

# Peer Feedback Comments and Revision Strategies of L2 Writers: Insights from the Analyses of Their Reflection Texts

Yukako Hatakeyama

## Abstract

This paper aims to investigate what types of feedback comments pre-intermediate to intermediate learners of English in a women's university in Japan generate in peer review. It also intends to clarify how the learners revise their essays. The participants consisted of two groups: One did peer review five times, and the other received automated writing evaluation (AWE) feedback four times before one peer-review occasion. The participants' reflection texts were used as data for analyses. The findings suggest that second language (L2) learners provide more critical feedback than compliments and more comments on global aspects of essays than local points in both groups. Time-series analyses and comparison with a control group that received AWE feedback showed that the number of peer-review opportunities contributed to increased critical comments and the comments on global points. It was also found that more specific and writer-friendly feedback comments consisting of two or three steps increased as the participants became more experienced in peer review. In both groups, the rate of the writers' incorporation of critical comments into their revisions was found to be high, and they also used their own ideas and judgments to revise their drafts. Implications of the findings and limitations of this research are discussed.

**Keywords:** peer review, L2 writing, feedback comments, AWE, revision strategies

## 1. Introduction and Literature Review

Writing in academic contexts is a complex and demanding activity. Indeed, it requires various skills and knowledge: content knowledge, rhetorical understanding, genre awareness, control of linguistic and extralinguistic features, and strategies for managing the writing tasks, to name a few (Ferris, 2018a). This is already a challenge for those writing in their first language (L1). Yet second language (L2) writers face additional challenges since they are still in the middle of second language acquisition.

## 1.1 L1 and L2 Writing Pedagogy

Even though there are differences between L1 and L2 writers, the writing pedagogy for both shares characteristics, and L1 writing pedagogy has significantly benefited the teaching practices and research of L2 writing. Notable trends include the process approach, viewing “writing as a process of developing organization as well as meaning,” which seems to have gained wide acceptance in the L2 writing pedagogy (Matsuda, 2003, p.21). The process approach allows writers to learn from writing multiple drafts and obtaining feedback by helping them brainstorm ideas and plan their writing and guiding them through the revision and editing stages (Ferris, 2018b). Consequently, it contributes to “building confidence in L2 writers” (pp.91-92).

The influence of L1 writing pedagogy on its L2 counterpart is not limited to the process approach. For one thing, the value of reflection and metacognition is recognized. Reflective writing tasks help writers learn about writing rather than merely completing assignments (Ferris, 2018a). These tasks include “goal-setting,” “self-evaluation of progress made,” and “real-time analysis of what the student has learned from completing a specific writing task” (p.92). For another thing, collaborative activities are widely conducted in L2 writing classes, with the advancement of technology that enhanced co-authoring and co-editing pieces of writing (Liu & Hansen Edwards, 2018; Sommers, 2013).

## 1.2 Peer Review

One essential process of writing that is collaborative is peer review or peer feedback. Peer review “gives writers opportunities to discuss their texts and discover others’ interpretations of them” (Hyland & Hyland, 2019, p.7). Peer review uses “learners as sources of information and as interactants for each other in such a way that learners assume the roles and responsibilities more typically taken on by a formally trained teacher, tutor, or editor in commenting on and critiquing each other’s draft in both written and oral formats in the process of writing” (Liu & Hansen-Edwards, 2018, p.1). Peer Review is supported by various theories, such as sociocultural theory (including Vygotsky’s concept of Zone of Proximal Development (ZPD), Vygotsky, 1978) and the interaction hypothesis (Long, 1996) in second language acquisition (SLA).

### 1.2.1 Benefits of Peer Review

Numerous potential benefits of peer review on L2 writers have been explored to date. Research findings suggest that peer review raises the writers’ awareness of the readers (Mendonça & Johnson, 1994; Rollinson, 2005; Tsui & Ng, 2000) and enables the writers to receive another and different kind of feedback in addition to that given by the teacher (Ferris, 2018c; Rollinson, 2005). Further, it provides opportunities for meaningful interaction with peers and collaborative learning (Mendonça & Johnson, 1994; Tsui & Ng, 2000). It has also been claimed that peer review encourages and motivates L2 writers since feedback comments from peers, who share the same experience of writing, are supportive and less authoritative compared to teacher feedback (Rollinson, 2005).

Peer review benefits not only the writers but also the readers. Lundstrom and Baker (2009) examined which—giving or receiving feedback—is more beneficial to improving students’ writing and found that the gains in writing ability of the givers of feedback were more significant than those of the receivers. By

reading their peers' drafts, L2 writers' awareness of their strengths and weaknesses was also found to be raised (Tsui & Ng, 2000). Furthermore, reviewing peers' drafts helps them learn critical thinking skills to become critical readers and revisers of their own writing (Berg, 1999; Rollinson, 2005; Sommers, 2013; Yu & Fu, 2017), which leads to the improvement of their text quality.

### 1.2.2 Concerns about Peer Review

Even with the above-stated benefits, concerns about peer feedback have been discussed. Given its time-consuming nature, whether peer-review is worth conducting with time being spent has been discussed (Rollinson, 2005). Indeed, this activity involves reading a draft, making notes, and communicating with the writer, which, altogether, takes up a considerable amount of time.

The first concern is the writers' reluctance to engage in peer review. This reluctance could come from their lack of enough grammatical knowledge or their fear of giving inappropriate comments (Srichanyachon, 2011) and their doubt on the peers' ability to provide feedback (Tsui & Ng, 2000). However, according to Tsui and Ng (2000), regardless of the degree of uptake of peers' comments, L2 writers regarded peer comments as having certain roles to play. Similarly, Hatakeyama (2010) reports that L2 students at a bilingual university in Japan perceive peer review favorably, feel that getting feedback from another perspective is stimulating, and find it interesting to learn by reading peers' papers.

