

# Language Input Effects on L2 Composition Peer review Feedback

**Clay Williams\***

*Akita International University*

## Abstract

---

To date, very little research on L2 peer review of L2 essay composition has focused on the issue of whether the choice of L1 or L2 use in the peer review process facilitates or hinders students' transmission and reception of productive commentary on their essay drafts. This study was designed to ascertain the impact of language choice in written peer review sessions on both the types of commentary made and essay authors' propensity to incorporate peer commentary into subsequent drafts. Data was collected from essay assignments in an English writing course at an English-medium university in Japan using written peer reviews conducted in both Japanese and English in a cross-balanced design. The data were analyzed to measure the impact of the language used on the number and types of comments made, as well as to ascertain the relative impact of peer commentary on the subsequent revisions. The study found sizeable differences according to the language a peer review was conducted in, suggesting that L2 written peer reviews may be more beneficial at identifying/rectifying paragraph-level and structural issues, whereas L1 peer review was slightly better for correcting rhetorical and logical issues.

---

**Keywords:** L2 writing; peer review; essay writing pedagogy; L1/L2 effects

---

## Introduction

Given the increased prominence of English communication in the East Asian educational domain in recent years, it is hardly surprising that more and more universities are offering courses on advanced English essay writing. Increasing competition in the job market has increased the need for demonstrated English skills when applying for employment (Reed, 2002). Recent years have witnessed a sudden growth in the number of tertiary-level programs and whole institutions wherein English is the lingua-franca for all instruction. While a few institutions have long and illustrious histories (e.g., International Christian University in Japan), the last decade has seen the birth of such institutions as Akita International University (Japan), University of Nottingham Ningbo (China), Tan Tao University (Vietnam), and Xing Wei College (China), all offering full degree programs taught entirely in English. In addition to entirely English-medium institutions, other established regional universities have begun offering specific degree programs taught largely or entirely in English, such as Waseda University (Japan), Ho Chi Minh International University (Vietnam), and Asian Pacific International University (Thailand). Outside of the Asia-Pacific region, likewise, there is a global trend of increased number and prominence of English-medium programs (Dearden, 2014). According to an ICEF Monitor report (Trend alert, 2012), English is already the lingua franca of many Middle Eastern universities, there are over 4500 courses being taught in English across continental Europe, and the number of English-medium courses/programs is on the rise in Africa and South America, as well.

\* Email: [williams@aiu.ac.jp](mailto:williams@aiu.ac.jp). Tel.: +81-18-886-5820. Address: English Language Teaching Practices, Graduate School of Global Communication and Language, Akita International University, Yuwa, Akita-city 010-1292, JAPAN

This steep rise in English-medium content course availability has had a deep impact on both the type and level of English language preparatory courses being offered, and the area of academic writing has been no exception to this phenomenon. Whereas, not very many years ago, most L2 English writing programs functioned at a low level, mostly focusing on issues of grammatical accuracy, these new English-medium programs require writing courses which prepare students to write for English-medium content courses across a broad array of academic disciplines. This has effectively raised the standards for writing course content and curricular goals in East Asia, as has the increased English proficiency of the students enrolled in such programs. These curricular changes have resulted in many of these preparatory courses consciously modelling themselves after the sorts of writing programs that L2 English international students would encounter in universities in L1 English countries. This gives rise to some unique issues, as most prior research in L2 English writing examining high-proficiency writing has come from the ESL context, and not an EFL context. The sudden proliferation of advanced-level writing programs in EFL contexts—wherein students come from the same L1 and national background—enables an opportunity for experimentation in pedagogical design and efficacy that would not be possible in lower proficiency level nor with students with mixed L1s and nationalities.

This paper investigates peer review methodology in light of the new classroom dynamics created by the proliferation of high-level (i.e., focusing on production of academic and/or professional essay writing skills as opposed to a more general focus on sentence and paragraph-level vocabulary and grammar skills) L2-English writing courses in East Asia. As many academic writing courses make use of written peer review as part of the drafting/revision process for essay writing, the question arises as to whether there would be any immediate advantages, disadvantages, or qualitative differences derived from conducting peer reviews in the students' L1 or L2.

