were taught about the structure and features of individual sections of science reports prior to writing their own, meaning that they had a guide to the

structure of each section in terms of obligatory features or *moves* (Swales, 1990). Prior to beginning each peer feedback session, students were briefly reminded to focus on all of these features i.e., content, structure, register related features, and formatting. In this way, students learned to focus on a variety of features in their writing.

Peer feedback dyads were self-initiated (i.e., students decided who they would sit next to in class and this formed the dyad) unless students were working in groups for their research, in which case they did not do peer feedback together. Peer feedback sessions took around 30–40 minutes. During this time peers swapped papers, read quietly once through the paper, then read again and made comments on various aspects of the text. Peers could discuss aspects of each others' writing at any time but typically they read and made notes quietly for 10–15 minutes before discussing. Students used either a laptop computer with touch-screen pen, or an iPad with an annotating application for marking up peers' papers.

Participants

Two regular first-year writing classes that were held in a technology-equipped classroom were selected for the study. All learners were native speakers of Japanese and had studied English in their secondary education for approximately six years. Participants signed forms agreeing to allow their work to be used in the research. Students who did not submit a sufficient number of texts were not included in the study (i.e., both drafts of at least two sections of the report). A total of 18 students' writing was collected for use in the analysis. An additional 3 students' suggested revisions were included though their own writing was not as they failed to meet the submission criteria.

Texts (Initial/Revised)

Three separate texts were collected: initial drafts, annotated initial drafts and revised drafts. Inital and revised drafts were composed by the 'writer', and the initial drafts were annotated by the 'reviewer'. All texts were collected electronically from students over a period of five weeks; students wrote, revised and reviewed texts as part of their primary assignment (a report worth 50% of the course grade). The report included four sec-

tions: Introduction, Method, Results and Discussion (each section is classified as an individual 'text'). Mean individual text length was 196 words (Standard Deviation = 72). Topics were varied and fell under various scientific disciplines including biology, physics, psychology, chemistry and engineering; however, all followed the same format based around a simple experiment. The total number of texts collected was 120 (60 initial and 60 revised drafts), which were made up of revised/initial drafts of 16 Introductions, 17 Methods, 14 Results and 13 Discussions.

Reviewed texts were annotated by peers using iPads and laptop computers and printed out at the end of the session for the writer to take home for revision. These were also emailed directly to the teacher at the end of the class. Students were also asked to submit these annotated texts to the teacher. Due to technical problems, however, only a small number of annotated texts were received, meaning only a small subset of the total texts can be analysed in terms of the suggested revisions. A total of 36 annotated texts were collected for use in the present study (Seven Introductions, 13 Methods, 11 Results, five Discussions) from a total of 21 peers.

Coding Scheme and Procedure

Actual revisions made and those suggested by peers were both coded using the same procedure and materials. The coding scheme used in the present research was adapted from those used in previous revision studies (Faigley & Witte, 1981; cf. Berg, 1999 and Suzuki, 2008). Faigley and Witte's coding scheme is useful because it distinguishes between two levels: surface changes and text-based changes. The former include revisions that do not change the meaning of the text and the latter include revisions that do change the meaning of the text. In turn, these two levels are divided into four categories: formal changes (e.g., spelling, tense), meaning-preserving changes (e.g., word choice, active to passive sentence changes) and meaning-related changes (e.g., content and rhetoric-related revisions) at a microstructure level and meaning-related changes at a macrostructure level. We preserved the distinction between surface- and text-(meaning) related changes and within surface revisions the distinction between formal and meaning-preserving changes, but we did not distinguish between micro- and macro-structure levels as this

appeared somewhat difficult to operationalize. Instead we used the distinction between word, sentence, and paragraph level changes used by Suzuki (2008) for categorizing revisions within the meaning-related and non-meaning-related revisions categories (Table 1).

In addition, we used a separate coding scheme for register-related revisions, that is, we dual-coded revisions for register after coding them using the main coding scheme[1]. This allowed us to assess more precisely if and how peer feedback results in revisions related to register, which is an important aspect of our academic English course. Register-related revisions are divided into three categories based on teaching points in the course: *modality, subject-orientation, lexical choice.* Modality refers to the use of hedging in academic discourse, for example the use of *may*, *appears to,* and *possibly* to weaken the strength of claims. Subject-orientation refers primarily to the use of personal pronouns, which are often avoided in formal academic writing (particularly *you* and *we* used as a general reference to *the reader* or *people in general*) and to the use of passive as opposed to active sentences. Finally, lexical choice refers to the tendency for academic writing to use words that are more objective, formal and specific, and to avoid words that are subjective, informal and general. This coding scheme was used in addition to the general coding scheme to assess the number and type of register-related revisions made by participants (see Table 1 below).