Second, the effectiveness of peer review, especially in the Asian context, has been of concern, given its cultural backgrounds and educational practices. According to Carson and Nelson (1996), Chinese participants provided feedback that pleases the writer rather than improving the writing; they were found reluctant to initiate especially critical comments due to their intention of maintaining group harmony. However, Ferris (2018b) claims, "student discomfort should not dissuade L2 writing teachers from using collaborative group work as an integral part of the class" (p.101). It is because "the ability to collaborate productively around a piece of writing or other form of communication has become a core skill for students and future professionals" (p.102). Indeed, many studies report peer review in Asian contexts, including Japan (e.g., Hatakeyama, 2010; Smith, 2019). Chinese EFL university students were found to provide mainly critical and constructive feedback (Yu & Hu, 2017). Studies show successful peer-review outcomes by conducting training in Taiwanese EFL university classrooms (Min, 2005, 2006).

### 1.2.3 Foci of Peer Review: Global or Local

One topic that has been calling the attention of researchers and teachers is the foci of peer comments. A study reports a student's behavior of critiquing peers' drafts by primarily paying attention to grammatical errors while the other focused more on idea development, a variety of expressions, and sentence structures (Yu & Hu, 2017). Indeed, error-free writing does not necessarily mean good academic writing, and other and non-grammar-related elements are also essential factors in assessing the quality of writing. For instance, although in a timed-writing testing context, Educational Testing Service® (n.d.) provides criteria to evaluate writing in TOEFL iBT® tests: a) writing topic and task completion, b) organization and development of ideas with explanations, exemplification, and details, c) unity, progression, and coherence, d) language use, syntactic variety, appropriate word choice, and idiomaticity. Useful feedback comments should not focus

only on local grammatical errors but also on other aspects.

Further, Liu and Hansen-Edwards (2018) claim that feedback comments that “address global issues such as content and rhetoric as specifically as possible” are most helpful (p.58). Global issues here include content (topic development and clarification of content) and rhetoric (organization patterns and the effectiveness of topic sentences, thesis statements, concluding paragraphs, concluding sentences), and global grammatical errors that interfere with the intended meaning. L2 writers who have been trained in peer review are found to spot “content and rhetorical problems as well as grammar and style issues” (p.58).

#### 1.2.4 Adoption of Reviewers’ Comments into Revisions

Another point of interest in peer review is how much reviewers’ comments are incorporated in the writers’ revisions. Some L2 writers are reluctant to incorporate peer feedback into their revisions (Mendonça & Johnson, 1994). Tsui and Ng (2000) found that many participants favored teacher feedback more than peer feedback. Min (2006) reports that writers incorporated less than half of peer comments in revisions before intensive peer-review training. Liou and Peng (2009) show that the adoption rates of revision-oriented peer comments remained just below 50%, both before and after peer-review training. Among many elements influencing the adoption rate, improving the quality of peer comments may be one crucial factor besides other features such as writer perceptions and the timing of peer and teacher feedback.

Although the quality of peer comments is of primary importance, additional points need to be taken into account. Perhaps it could be maintained that L2 writers’ selective incorporation of suggestions may not automatically mean distrust of their peers. Instead, it could be regarded that L2 writers’ ownership of text has been fostered. It is possible that they “feel that they have autonomy over their own text and can make their own decisions on whether they should take the comments on board or not” (Tsui & Ng, 2000, p.167). If so, sources of revisions other than peer comments are worth investigating in that it could contribute to a better understanding of the L2 writers’ revision strategies.

#### 1.2.5 Peer Review Training

As mentioned above, conducting training before peer review is an essential means to realize effective peer review. As Liu and Hansen-Edwards (2018) state, “It is not realistic to expect that students will be able to read effectively and make constructive comments without instruction in how to do so. Nor is it reasonable to expect that students can respond constructively to the comments made by their peers on their papers without instruction” (p.108). Thus, training is vital in forming students’ perceptions toward peer review, the types of feedback they make, and the extent to which they incorporate peer suggestions.

Many successful instances of peer review training have been reported to date. After providing training consisting of 11 steps—ranging in time from 5 to 45 minutes each—to ESL students in a university-based intensive English program, they made more meaning-type changes in their revisions and produced higher quality texts than their counterparts without any training (Berg, 1999). Min (2005) provided students with intensive training—consisting of four hours of in-class demonstration and modeling and one hour of individual conference per student—to improve vague or rubberstamp feedback comments they produce. It was found that, after the training, the students generated more comments that facilitate revisions and made

more relevant and specific comments on global points. After the same peer review training, the students incorporated a higher number of reviewers' comments into revisions. The revisions with enhanced quality were higher in number than before the training (Min, 2006). The reviewers' stances have also changed from a dominant prescriptive stance to a more collaborative stance (Min, 2008). Smith (2019) observes that the implementation of a four-step cycle of the peer review process—reviewer's observation comments, reviewer's advice, reviewer's praise/encouragement, and reviewee's re-evaluation of the essay—facilitated effective and smooth peer review and the development of revision skills on the writer.

Indeed, successful peer review training benefits both writers and readers because writers can get more specific and constructive feedback. Reviewers will also provide useful feedback on both global and local points, leading to a higher adoption rate of their comments by writers. Even with these benefits, in some writing courses, it may not be feasible to conduct intensive peer review training as described in the literature due to various constraints. Thus, whether reviewers could learn to provide feedback without intensive training but with some intervention and multiple peer-review opportunities is a question worth investigating.

### 1.3 Automated Writing Evaluation

Automated writing evaluation, or AWE, is another source of feedback in addition to peer and teacher feedback. AWE programs have been developed to meet the challenges that writing instructors face: the time and skill required to evaluate students' multiple drafts (Warschauer & Ware, 2006). AWE programs are now commercially available with functions of giving both numerical scores and evaluative feedback.