## **Literature Review**

### **Peer Review**

One of the fundamental, main-stay pedagogical approaches to most writing composition programs and courses is the concept of peer review. Originally developed for use in L1 composition teaching, and fitting squarely into the cooperative learning strategies domain, this has been a time-tested teaching technique, and its positive effects on L1 composition student performance have been reported in studies too voluminous to fully catalogue here (e.g., Beaven, 1977; Gere & Abbot, 1985). In the L2 composition domain, as well, it has a distinguished track-record, and many researchers have attested to its effectiveness in application to L2 writing courses (e.g., Allison & Ng, 1992; Arndt, 1993; Keh, 1990; Lockhart & Ng, 1993; Tsui & Ng, 2000). The perceived benefits of peer review are many: it is thought to help students in developing the ability to appropriately analyze and revise their own writing (Zhang, 1995); it may increase learner participation (Mendonça & Johnson, 1994); and it is supposed that peer feedback is less threatening than teacher feedback (Jacobs, Curtis, Braine, & Huang, 1998). However, these views have not gone without challenge. Nelson and Murphy (1993) observed significantly more instances of students being overly focused on surface problems at the expense of deeper textual issues, as well as greater reluctance to implement peer commentary on the part of L2 writers as opposed to L1 writers. Their status as second language speakers is assumed to make students more reticent towards accepting classmate feedback. Additionally, the tendency towards teacher-centered classroom practices and general deference extended to teachers in East Asia can make students wary of accepting peer commentary as authoritative, or even as helpful (Ferris, 2003; Nelson & Murphy, 1993). Wu (2006) found in a study on Chinese L1 students of English that teacher feedback had a measurably greater impact on student writing performance than did peer feedback.

There has been much investigation delving into various aspects of peer review in the L2-English context which has enlightened our understanding of the strengths and potential limitations of peer review activities in L2-English composition courses. For instance, Liao and Lo (2012) found that the relative quality and types of peer review commentary was largely dependent upon students' L2 proficiency levels. While both high and low-level proficiency learners' comments were dominantly used to identify problems, higher-level proficiency reviewers' comments provided significantly more detail in both discussion of the problem and in suggesting means of

improvement. Salih (2013) conducted post-peer-review debriefing sessions and interviews to compare student expectations for peer review comments with the actual patterns emerging from peer review activities. They found that, despite the writers' expectations that peer reviewers would focus on grammatical correction, in reality, the dominant type of comments delivered were regarding essay structure, and reviewers recounted their primary focus was in maintaining the clarity of the feedback. Yu and Lee (2015) investigated the factors which determine individual participation in group peer feedback activities, and they found that the primary determining factor was student motivation, which itself was affected directly by sociocultural context. Some of the studies on peer review in L2 contexts have turned up less definitive results, but are nevertheless valuable in shaping our understanding of the pedagogical value of the activity. For instance, LoCastro (2000), investigating whether peer review would follow the dominant discursive norms of the L1 and L2, found the results inconclusive; however, the study also revealed that there seemed to be a clear effect of instruction in dropping L1 discursive behaviors.

One of the more extensively-covered aspects of peer review in the L2-English classroom domain has been the issue of the extent to which recent technological advancements could significantly improve students' performance and acceptance of classmates' suggestions. Much of the research has found computer-mediated learning to be an effective means of L2 study, as it can lower affective variables and enhance motivation (e.g., Coniam & Wong, 2004; Strenski, Feagin, & Singer, 2005), there has naturally been some curiosity as to whether this would impact the efficacy of peer review in L2 writing classes. Some researchers (e.g., Crank, 2002; Liu & Sadler, 2003) have found significant improvements to peer review performance when peer review is conducted via computer platforms in asynchronous computer-mediated communication; whether or not those effects are permanent has come under question. Xu (2007) found that performance boosts fade over time, and suggests that the temporary increase in productivity simply reflects student curiosity and excitement over the new technology, and diminishes as students become accustomed to its use.