Two researchers (the present authors) compared initial and revised drafts, identified revisions, and coded them according to the above classifications. Prior to coding the full set of texts, the two researchers practiced coding a number of texts and discussed discrepancies until reaching agreement. Then, the researchers coded all texts individually. After coding, the total numbers of revisions classified by type were recorded for each researcher and interrater reliability was calculated. The Kappa statistic was calculated as a measure of ratings similarity while accounting for the possibility of chance agreement, and gave a moderate similarity score of 0.65. All discrepancies were discussed until 100% agreement was attained. The above procedure was repeated for coding suggested revisions in the 36 annotated texts. Interrater reliability was measured as 0.75, and 100% agreement was reached through discussion of discrepancies.

Revisions			
Surface		**Text**	**Additional (Dual-coded)**
Formal Revisions	**Meaning-Preserving Revisions**	**Meaning-Related Revisions**	**Register-Related Revisions**
1. Mechanics: *punctuation, capitalization, spelling, spacing*	**1. Word / phrase level:** *e.g., word choice (on the other hand - however)*	**1. Word / phrase level** *e.g., word choice*	**1. Modality / Tense** (related to rhetoric)
2. Word Form: *singular-plural, subject-verb agreement*	**2. Sentence / clause level** *e.g., active-passive changes, joining sentences with coordinators*	**2. Sentence / clause level** *e.g., adding sentences or clauses with new information*	**2. Passive / Pronouns** (related to objectivity)
	3. Discourse level (within and across paragraphs): *e.g., moving sentences, rearranging sentences and paragraphs*	**3. Discourse level (within and across paragraphs)** *e.g., adding or rewriting paragraphs*	**3. Word choice:** informal-to-formal, general-to-specific, subjective-to-objective

Table 1: Coding scheme for revisions

Results

Analyses of suggested revisions and revisions are presented first in this section, followed by the analysis of text complexity. For the former analyses, we firstly present the percentage of suggestions/revisions categorised by type and then use statistical measures to assess whether observed differences are statistically reliable. Importantly, the number of suggestions/revisions are the response variables and are count, rather than continuous, data. Because we are interested in the type of suggestions/revisions, three counts are used for each text, that is, counts of formal, meaning-preserving and meaning-related suggestions/revisions[2]. By looking at the distributions of revisions (both actual and suggested; Figures 1a and 1b) it is clear that zero is highly frequent in the count data. This means that when analysing the data statistically, a normal distribution is not appropriate and instead an alternative distribution must be used. Goodness of fit test revealed that a negative binomial distribution was the most appropriate for the data. We also wanted to include text type (in other words, the specific part-genre of the research paper: Introduction, Method, Results, Discussion) as a predictor variable because this may influence the number of suggestions/revisions made. Consequently, generalized linear models were used, as

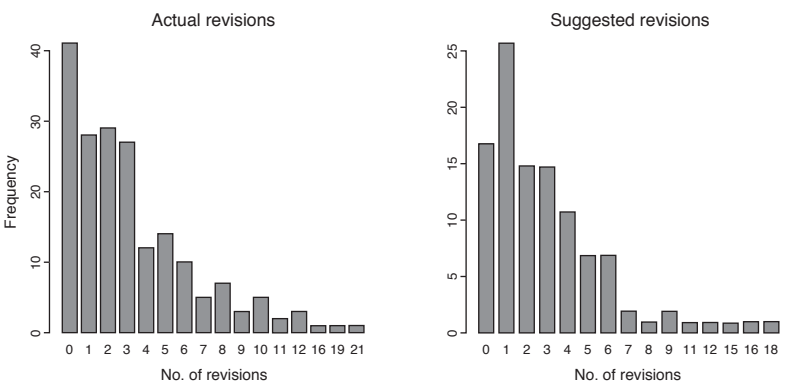

Figure 1a (Left): Distribution of counts for revisions made with the number of revisions on the x-axis and the frequency on the y-axis; Figure 1b (Right): Distribution of counts for suggested revisions made with the number of revisions on the x-axis and the frequency on the y-axis

these are suitable for count data with multiple independent variables, and the selected distribution was negative binomial. The analyses were conducted using the glm.nb function available in the MASS package in R open source software (R Core Development Team, 2010).