The overall validity of AWE programs, in terms of the correlations between computer-generated scores and human rater scores, is high in large standardized tests (Warschauer & Ware, 2006). Yet research in classroom contexts is in the ongoing process of being accumulated. Chen and Cheng (2008) report how AWE was used with advanced English learners in Taiwan—third-year university English majors. It turned out that AWE use was not perceived positively by the learners, in part due to the limitations in its function, such as favoring lengthiness, failing to recognize incoherent or illogical writing, and generating formulaic and unspecific feedback. Writing instructors' pedagogical practices were found to affect learners' perceptions as well. Moreover, the integration of AWE and teacher assessment was found to offer practical benefits of enabling teachers to focus more on meaning and reduce the impact of AWE limitations. Since learning goals and expectations of the advanced learners could have been beyond AWE feedback, the researchers discuss that learner proficiency and learning goals (e.g., acquiring fundamental or advanced writing skills) need to be considered in implementing AWE.

Li et al. (2014) researched 67 first-year ESL students of various majors in a U.S. university. The correlation between the feedback of AWE and instructors showed mixed results. While the moderate correlation was found in the first assignment, the low and insignificant correlation was found in the latter writing assignment, which could be attributed to the skewed and narrowly-ranged distribution of AWE scores in the latter. Interestingly, instructors' perception of high trust on low AWE scores and low trust on high AWE scores—that high AWE scores do not necessarily mean good writing described by the instructor—coincides with the results. It was also found that the students showed higher trust in AWE

scores than in Chen and Cheng's (2008) study.

These two studies are part of the growing body of research. However, the findings imply that writing instructors need to bear in mind the possible limitations of AWE scoring in classroom contexts, such as its weight on accuracy and organization rather than content and limited ability to differentiate between written texts of relatively high quality. Nevertheless, AWE's strength in detecting grammatical and spelling errors can be effectively used to raise awareness of L2 learners on these aspects and help instructors focus on more global and meaning-related points in reviewing and assessing learners' texts. Thus, instructors need to consider where AWE should be placed in the writing process, whether and how to incorporate human feedback with AWE scoring, and whether and how AWE scores are put into the assessment.

## **2. This Study**

This study's primary purpose is to explore the behavior of Japanese female university students—L2 English learners—during and after peer review. More specifically, it first aims to determine whether L2 reviewers in an Asian EFL context give critical feedback (e.g., Min, 2005, 2006; Yu & Hu, 2017) or mostly praise (e.g., Carson and Nelson, 1996) in peer review. It also intends to clarify what types of feedback comments reviewers—without intensive training—provide, drawing on Min's (2005) four steps—clarification of writers' intentions, problem identification, explanation of the problems, and specific suggestions. Third, it will examine whether the number of peer-review opportunities affects the reviewers' feedback comments in the above two points. The fourth objective is to clarify whether the reviewers provide comments on global or local points. The fifth goal is to examine the degree to which peer comments are incorporated into writers' revisions. Finally, it also tries to identify the sources of revisions of writers' texts.

The six research questions are summarized below.

- 1) Do female university students in an Asian EFL context give critical feedback in peer review?
- 2) Do they provide specific and writer-friendly feedback comments without intensive training?
- 3) Does the number of peer-review opportunities affect the characteristics of feedback comments?
- 4) Do they provide more feedback on global points or local points?
- 5) To what degree are peer comments incorporated into L2 writers' revisions?
- 6) What are the sources of the L2 writers' revisions?

As for the first and second points, the prediction was that the reviewers would provide critical feedback to a certain extent (cf. Min, 2005, 2006) and that the reviewers mostly provided less specific comments. The influence of peer review opportunities was anticipated to be positive in giving critical feedback. It was expected that the reviewers would focus on global points rather than local points. Finally, the incorporation rate of peer comments was expected to be high, and the sources of revisions were expected to be mostly peer feedback comments.



### **3. Method**

#### **3.1 The Course**

Intermediate English (S/W) is a course for sophomore students to enhance their speaking and writing skills in a high-level women's university in Tokyo. In this course, the students write five essays of different types—descriptive, narrative, comparison-contrast, cause-effect, and opinion—in one semester over a period of 13 weeks. Each piece is about 300 to 450 words in length and becomes longer as the semester proceeded. They also have opportunities for spoken interaction and presentation.

In the first semester of the academic year 2020, I taught two classes of Intermediate English (S/W) to mostly sophomore students aged presumably around 20. All lessons were given online, using Zoom, a cloud-based online video conferencing service. Since instructors are given discretion on grade components and their relative weight, in my class, the students' grades were largely determined by the quality of their writing assignments (10% X 5 essays) and reflections (10%).

#### **3.2 Participants**

Of the two classes, one class mainly consisted of humanity majors (Group A), and the other life science majors (Group B). Those in the two classes' English proficiency was presumably equivalent and low-intermediate to intermediate; both groups belonged to the lowest of the three levels at this university. The results of the TOEFL ITP test—that the students had taken upon entering university in April 2019—show those in Group A obtained somewhat lower scores (average: 439; standard deviation: 23.7; range: 397-467) than their Group B counterparts (average: 458; standard deviation: 13.7; range: 433-477), in part due to the two outliers who scored around 400. The scores translate that many belonged to the upper level of CEFR A2 (ITP score range: 337-459), with some to the lower level of B1 (460-542), at the time of test-taking (ETS, n.d.). Thus, it could be stated that they were lower-intermediate to intermediate in proficiency at the research time.

The participants in this study were mostly sophomores who belonged to either of my two Intermediate English (S/W) classes (Group A or Group B) in the first semester of the academic year 2020 and those who gave consent to their participation in this research. The research outline was explained orally to all the students in two classes, and it was emphasized that participation in this research is entirely voluntary and does not affect their grades. A link was provided to the consent form using Google Forms, which gives the same explanatory information as my verbal explanation. A total of 29 students—12 students from Group A and 17 students from Group B—agreed to participate.