### **The Issue of Language Medium in Peer Review**

The motivating question for this study was whether the language in which written peer essay review was conducted would prompt students to write better commentary and/or be more accepting of peer suggestions (as evidenced by inclusion in subsequent drafts). As basic as the question appears on its face, it appears that there has been almost no research to date touching upon the issue. While it would be impossible to state conclusively why something does not receive broader treatment in the academic community, the author's suspicion is that the issue has been ignored because, at least in East Asia, the issue was mostly moot until relatively recently. As teaching practices tend to be pragmatic at their core, if a group of students was not very skilled in the L2, it would be unsurprising for an instructor to allow students to use their L1 for peer review purposes, so as to enable them to fully explain their thoughts on each other's writing. Likewise, if students had the requisite skills to conduct a meaningful peer review in the L2, many teachers might be hesitant to allow students to conduct the review in their L1, and thus to lose such a precious opportunity to use and expand upon L2 writing skills. In the ESL context, likewise, there often exists the need to conduct peer review in L2 English regardless of student proficiency because classes are frequently composed of mixed groups of students from various language/cultural backgrounds, with English being the only common vehicle of communication. It is only in the new, emerging context of the English-medium international university, wherein monolithic blocks of students from the same L1-background study through the medium of L2-English, that the question becomes valuable. These students unquestionably have the skills to conduct L2 peer review; however, as they use English day-in and day-out during their university experience, there would likely be less compulsion to make every minute of class time count. Classes are also overwhelmingly populated by students of similar cultural and linguistic backgrounds, so use of L1 for peer review becomes a realistic option. As both options seem equally practical and valid, it becomes necessary to consider whether or not there would be any qualitative differences in feedback and revision performance between written peer review conducted in students' L1 or L2.

While prevailing practices have emphasized use of L2 in written peer review sessions, there are reasons to suspect that L1 peer commentary may in some ways be more effective towards producing a better final essay. Studies such as that by Nelson and Murphy (1993), as well as Wu (2006), indicate some level of resistance by L2

writers to take the critiques of other L2 writers very seriously. While writers can understandably be concerned about the limitations of subject expertise and knowledge by peer reviewers, these negative perceptions are sometimes further exacerbated by reviewers' lack of L2 skill constraining their ability to articulate legitimate points about the writings they are tasked with reviewing. Simply put, if peer reviewers respond in the shared L1, they may be able to express their concerns about the essay's structure, coherence, logic, etc. with considerably greater coherence and detail than is allowed by the L2. This greater specificity by reviewers could result in a stronger argument for change, thereby more likely being accepted into subsequent revisions by the original author.

There appears to have been little study of language effects on peer review. One exception was Huang's (1996) study of oral peer review in L2 English classes. Huang found an asymmetry in the foci of students' commentary on each other's papers, depending on the language used to moderate the discussions. Peer feedback delivered in the students' L1 (Mandarin Chinese) were more specific, and focused mostly on issues of language usage. By contrast, while groups delivering feedback in the L2 (English) were more general in their commentary, they nevertheless managed to discuss a broader range of issues (e.g., language use, essay reasoning, and rhetorical strategies). Both languages displayed positive effects on student feedback sessions: the use of L1 was perceived as being more effective in eliciting deeper commentary and appeals to implement peer feedback; however, the L2 sessions, while not as focused, nevertheless elicited more communal support among students. Since this study focused only on *oral feedback*, it is still unknown whether language choice in *written feedback* elicited via peer review would lead to differences either in the types of comments made or in its persuasiveness (as measured by prompting the authors to change their papers in subsequent drafts).

### Research Questions

The study described herein was designed to ascertain whether advanced students of English composition would show any significant differences in peer review performance depending on whether the peer reviews were conducted in their L1 or L2. The specific questions which this study sought to answer are:

- 1) What is the relationship between the language that peer feedback is delivered in and the type of feedback delivered?
- 2) In individual feedback categories, what is the relationship between the language that peer feedback is delivered in and the degree to which authors are willing to accept comments and integrate them into subsequent revisions?
- 3) In holistic categories (based upon Huang, 1996) of mechanical vs. rhetorical commentary, is there a relationship between the language used for peer review and the types of comments that authors are more willing to accept and integrate into subsequent revisions?

### Methods

#### Participants

The study was conducted at a small, English-medium international university in northern Japan. At this university, all students take 1-3 semesters of foundational intensive English coursework before being mainstreamed into regular content (degree-seeking) coursework. Upon mainstreaming, the students are still required to take two English composition-writing courses as graduation requirements. These courses are usually taken within the first two semesters after finishing the foundations program and beginning their degree programs. The participants in this study were 39 students in two different sections (20 in one class section, and 19 in the other) of the first required composition-writing course. All students had Japanese as their L1, and all had been enrolled at the university for 1-3 semesters prior to taking the course (most were in their second semester). At the time, TOEFL IPT scores of at least 500 were required in order to exit the foundation classes and begin regular coursework; however, most students had already exceeded that minimum score before embarking on the

foundations coursework, and exit criteria have since been modified. While individual TOEFL scores are protected by law, and thus inaccessible to the researcher, at the time of the study, the average TOEFL IPT score in such writing classes would have been around 530, with a range from a low of 500 (as the minimum score for entry) to a high of 650. The students had all taken at least one academic writing course focused on production of multi-paragraph essays in the preceding semester, and as part of that course, had already gained some experience with peer review procedures.