Suggested Revisions

The total numbers of suggested revisions made in each category as well as the proportions (%) are shown in Table 2 below. The percentage of suggested revisions made for formal errors was 45%, that for meaning-preserving errors was 34%, and that for meaning-related suggested revisions was 19%, with 2% being unclassifiable. Formal suggested revisions made up almost half of the total, which shows that reviewers were often making suggested revisions regarding formatting-related (e.g., capitalization) or more often low-level grammatical errors (e.g., articles, agreement, tense). Also, word-level suggested revisions were the most common, followed by sentence-level suggested revisions. Register-related suggested revisions made up 12% of the total of suggested revisions, indicating that reviewers were able to notice and point out register-related errors. The fewest register-related suggested revisions were made for modality, while the most were made for sentence-level issues regarding objectivity and for word-level register-related issues.

A generalized linear model was fitted with revision type (formal, meaning-preserving, meaning-related) and text type (Introduction, Method, Results, Discussion) as factorial predictor variables and the number of suggested revisions made as the response variable. The final model revealed a significant effect of revision type only ($p<.05$), with meaning related revisions being the significant factor level. Figure 2 below show the means and standard deviations for each type of suggested revision. The number of suggestions did not differ according to text type ($p>.1$).

	Total number of suggested revisions	Percentage of total suggested revisions (%)
Formal revisions	153	**45**
F1 – Mechanics	50	15
F2 – Word Forms	103	30
Meaning-preserving revisions	114	**34**
MP1 – Word level	99	29
MP2 – Sentence level	9	3
MP3 – Paragraph/ paragraph level	6	2
Meaning-related revisions	70	**19**
MR1 – Word level	45	12
MR2 – Sentence level	25	7
MR3 – Paragraph/ paragraph level	0	0
Unclassifiable	2	**2**
Total		*100*
Register-related revisions	39	**12**
R1 – Modality	1	3
R2 – Pronouns, voice	22	56
R3 – Objectivity, specificity, formality	16	41

Table 2: Number and type of suggested revisions made by peers during peer feedback

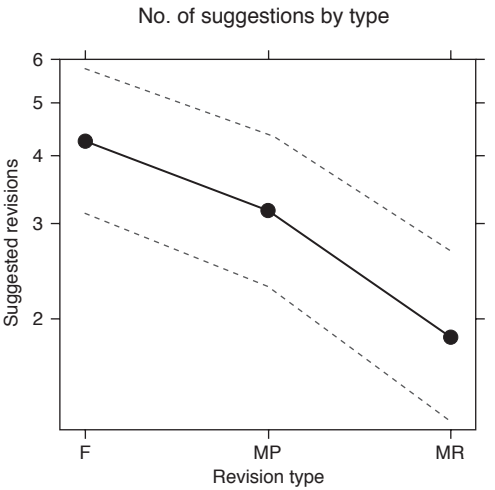

Figure 2: The mean number of suggested revisions per text are shown on the vertical axis, with the three revision types (formal, meaning-preserving and meaning-related) shown on the horizontal axis. The standard error is shown by the broken lines around the mean data points.

Revisions

The total numbers of revisions made in each category, as well as the percentages of the total number of revisions are shown in Table 3. The percentage of revisions made for formal errors was 32%, that for meaning-preserving errors was 30%, and that for meaning-related revisions was 38%. This shows that overall writers made a similar number of revisions for each broad category of revision. Looking at the word, sentence and paragraph level revisions, it is clear that while very few paragraph-level revisions were made, there were more revisions at the sentence level but the most at the word level for both meaning-preserving and meaning-related revisions. Similar to the suggested revisions, few actual revisions were made at the discourse level. Both actual and suggested revisions tended to be more focused on non-meaning-related errors. However, more meaning related revisions were made than those suggested, which shows that revisions were not completely dependent on suggestions made by peers.