#### **3.3 The Class Flow**

The two classes used the same textbook, *Effective Academic Writing 2*, published by Oxford University Press. Lessons proceeded following the textbook sequence, where one unit corresponds to each of the five types of essays. At the beginning of each unit, the learning outcomes (LOs) of this unit were explained. LOs included using sensual information to describe the food (in descriptive essays), using time-order signals effectively (in narrative essays), using expressions for comparison and contrast (in comparison-contrast

essays), and introducing counter-arguments and refutations (in opinion essays). Students also read and analyzed two sample essays in each unit to have a concrete understanding of their rhetorical characteristics, including structure, target grammar, and expressions. Explanation of these rhetorical and grammatical foci was given as well. Speaking activities and textbook exercises were also done in the lesson.

In terms of writing, the process writing approach was adopted. Students first brainstormed topics for their essays and shared them with their partner(s). They then engaged in pre-writing activities such as drawing mind-maps, listing the main points, or drawing Venn diagrams. They also had opportunities to explain their ideas and pre-writing work with peers. After creating the outlines, they shared them with their partner(s) and started writing a draft. Upon completing their drafts, they checked their texts using Grammarly, a free grammar-checking software. This helped them have confidence that their drafts are mostly error-free and help the reviewers focus more on global issues. After a week, in the next lesson, they received feedback either by 1) peer review (Group A) or 2) AWE (Group B). Subsequently, a few students shared the feedback they received with the class. Students submitted their revised final essays within a week. They also wrote “A Note to the Teacher,” a short reflection note submitted in Google Forms, explaining what peer feedback they received and how they revised their essays, which was adapted from Sommers’ (2013) “Dear Reader Letter.” Students could choose either English or Japanese to write this note, which was to ease their burden and help them freely express their thoughts. It turned out that about 90% of entries were in Japanese. For the visual explanation of the flow of lessons and writing processes, see Figure 1.

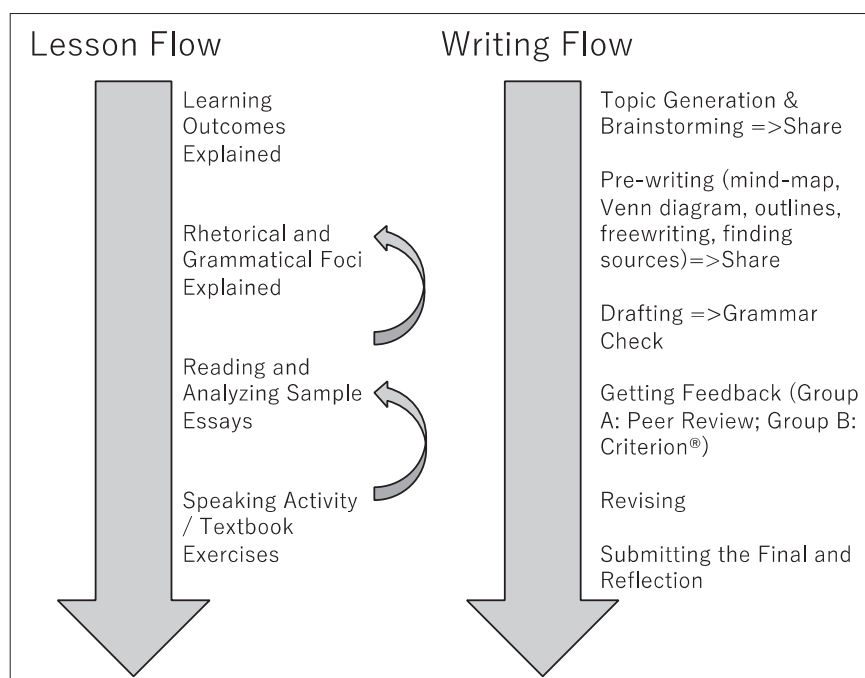


Figure 1: A sample flow of lessons and writing processes

### 3.4 Two Treatment Groups

As briefly mentioned in the previous section, both Group A and B received the same lessons and did the same activities except how they received feedback on their first four essays. Group A students conducted peer



review in which they gave and received feedback to and from their peers. In contrast, Group B students had their drafts checked by Criterion®—an online writing evaluation service (AWE) provided by Educational Testing Service (ETS)—and shared the feedback in small groups. Both groups did peer review in the last essay assignment.

### 3.5 Peer Review Procedure

In this study, peer review was done in class using Zoom. No formal and intensive peer-review training was conducted, but before the first peer review session, the benefits of peer review and what the readers and writers need to keep in mind were communicated. The idea that the readers are expected to tell what they liked candidly, noticed, didn't understand, or wanted to know was emphasized. It was also underscored that whether the writers follow the reviewers' advice is up to the writers.

In each peer review session, the peer-review checklist, which reflects the LOs of each unit, was distributed (online) to students and explained. In the checklist, space to note the best parts/features of the drafts and suggestions for the writers had been created. After the checkpoints were confirmed, the students were randomly put into pairs in Zoom breakout rooms and exchanged their drafts using Zoom or email. About 15-20 minutes were given to read the partners' drafts and take brief notes in the checklist, and an additional five minutes were secured for oral feedback in either English or Japanese. During this five-minute oral feedback session, the writers took notes of the reviewers' feedback comments. After the session, a few students shared with the class the feedback they received.

### 3.6 AWE (Criterion®) Feedback Procedure

Group B students received feedback from Criterion® on the first four essays; they peer-reviewed the last (i.e., fifth) essays only. To receive feedback, the students copied their drafts and pasted them in the designated space in Criterion®. After receiving feedback on such points as the text structure and repetition of words, they shared the feedback in small groups in Zoom breakout rooms. Here, some asked their peers questions to solve the issues pointed out by Criterion®, and some showed and explained the feedback they received by sharing their screens. After the breakout session, a few students shared with the class the feedback they received.