### **Data Collection**

The course was designed as an introduction to analytical, academic essay writing. For each assignment, students would read a handful of essays and/or short stories (grouped by a theme) and develop a unique interpretation through literary analysis. They would then write a 4-5-page essay arguing/defending their interpretation. Each essay-drafting period would involve two opportunities for peer review. The peer review sessions were conducted in-class, and were facilitated by a 16-question response sheet (see Appendix 1) asking a mix of short- and long-answer questions focusing on paragraph-level and essay-level critique and response. All peer review during the first essay was conducted in English, and was preceded by explanations, examples, and group work designed to get students familiar with both the form and the peer review dynamic, as well as to firmly define the expectations and boundaries of acceptable and unacceptable peer commentary.<sup>1</sup>

The experiment started from the second required essay, and lasted through the third and final essay (for a total of two essays and four peer review sessions). The essay response sheet was translated by a bilingual native speaker of Japanese, and from the first peer review session for paper #2, class “A” was given the Japanese peer review sheet (see Appendix 2) and was encouraged to respond to each other in Japanese, whereas class “B” continued to peer review in English. This was held constant through both peer review sessions for paper #2. For the third paper, the languages used for the peer review in each class were switched, with class “A” conducting both peer reviews in English, and class “B” conducting both peer reviews in Japanese, thus counterbalancing the presentation of English and Japanese peer review. The course instructor explained to both classes that this peer review commentary was being used in a study to determine what (if any) impact the language of the peer review would have on their performance, and encouraged students to respond in the appropriate language; students were not given any additional information on how the data collected would be analyzed. All peer review sheets, corresponding draft essays, and final essays were handed in at the end of each assignment period. Photocopies were made to enable final grades and instructor feedback to be distributed to the students while the researcher retained the original forms. Students were told that, upon request, they could reclaim their original copies, though none followed up on this offer.

### **Analytical Procedure**

After the end of the semester, all documents were analyzed by a team of two bilingual L1-Japanese/L2-English graduate students according to a prepared rubric. First, they focused on identifying all peer review comments according to the types listed in Table 1. The first 5 categories focus on word, sentence, and paragraph-level issues of structure. They include mechanical issues (e.g., highlighting a misspelled word or incorrect grammar), introduction/thesis (e.g., one of the most common comments regarded theses which were either overly broad or failed to make an argument), topic statements (e.g., pointing out incorrectly placed, missing, or topic statements irrelevant to the following paragraph content), body paragraphs (e.g., unclear structure or examples that do not clearly support the topic), and conclusions (e.g., failure to restate one’s position). Categories 6-9 focus on more global issues of coherence and appropriateness. These include phrasing issues (e.g., issues of word choice and degree of formality/informality in writing), logic issues (e.g., non-sequitur arguments), persuasiveness issues (i.e., an inability to convince the audience of one’s point), and global comments (i.e., the reviewer’s overall view of the paper). As such, while the feedback types are subcategorized more extensively, the first 5 categories can be considered to correspond directly to Huang’s (1996) category of language/accuracy focus, and categories 6-9 correspond to Huang’s second (broader) category of rhetorical/logical focus. Category 10 measures the extent to which the peer review exercise was used for purely social reasons (e.g., “Hi! How’s it going?”), and as such,

stands largely outside of either category. The two student evaluators worked together and were required to agree on classifications. In the event of disagreements, the author would make the tie-breaking vote. Next, all instances of each category were cataloged and numbered according to language used (i.e., Japanese or English).

An important caveat needs to be mentioned before continuing to the results. Seven students were dropped from the study and results from their work were not factored into the analysis. Six of the seven were rejected because they had missed one or more peer review sessions and/or failed to hand in one of the required papers. Results are only tabulated using students who attended all peer review sessions and turned in all required work. The other student responded almost entirely in the wrong language during one of the peer review sessions, and was thus omitted from the analysis.