Register-related revisions made up 10% of the total number of revisions, which shows learners paid attention to register

when revising. The highest percentages of register related revisions were for word-level changes related to objectivity, specificity and formality (48%) and for sentence-level revisions related to objectivity, such as removal of personal pronouns and changing active to passive sentences (38%), while the least revisions were made regarding modality (13%). The percentages of register-related revisions (10%) and suggested revisions (12%) were very similar, suggesting that during peer feedback and the actual revision process, learners were partially focused on register-related issues.

	Total number of revisions	Percentage of total revisions (%)	Percentage of suggested revisions incorporated by writers (%)
Formal revisions	197	**32**	**69**
F1 – Mechanics	80	13	70
F2 – Word forms	117	19	69
Meaning-preserving revisions	190	**30**	**66**
MP1 – Word level	145	23	70
MP2 – Sentence level	32	5	44
MP3 – Paragraph/ paragraph level	13	2	83
Meaning-related revisions	240	**38**	**64**
MR1 – Word level	132	21	71
MR2 – Sentence level	99	16	56
MR3 – Paragraph/ paragraph level	9	1	0
Total	*627*	*100*	*100*
Register-related revisions	64	**100**	**59**
R1 – Modality	8	13	0
R2 – Pronouns, voice	25	39	55
R3 – Objectivity, specificity, formality	31	48	63

Table 3: Number and type of revisions made following peer feedback

A generalized linear model was fitted as for the suggestions described previously. The number of revisions did not differ significantly by revision type ($p>.1$) but differed according to text type ($p<.001$). Figure 3 illustrates the mean number of revisions made for each text type. As can be seen the greatest difference is between the Method and Discussion sections, which contain the most and least revisions, respectively. Taken together, while the type of revision was significantly different for suggestions, the type of text was significantly different for the revisions; these results are taken up again in the Discussion.

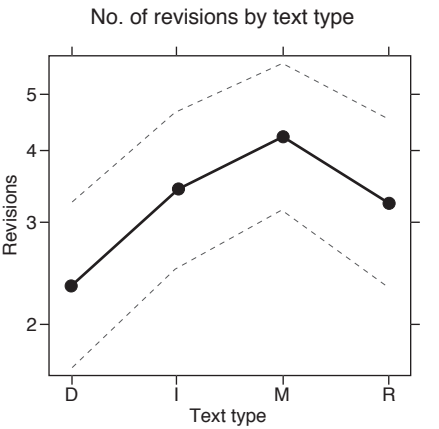

Figure 3: The mean number of revisions per text are shown on the vertical axis, with the four text types (Discussion, Introduction, Method, Results) shown on the horizontal axis. The standard error is shown by the broken lines around the mean data points.

The percentage of suggested revisions that were *actually incorporated* into the revised drafts was 69% for formal revisions, 66% for meaning-preserving revisions and 64% for meaning-related revisions. The overall mean of suggested revisions incorporated into drafts was 66%. For register-related revisions, 55% of suggestions regarding voice/pronoun use were revised and 63% of suggestions to improve the objectivity, specificity and formality of texts were actually incorporated into the revised texts. Though there were three suggested revisions for the modality category, none of these were incorporated into the second drafts[3]. Overall, 59% of register-related suggestions were incorporated into revised drafts. Taken together, these results reveal

that around a third of suggestions are acted upon by writers, indicating that writers evaluate suggestions and do not blindly incorporate all of them; this issue is taken up again in the Discussion.

General Measures of Text Complexity

The two drafts (initial and revised) were compiled into two corpora for an analysis of text complexity. A number of indices that are illustrative of general text complexity were selected from those available in the computational linguistics tool Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). These included the following seven measures: *number of words per text, number of sentences per text, number of paragraphs per text, number of syllables per word, words per sentence, sentences per paragraph,* and *number of word types per text*. One-way within-subjects Analysis of Variance (ANOVA)[4] were conducted separately with each measure as the response variable and the draft (initial or revised) as a categorical predictor variable (Table 4). Of the seven measures, *number of sentences per text, sentences per paragraph, paragraphs per text* and *average number of syllables per word* were not significant ($p>.05$). The measures that were significant ($p<.05$) showed that revised