### 3.7 Data Collection

The participants' short reflection texts ("A Note to the Teacher") submitted in Google Forms upon completion of their writing tasks were used as data for analysis. For Group A students, who did peer review throughout the semester, two different prompts were used for all five essays: 1) What kind of feedback/advice did you receive from your partner? 2) How did you revise/change your essay? For the students in Group B, who received feedback four times from Criterion®, the above prompts were used only for the last essay. The participants' reflection texts were compiled, and the names and student IDs associated with the texts were deleted to make the data anonymous. In this way, the data sets have been prepared. The entire procedure had been approved by the university's Humanities Research Ethics Committee on May 29, 2020 (Number 2020-9).

### 3.8 Data Analysis

To answer the first to fourth research questions, the participants' answers to the first prompt—"what kind of feedback/advice did you receive from your partner?"—were manually coded and analyzed. Five sets of reflection texts were obtained from Group A participants, one set from Group B participants. When a participant received comments including more than one point, their comments were divided by point and coded. The first criterion for coding was whether the feedback was about good points (compliments) or problems/suggestions. The second criterion was inspired by Min's (2005) four-step criteria exploring the characteristics of reviewer's comments in giving feedback (i.e., the second research question): [clarification], [problem], [explanation], and [suggestion]. The data summarized according to the two criteria were compared in time series and by groups to answer the third research question investigating the effects of the number of peer-review opportunities on the characteristics of feedback comments. Concerning the fourth research question, the feedback comments were categorized using Liu and Hansen-Edwards's (2018) accounts regarding global or local points.

The participants' responses to the second prompt—"how did you revise/change your essay?"—were manually coded to answer the fourth and fifth research questions.

## 4. Results

### 4.1 Types of Peer Comments: Compliments or Critical Comments

Table 1 summarizes the types of feedback the participants received; the participants' feedback—reported by the writers (i.e., the receivers of feedback)—was categorized into compliments and critical comments. The compliments' ratio was 35% on average, ranging from 21% to 42% in Group A, and the percentage of critical comments was 65% on average, ranging from 58% to 79%. A time-series analysis reveals that the ratio of critical comments gradually increased. Although the proportion of critical comments (56%) was higher than that of compliments (44%), the difference turned out to be less salient in Group B. Interestingly, the compliments-critical comments ratio of the first essay of Group A (42% vs. 58%) was equivalent to that of the last essay of Group B (44% vs. 56%).

Table 1

*Types of Peer Feedback: Compliments or Critical Comments*

Types of Essays	Group A						Group B							
	Descriptive (N=12)		Narrative (N=11)		Comparison-Contrast (N=12)		Cause-Effect (N=12)		Opinion (N=11)		Group A Total		Opinion (N=17)	
Compliments (%)	8	42%	7	37%	9	38%	8	35%	4	21%	36	35%	15	44%
Critical Comments (%)	11	58%	12	63%	15	63%	15	65%	15	79%	68	65%	19	56%
Total # of Comments	19		19		24		23		19		104		34	
Points / Comments	1.58		1.73		2.00		1.92		1.58		1.79		2.00	

### 4.2 Characteristics of Feedback Comments in Light of Min's Four-Step Criteria

Table 2 shows the breakdown of critical comments using Min's (2005) four-step criteria: clarifying writers' intentions [clarification], identifying problems [problem], explaining the nature of the problems

[explanation], and giving suggestions [suggestion]. It also shows the ratio of 1-step comments to 2/3-step comments following Min (2005).

In terms of the breakdown of critical comments, Group A participants were found to receive suggestions [suggestion] most often (52% on average), followed by problem identification [problem] (36% on average). There were only a few clarification comments [clarification] (2%), and the frequency of explanations of the identified problem was not high (10%). As for Group B participants, the top two types of comments turned out to be the same as those in Group A, but their frequency was reversed ([problem] 52%; [suggestion] 29%). It is also worth noting that, in Group A, the ratio of suggestions [suggestion] out of all four types remained over 60% in the first three essays (descriptive, narrative, and comparison-contrast), whereas that of cause-effect essays dropped significantly to 42% and of opinion essays further down to 32%. Interestingly, when Group A and B data of opinion essays only were compared, the breakdown of reviewers' comments types was quite similar.

As for the 1-step comments to 2/3-step comments ratio, in Group A, it was around 5 in the first two essays, reached the highest point of 15 in the third essay, and then went down to 1.0 and 1.8, respectively (the fourth and fifth essays). This was driven by the sudden increase of 2/3-step comments in the last two essays. On the other hand, the fifth essay's ratio in Group B was as high as 8.5. This implies that, in Group A, an increase in the comments involving more than one step took place after the middle of the semester, but 1-step comments were overwhelmingly dominant in Group B, which had only one peer-review opportunity.

Table 2

*Breakdown of Critical Comments and Ratio of 1 Step vs 2-3 Step Comments from Min (2005)*

Types of Essays	Group A						Group B	
	Descriptive (N=12)	Narrative (N=11)	Comparison-Contrast (N=12)	Cause-Effect (N=12)	Opinion (N=11)	Group A Total	Opinion (N=17)	
Clarification (%)	1 7%	0 0%	0 0%	1 4%	0 0%	2 2%	1 5%	
Problem (%)	2 14%	4 29%	5 28%	11 46%	10 53%	32 36%	11 52%	
Explanation (%)	2 14%	1 7%	1 6%	2 8%	3 16%	9 10%	3 14%	
Suggestion (%)	9 64%	9 64%	12 67%	10 42%	6 32%	46 52%	6 29%	
Total # of CC	14	14	18	24	19	89	21	
CC / Participant	1.17	1.27	1.50	2.00	1.73	1.53	1.24	
# of 1-Step Comments	9	10	15	7	9	50	17	
# of 2-3 Step Comments	2	2	1	7	5	17	2	
1 Step / 2-3 Steps	4.50	5.00	15.00	1.00	1.80	2.94	8.50	

Note. CC refers to critical comments

#### 4.3 Foci of Critical Peer Comments: Global or Local

Table 3 displays the breakdown of the foci of critical peer comments. It was found that there were more critical comments on global issues than local issues. Although the critical comments on global points were predominant in all essays, the breakdown reveals that the foci differed by assignments. For example, many comments on content were found in the first three essays in Group A, and the comments regarding rhetoric were very high in the last essay. Besides, there were many comments on local issues in the three essays in the middle of the term. From the comparison of feedback comments on the fifth essays between Group A and B, Group A participants were found to generate more global comments (93%) than their Group B counterparts (68%).

Table 3

*Foci of Critical Peer Comments: Global or Local*

Types of Essays	Group A						Group B	
	Descriptive (N=12)	Narrative (N=11)	Comparison-Contrast (N=12)	Cause-Effect (N=12)	Opinion (N=11)	Group A Total (N=58)	Opinion (N=17)	
Content	7 64%	6 50%	6 40%	2 13%	2 13%	23 34%	5 26%	
Rhetoric	3 27%	2 17%	1 7%	3 20%	9 60%	18 26%	7 37%	
Global Grammar & Word Choice	0 0%	0 0%	2 13%	2 13%	3 20%	7 10%	1 5%	
Local Grammar & Word Choice	1 9%	4 33%	6 40%	6 40%	1 7%	18 26%	3 16%	
Style	0 0%	0 0%	0 0%	2 13%	0 0%	2 3%	3 16%	
Total Critical Comments	11	12	15	15	15	68	19	
Global Points	10 91%	8 67%	9 60%	7 47%	14 93%	48 71%	13 68%	
Local Points	1 9%	4 33%	6 40%	8 53%	1 7%	20 29%	6 32%	

Note: Content refers to topic development and clarification of content.

Rhetoric refers to organization patterns and effectiveness of topic sentences, thesis statements, concluding paragraphs, and concluding sentences.

Global Grammar and Word Choice refers to the points that interfere or change the intended meaning.

Local Grammar and Word Choice refers to the points that do not interfere with or change the intended meaning.

Style refers to the points related to punctuation, formatting, and references.

#### 4.4 Ratio of Critical Peer Comments Incorporated into Writers' Revisions

As can be seen in Table 4, more than 80% of the critical peer feedback comments were found to be incorporated into the writers' revisions in all essays in both Group A and B. There was no consistent trend in the changes of ratio in Group A; the rate fluctuated between 80% and 100%. The average ratio was 85% in Group A and 89% in Group B.

Table 4

*Percentage of Critical Comments Incorporated in Writers' Revisions*

Types of Essays	Group A					Group B	
	Descriptive (N=12)	Narrative (N=11)	Comparison-Contrast (N=12)	Cause-Effect (N=12)	Opinion (N=11)	Group A Total (N=58)	Opinion (N=17)
# of Critical Comments (CC)	11	12	15	15	15	68	19
# of Revisions based on CC	9	12	12	13	12	58	17
% of CC Incorporated in Revisions	82%	100%	80%	87%	80%	85%	89%

#### 4.4 Sources of Writers' Revisions

Table 5 summarizes the sources of revisions made by the writers. In both groups, feedback from peers accounted for more than half of the revisions (average of Group A 57%; Group B 52%), followed by the changes made by themselves (average of Group A 37%; Group B 39%). Interestingly, some of the revisions—although the ratio is not high (average of Group A 5%; Group B 3%)—were triggered by reading their partners' drafts, which tended to occur more often in the participants' first and second peer review opportunities in Group A. There were a few specific references to class content (one in Group A; two in Group B), all of which took place in the participants' first peer review opportunities in both groups.

Table 5

*Sources of Writers' Revisions*

Types of Essays	Group A						Group B
	Descriptive (N=12)	Narrative (N=11)	Comparison-Contrast (N=12)	Cause-Effect (N=12)	Opinion (N=11)	Group A Total (N=58)	Opinion (N=17)
Feedback from Peers	9 53%	12 48%	12 60%	13 65%	12 63%	58 57%	17 52%
Reading Partners' Drafts	2 12%	2 8%	0 0%	1 5%	0 0%	5 5%	1 3%
Class Content	1 6%	0 0%	0 0%	0 0%	0 0%	1 1%	2 6%
Writer (Self)	5 29%	11 44%	8 40%	6 30%	7 37%	37 37%	13 39%
Total Revisions	17	25	20	20	19	101	33
Revisions / Participant	1.42	2.27	1.67	1.67	1.73	1.74	1.94

## 5. Discussion

### 5.1 Types of Peer Comments

The finding that the participants received more critical comments than compliments in peer review counters Carson and Nelson (1996) and are in line with the studies in Asian Contexts (e.g., Min, 2005; Yu & Hu, 2017). It is also worth noting that in Group A, which had multiple peer review opportunities, the ratio of critical comments increased over time as the participants became more experienced. It is possible that even though there was no formal and intensive training, the act and experience of peer review itself—together with opportunities of sharing of the content of peer review with the class after each review—served as quasi-training. Teacher feedback with critical comments and compliments, which was given on the final version of each essay, could have influenced the nature and characteristics of peer feedback comments on the subsequent drafts. Another possible explanation is that since abundant pair and group work opportunities were provided in lessons throughout the semester, the students had become familiar with one another and less hesitant to share their honest and candid ideas.