	Initial	Revised	Mean Sq	F (1, 17)	p (>F)
Number of words	190.0 (75.6)	201.8 (68.3)	4393.10	9.762	0.006**
Number of word types	88.3 (29.9)	93.4 (27.6)	828.01	9.173	0.008**
Words per sentence	17.5 (4.7)	18.2 (4.9)	14.60	4.845	0.042*
Number of sentences	11.4 (5.0)	11.8 (4.8)	4.96	2.456	0.136
Sentences per paragraph	3.9 (2.1)	4.2 (2.3)	1.68	0.920	0.351
Number of paragraphs	3.5 (2.6)	3.5 (2.5)	0.03	0.049	0.828
Syllables per word	1.5 (0.1)	1.5 (0.1)	0.00	0.012	0.916

Note: Standard deviations are in parentheses; asterisks indicate the degree of probability **= $p<.01$, *=$p<.05$

Table 4: Analysis of variance table for text complexity measures and draft type (initial/revised)

texts had a greater average number of words per text, words per sentence and word types per text. The increased complexity of revised texts compared to initial texts thus appears to be geared towards local, word-level revisions primarily as opposed to inter-sentence- and inter-paragraph-level changes. This suggests that writers tended to focus on localized revisions as opposed to more global, paragraph-level revisions, supporting the observation from the main analysis that most revisions tend to be at the word and sentence rather than paragraph level.

Discussion

The analysis of suggested revisions showed that reviewers tended to focus on surface-level revisions and meaning-preserving revisions more than meaning-related revisions. Thus, when reviewing texts, although revisions at all levels are valid and necessary, it appears that learners were taking a potentially less critical approach to the appraisal of their peers' writing by avoiding a focus on meaning. Moreover, these revisions were also restricted primarily to the word and sentence levels, which supports the finding that learners were not appraising the overall rhetorical structure of the texts and were localising their attention to sentence-by-sentence suggested revisions. Training that focuses on content-related issues and raising questions when meaning is ambiguous may be beneficial before students begin peer revision activities.

The analysis of revisions made following peer feedback showed that learners made roughly similar numbers of revisions at the surface-, meaning-preserving and meaning-related levels. Given that learners were instructed to revise at these various levels through training activities, such as noting exercises, text analysis and observing examples of peer feedback, this is probably unsurprising. However, these findings do show that appropriate training can lead to consideration of revisions in various domains (content, structure, language, and formatting). Previous research showed that content-related negotiations were more numerous following peer feedback than following self-review (think-aloud methodology was used to assess negotiations in self-review), leading to the conclusion that peer feedback could be used for content-related revision and self-revision could be

used for other textual revisions (Suzuki, 2008). Our results demonstrate that with appropriate training, revisions are made for both meaning-related and non-meaning-related issues. Register-related revisions were also made (10% of the total revisions), which shows that training learners in register prior to peer feedback can lead to revisions at this level. Thus, these results indicate benefits of peer revision in helping learners to use appropriate language and register, which is a cornerstone of proficient academic writing.

An interesting finding is that while 38% of revisions were meaning related, only 19% of suggested revisions were similarly focused on content. This suggests that peer feedback may lead to increased engagement with content on the writers' part, even if meaning-related suggestions were not made on the text. This is line with previous research which showed peer feedback led to increased focus on content-related issues compared to self-review (Suzuki, 2008). Alternatively, the greater number of revisions compared to suggestions could be because writers were still modifying and adding content to their writing during the revision period (e.g., adding and modifying details of the experimental procedure during the revision period). The numbers of register-related revisions and suggestions were comparable (10% and 12%, respectively), thus showing that both writer and reviewer considered register during the peer feedback process.

Around two-thirds of suggestions were actually incorporated in the revised texts. This is slightly more than in Mendonca and Johnson's (1994) study where just over half of peers' revisions were incorporated, intimating that the reviewers' suggestions in the present study were potentially of higher quality (or the writers were better able to understand and utilise the feedback). However, the question arises as to why a third of suggested revisions were not incorporated. There are a number of possibilities: firstly, the suggested revisions were inaccurate, though from observing the data, these appeared to be rare. Previous work has also shown that overall few of peers' suggestions tend to be misleading (Jacobs, 1989). Also, a number of revisions were of the optional variety, such as suggestions for changing active sentences to passive ones. Previous studies have also shown that optional revisions are suggested, though are few in number. Ferris (2006) showed that 7% of writing instructors'

error corrections were coded by other researchers as 'unnecessary' (while 3.6% were incorrect and 89.4% were correct)[5]. Secondly, the writer may have simply failed to notice the suggested revision due to haste in preparing the second draft (for example, only the first part of one of our texts was revised and the latter half was not revised at all). Thirdly, large-scale rewriting could mean that a particular suggested revision becomes no longer applicable, leading to it not being revised, though this is unlikely as such large scale revisions were rare (as shown by the lack of discourse-level revisions). These explanations combined may account for the non-incorporated suggested revisions, but this is an area for future research.

One key finding is that revisions tended to be made at the word and sentence level as opposed to the paragraph level. This is in line with previous findings (Berg, 1999; McGroarty & Zhu, 2002; Suzuki, 2008). Regarding the complexity of texts following peer feedback, we found that revised texts were significantly different from initial drafts and that these differences were also primarily related to complexity at the word and sentence level, as opposed to the paragraph and inter-paragraph level. The revised texts in the present study contained more words per text, more words per sentence and a greater number of word types. Previous studies have shown that such features are indicative of more complex texts, such as those written for advanced readers when compared to simplified texts written for lower proficiency learners (Crossley, Allen & McNamara, 2011). These features of text complexity are related to cognitive processing of texts, influencing readers' ability to parse texts efficiently. In the present educational context, one aim is to help learners to better appropriate complex texts, moving towards a formal, objective and specific style of writing that is suitable for academic texts. These results show that through the process of peer feedback learners can and do improve the complexity of their writing. However, we also found that the measures that were significant were restricted to the word and sentence level, which shows learners concentrated their revisions at these levels as opposed to the intra- and inter-paragraph levels. There are a number of possible reasons why this may be the case. One reason for the lack of revisions at the paragraph level may be that such revisions are more difficult as they require reviewers to process the whole text and relate infor-

mation across paragraphs. Revisions at this level typically involve moving sentences within and across paragraphs and adding new paragraphs and are concerned with the development of rhetorical arguments within and across paragraphs. Another reason for a lack of such revisions may be that in the present study learners were provided with a genre model of each section of a science article (e.g., Introduction section), which may have constrained the moving around of paragraphs and sentences within paragraphs. As this was the first time learners had written a research paper, it is understandable that they followed the genre model closely and thus the rhetorical patterning remained relatively fixed across drafts.

Finally, regarding the main effect of text type, the overall pattern of the number of revisions shows an increase (Introduction, Method) then decrease (Results, Discussion) over time. Three reasons are suggested for these possible differences: a general fatigue effect, rhetorical and content-related complexity, and diminished effects of training. As learners take around sixteen classes per semester it is possible that as the semester progressed they became overburdened with other subjects and thus spent less time revising their later texts. Alternatively, the rhetorical complexity of the Discussion section as opposed to the relative simple rhetorical structure of the Method section may have impacted the number of revisions made, due to an insufficient understanding of how to revise this section. Finally, the peer feedback training session was given in the sixth week of the course and peer feedback for the Discussion section was done in the tenth week; this gap between training and feedback may have resulted in fewer suggested and actual revisions at the later stages. In regard to the latter two explanations for the reduction of revisions in the later sections, both refresher and additional (text type specific) training may be required prior to peer feedback of Discussion sections. Specific training of peer feedback for rhetorically complex text types may thus prove to be a successful intervention in the aim of increasing the number of revisions made in these sections.

Limitations and Future Directions

The present study was intended to provide a preliminary overview of the revision process, how writers and reviewers interact

with the text and whether text complexity develops as a result. However, there are a number of limitations that should be addressed in follow-up studies. Perhaps most importantly is whether the suggestions and revisions were actually beneficial; in other words, in the present study it was not determined if they actually led to improved texts. Having raters assess the quality of suggestions and revisions is a possible way to investigate this; however, a more comprehensive method would be to have the texts rated for quality by trained raters. These are targets for future research.

In addition, the complexity analysis could also be complemented with a human-rater quality analysis to confirm whether increases in complexity also lead to more register-appropriate texts, rather than relying on inference. In terms of measuring text complexity, a greater range of indices are available that may further help to describe the types of textual changes that occur through the process of peer revision. Measures such as word frequency, noun phrase density and other measures of syntactical complexity would help to achieve this aim.

Finally, a qualitative analysis of *why* writers choose to incorporate suggestions or not would be highly informative. This is an interesting issue that may relate to the interpersonal dynamics of peer feedback dyads, as well as the quality and type of feedback provided.