The participants' readiness for peer review could account for the predominance of critical comments over compliments. The education at this university, which emphasizes active learning, conducted in a relatively small class size, could have led students to think critically and give constructive feedback actively. Besides, the checkpoints in the peer review checklist and the teacher's explanation could have played a role.

Group B participants also received more critical comments than praises with smaller differences. Even though they experienced peer review only once, perhaps the interaction among students through pair and group work activities in online lessons could have contributed to having the atmosphere of exchanging ideas frankly. Maybe the fact that they had received feedback from Criterion® on four of the five essays before engaging in the peer review—together with the student discussions associated with it—could have served as the simulation for peer review. As for the smaller difference in ratio between critical comments and compliments compared to Group A, the number of peer review opportunities could be one explanation. They might have been somewhat hesitant to give mostly critical feedback from the very first peer review opportunity. Other factors, such as student character and majors, are worth exploring through more in-depth, qualitative studies.

## 5.2 Reviewers' Comment Types

Classification of the reviewers' comment types, drawing on Min's (2005) study, revealed that critical peer comments were predominantly problem identification [problem] and suggestions [suggestion] in both groups.

It is also interesting to note that the total number of critical comments per participant shows the upward trend in Group A. As they became more experienced in peer review, they may have become better able to spot points for improvement in the writers' drafts within a limited time. Interestingly, in the fourth essay, the contributing factor was the increase in 2/3-step comments instead of 1-step comments. The combination of "[problem] and [explanation]" comments and "[explanation] and [suggestion]" comments increased sharply in the fourth review. It is possible that through multiple peer review experiences and stimulation from peers and class, the way they provide feedback has become more specific and writer-friendly.

A closer look at the comment types in Group A over time reveals unique phenomena. The ratio of [suggestion] on the first three types of essays is roughly equivalent, around 65%. In contrast, the proportion of [suggestion] in the fourth essay's feedback dropped to 42%, and further down to 28% in the fifth essay. The topics of the first three essays were 1) a dish or food (descriptive essays), 2) a memorable experience (narrative essays), and 3) the similarities and differences between the two places or things (comparison-contrast essays). In writing about these topics, the writers could draw on their experience and knowledge, and the reviewers could give concrete suggestions. However, in the last two essays, they had to write the causes or effects of a phenomenon and give an opinion about a controversial issue with a counter-argument and refutation. In these two essays, students were assigned to cite at least one outside source. Altogether, the two essays are more challenging for the writers and the reviewers. Therefore, it is likely that the reviewers could identify problems, such as the lack of counter-arguments and citations, but were not able or confident enough to give concrete suggestions. Indeed, the ratio of suggestions in both groups on opinion essays is quite similar, which implies that both groups' participants might have faced similar challenges. Giving concrete suggestions might have been beyond their capacity, with the time constraint of 20 minutes reviewing the drafts for those with pre-intermediate to intermediate proficiency in English.

Group B, which had only one peer-review occasion, showed a few noticeable differences with Group A: the number of critical comments per participant and the ratio of 1-step comments over 2/3-step comments. The number of critical comments received per writer was smaller in Group B (1.24) than Group A (1.53) in the fifth essay. Moreover, the ratio of 1-step comments over 2/3 step comments was by far larger in Group B (8.50) than in Group A (1.80) in the same fifth essay. These differences could imply that Group B participants produced fewer and somewhat less specific and writer-unfriendly critical comments than Group A participants. Again, the differences in the number of peer-review opportunities could have played a role.

There were a few difficulties in analyzing the data. The first difficulty was in categorizing the comments. It was sometimes hard to determine which category one comment should fall into—problem identification or a suggestion. In these cases, the tone in the reflection text and/or the existence of concrete suggestions were examined. The second difficulty derives from the fact that the participants' reflection text, not the actual feedback comments, was used as data. It is possible that they have simplified the feedback content and omitted minor issues in writing their reflection texts. This means that even when a 1-step



feedback comment was confirmed in the reflection texts, a 2/3-step comment could have actually been provided. This illustrates the limitation of using the reflection texts, and additional methods to increase accuracy need to be sought.

### **5.3. Foci of Critical Peer Comments: Local or Global Issues**

The participants' use of Grammarly could partially explain finding more critical comments on global points than local points before conducting peer review. Using Grammarly could have contributed to making the drafts mostly error-free and enabled the reviewers to focus more on global points. The checklists' use, which led the reviewers to focus on global points, could be another explanation.

As content-related points, the reviewers mostly wrote about what needs to be added or what they wanted to know more. As for the rhetorical issues, the effectiveness of attention-getters and counter-arguments was addressed. The comments on local issues revolved around the use of synonyms or better word choice. In this way, as discussed in Liu and Hansen-Edwards (2018), these feedback comments that centers on content and rhetorical points, or global issues, were effective and helpful for the writers.

### **5.4 Incorporation of Critical Peer Comments into Writers' Revisions**

That the writers in both groups incorporated 80% or more of the reviewers' critical comments could reflect the writers' understanding of, and agreement with, the comments. Indeed, feedback from peers accounted for more than half of the writers' revisions. It could imply that the writers saw value in the peer comments as an essential source of advice.

However, the high incorporation rate should not be accepted at face value due to the possibilities that the writers did not refer to those comments which were incomprehensible or which they chose not to adopt. More simply, the participants might have chosen to write about only vital comments in their reflections. Even with these possibilities and limitations, this high rate could illustrate that fruitful peer review took place.

### **5.5 Sources of Writers' Revisions**

Although the primary source of the writers' revisions turned out to be critical peer comments, there were other sources of revisions for the writers. One important source was the writers themselves. They worked independently to improve their essays by using the resources available and synthesizing what they had learned. This signals that they make their decisions to improve their writing, as Tsui and Ng (2000) maintain.