In terms of pedagogy, the level of engagement with the discourse level of the text and the generally minimal focus on meaning-related issues when reviewing are issues that need addressing in the classroom by way of training and activities. A general re-evaluation of peer feedback training in terms of the amount, type and frequency of training would allow further consideration of these issues and possible intervention strategies.

Acknowledgements

The authors would like to thank Scott Crossley for generating the Coh-Metrix output, Paul Rossiter and Alex Gilmore for helpful comments and discussion at various stages of this research, and Akiko Nakazawa for assistance in the Komaba Active Learning Studio. Also, we would like to thank all of our students for their hard work in the writing classroom.

Notes

1. Register related revisions are those revisions that are coded as meaning-preserving, because revising words and clauses to be more register-appropriate rarely results in a change in meaning.
2. These three categories make up the levels of the revision type factor. We did not do the same for word-, sentence- and paragraph-level revisions, because primarily we are interested in differences according to revision type; rather the differences according to level of revision are discussed in relation to the complexity analysis.
3. The average number of suggested revisions by text type and revision type were not conducted because the small number of texts for each text type means comparisons were not reliable.
4. The dependent variables all had normal errors, meaning ANOVAs were appropriate for the statistical analyses.
5. This figure almost matches the number of incorrect and unnecessary corrections advised by instructors, though it is not clear if the two figures were related in this way. Thus, even with teacher feedback a small percentage of the feedback given is not utilised by the writers when revising their work. In her study, less than 10% of suggested corrections made by teachers were not acted upon by teachers.

References

Berg, C. E. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing, 8*, 215–237.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Chaudron, C. (1983). Evaluating writing: Effects of feedback on revision. ERIC Document Reproduction Service No. ED 227–706.

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.

Crossley, S. A., Allen, D., & McNamara, D. S. (2011). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16* (1), 89–108.

Crossley, S. A., Allen, D., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in A Foreign Language*, *23*, 84–101.

De Guerrero, M. C. M., & Villamil, O. S. (1994). Social-cognitive dimensions of interaction in L2 peer revision. *The Modern Language Journal, 78,* 484–496.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and*

Communication, 32, 400–414.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In Hyland, K. & Hyland, F. (Eds.) (2006). *Feedback in Second Language Writing: Contexts and Issues Applied Linguistics Series.* Cambridge University Press: Cambridge.

Graesser, A. C., McNamara, D. D., Louwerse, M. L., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36,* 193–202.

Guerrero, M. C. M. de, & Villamil, O. S. (1994). Social-cognitive dimensions of interaction in L2 peer revision. *The Modern Language Journal, 78*, 484–496.

Hyland, K. & Hyland, F. (Eds.) (2006). *Feedback in Second Language Writing: Contexts and Issues Applied Linguistics Series.* Cambridge University Press: Cambridge.

Jacobs, G. (1989). Miscorrection in peer feedback in writing class. *RELC Journal, 20* (1), 68–76.

Liu, J., & Hansen, J. G. (2002). *Peer response in second language writing classrooms.* Ann Arbor: Michigan 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.

McGroarty, M. & Zhu, W. (1997). Triangulation in classroom research: A study of peer revision. *Language Learning, 47* (1), pp. 1–43.

Mendonca, C., & Johnson, K. (1994). Peer review negotiations: revision activities in ESL writing instruction. *TESOL Quarterly*, *28*, 745–769.

Middleton, G., Allen, D. & Shibata, N. (2009). *An Introduction to Peer Review.* (DVD). KOMEX and the Department of English Language, College of Arts & Sciences, University of Tokyo.

Min, H. T. (2005). Training students to become successful peer reviewers. *System*, *33*, 293–308.

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*, 118–141.

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8* (3), 265–289.

R Development Core Team (2010). R: A language and environment for statistical computing version 2.11.1. http://www.R-project.org.

Suzuki, M. 2008. Japanese learners' self revisions and peer revisions of their written compositions in English. *TESOL Quarterly, 2*, 209–233.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge, England: Cambridge University Press.

Villamil, O., & De Guerrero, M. (1996). Peer revision in the L2 classroom: Social-cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing, 5*, 51–75.

Zamel, V. (1985). Responding to student writing, *TESOL Quarterly*, 19, 79–102.

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

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