A few instances of getting the hints by reading the partners' drafts were observed as well. One participant was impressed with her partner's detailed descriptions of a scene, which motivated her to drastically revise her draft. Another noticed that her partner's draft was reader-friendly, contrary to hers. These examples may underscore that peer review benefits the readers (e.g., Lundstrom & Baker, 2009; Sommers, 2013; Tsui & Ng, 2000).

## 6. Conclusion

This study revealed that the participants provided more critical comments than compliments in peer review. It also showed that critical comments were mainly of problem identification and suggestion types and that the participants tended to provide 1-step comments more often than 2/3-step comments. Time-series analyses and the comparison of two treatment groups showed that multiple peer review opportunities led to the increase of the percentage of critical comments and the ratio of 2/3-step comments over 1-step comments. The study also revealed that with these groups of participants in this context, critical comments on the drafts' global points were more often provided than those on local points. Moreover, the study showed the high incorporation rate of the reviewers' critical comments into the writers' revisions. Finally, it laid out that the writers also used their judgment and knowledge to improve their drafts.

This study has several limitations. First, since the participants' reflection texts are used as data, the reported behavior's accuracy is not guaranteed. In other words, the possibility of simplification, omission, exaggeration, underestimation cannot be excluded. Second, the analyses of actual revisions are not included in this study. Comparing the participants' drafts and final essays will help this end, which will most likely be the next research agenda. The third limitation is the lack of statistical analysis, mainly due to the first limitation described above. To obtain an in-depth understanding, conducting statistical analysis by comparing the participants' drafts and the final essays may be necessary. Furthermore, investigating the learners' perceptions of peer review and AWE through group interviews could provide further insights, and triangulation of data will be possible.

## Acknowledgments

This study was supported by JSPS KAKENHI (Grant Number 20K00772). I would also like to express my gratitude to the 29 participants in my writing classes for their hard work, agreeing to use their reflection texts as data, and providing insightful reflections. Finally, I am thankful for two anonymous reviewers whose invaluable comments helped me improve the quality of this paper.

## References

- Berg, E. C. (1999). The effects of trained peer review response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8, 215-241.
- Carson, J. G. & Nelson, G.L. (1996). Chinese students' perceptions of ELS peer response group interaction. *Journal of Second Language Writing*, 5, 1-19.
- Chen, C-F. E. & Cheng, W-Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94-112.
- Educational Testing Service (n.d.) *TOEFL iBT® Test Independent Writing Rubrics*. [https://www.ets.org/s/toefl/pdf/toefl\\_writing\\_rubrics.pdf](https://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf)
- Ferris, D. R. (2018a). Writing in a second language. In J.M. Newton, D.R. Ferris, C.C. Goh, W. Grabe, F.L. Stoller, & L.

- Vandergrift (Eds.). *Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking*. (pp. 75-88). Routledge.
- Ferris, D. R. (2018b). Building a writing curriculum. In J.M. Newton, D.R. Ferris, C.C. Goh, W. Grabe, F.L. Stoller, & L. Vandergrift (Eds.). *Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking*. (pp.89-105). Routledge.
- Ferris, D. R. (2018c). Writing instruction and assessment. In J.M. Newton, D.R. Ferris, C.C. Goh, W. Grabe, F.L. Stoller, & L. Vandergrift (Eds.). *Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking*. (pp.106-122). Routledge.
- Hatakeyama, Y. (2010). Changes made in a Theme Writing class: The importance of students' perceptions. *ICU Language Research Bulletin*, 25, Retrieved July 21, 2020, from <http://lrb.info.icu.ac.jp/home/volume-25-2010>
- Hyland, K. & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. In K. Hyland and F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (2nd ed., pp.1-22). Cambridge University Press.
- Li, Z., Link, S., Ma, H., Yang, H., & Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System*, 44, 66-78.
- Liou, H-C. & Peng, Z-Y. (2009). Training effects on computer-mediated peer review. *System*, 37, 514-525.
- Liu, J. & Hansen Edwards, J. G. (2018). *Peer response in second language writing classrooms*. (D. Belcher & J. Liu Eds.) (2<sup>nd</sup> ed.). The University of Michigan Press.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T.K. Bhatia (Eds.) *Handbook of language acquisition. Vol. 2: Second Language Acquisition* (pp. 413-468). Academic 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. <https://doi.org/10.1016/j.jslw.2008.06.002>
- Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp.15-34). Cambridge University Press.
- Mendonça, C. O. & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769.
- Min, H. T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293-308. <https://doi.org/10.1016/j.system.2004.11.003>
- Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of second language writing*, 15(2), 118-141. <https://doi.org/10.1016/j.jslw.2006.01.003>
- Min, H. T. (2008). Reviewer stances and writer perceptions in EFL peer review training. *English for Specific Purposes*, 27(3), 285-305. <https://doi.org/10.1016/j.esp.2008.02.002>
- Rollinson, P. (2005). Using peer feedback in the ELS writing class. *ELT Journal*, 59, 23-30.
- Savage, A. & Mayer, P. (2012). *Effective academic writing 2* (2nd ed.). Oxford University Press.
- Smith, G. (2019). Introducing a peer review system targeting shared understanding and actionable feedback. *ICU Language Research Bulletin*, 34 (pp.47-53). Retrieved July 21, 2020 from <http://lrb.info.icu.ac.jp/home/lrb-volume-34-2019-4>
- Srichanyachon, N. (2011). A comparative study of three revision methods in EFL writing. *Journal of College Teaching & Learning*, 8(9), 1-8.
- Sommers, N. (2013). *Responding to student writers*. Bedford/St. Martin's.
- Tsui, A. B. M. & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*,

9(2), 147-170.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Warschauer, M. & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 1-24.

Yu, S. & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing*, 33, 25-35. <http://doi.org/10.1016/j.asw.2017.03.004>