Table 1  
*Individual Feedback Categories in Peer Review*

| Word/ Sentence/ Paragraph-level Analysis | Global Issues<br>(Coherence/Appropriateness) | Social Commentary     |
|--|--|-----------------------|
| 1) Spelling & Grammar                    | 6) Phrasing Issues                           | 10) Personal Comments |
| 2) Introduction/thesis                   | 7) Logic Issues                              |                       |
| 3) Topic statements (in-body paragraphs) | 8) Persuasiveness Issues                     |                       |
| 4) Body Paragraph                        | 9) Global Comments                           |                       |
| 5) Concluding paragraph structure        |  |                       |

Once the peer review comments had been analyzed and categorized, the data was used to address the three motivating research questions. The first question—regarding the relationship between the language that peer feedback is delivered in and the type of feedback delivered—could be answered in two ways: either (1) by counting the number of subjects receiving explicit feedback according to each of the two categories used by Huang (1996) and then comparing (via a 2x2 Chi-square test) the number in each feedback category according to whether the feedback was delivered in Japanese or in English; or, (2) by counting the total number of instances of peer feedback in each of the two categories, and comparing by language (again, via a 2x2 Chi-square test). Both methods were employed here. When analyzing by ‘subject,’ it should be noted that it would be possible for one subject to populate more than one cell and thereby violate an assumption of the test, but subsequently, the unit of analysis was the instance of feedback, so the Chi-square test was still employed noting this limitation.

Research question number two—measuring any association between the language of peer review and the likelihood of individual categories of feedback being accepted into subsequent revisions—required an analysis of the draft vs the final form of the assignment (handed in for grading) to determine if the peer review comments had been accepted and used. With this information, the data could be analyzed via 2x2 Chi-square tests for each category individually.

In responding to the third question—concerning the relationship between the language of peer feedback and types of peer commentary which authors are more likely to accept and integrate into subsequent revisions—the data collected recording the number of instances of peer commentary and the number of comments which were implemented into subsequent paper drafts was simplified into the two broad categories of mechanical comments (sentence or paragraph-level) and logical/rhetorical comments (essay-level) according to Huang (1996). Once the data was compiled, a comparison of the degree of uptake in each category could be made in each language individually, and the categories of feedback could also be directly compared (individually) across languages.

## Results

### Research Question 1: Language Choice and Type of Feedback

The total number of subjects receiving each manner of feedback is shown in Table 2. Frequencies of mechanical and rhetorical feedback rendered in Japanese and English were compared via a Chi-squared test ( $\alpha = 0.05$ ). There was no significant difference in the numbers of subjects receiving each type of feedback:  $X^2(1, N = 156) = 0.06, p < 0.80$ .

Table 2

*Numbers of Students Receiving Feedback by Type and Language*

| Type of feedback   | English: Count # of Students receiving commentary by type | Japanese: Count # of Students receiving commentary by type |
|--|---|--|
| 1) Spelling & Grammar  | 11  | 17   |
| 2) Intro / thesis  | 12  | 9  |
| 3) Topic statements  | 11  | 10   |
| 4) Body paragraph structure  | 11  | 10   |
| 5) Concluding paragraph structure                                  | 10  | 9  |
| TOTAL: Word/Sentence/Paragraph-level Analysis (categories 1-5)     | 55  | 55   |
| 6) Phrasing issues   | 9   | 8  |
| 7) Logic issues  | 6   | 6  |
| 8) Persuasiveness issues   | 4   | 5  |
| 9) Global comments   | 5   | 3  |
| Total: Global Issues (Coherence/ Appropriateness) (categories 6-9) | 24  | 22   |
| 10) Social Commentary  |   |  |

The total number of incidents of each manner of feedback in each language is shown in Table 3. An analysis of the association between language used in the peer review and the feedback type according to Huang's (1996) two categories of word/sentence/paragraph-level analyses and global issues yielded no significant differences  $X^2(1, N = 234) = 0.39, p = 0.53$ , so while comments were somewhat more voluminous in Japanese than in English, the relative proportion of mechanical to rhetorical comments did not vary according to the language used in the peer review.

It is also worth noting that Japanese (L1) critiques of all types tended to be longer and more detailed. As a small example of the difference between the two, we can look at two examples, by the same reviewer, taken from the "logic issues" category. In a review written in English, the student writes simply, "Body paragraphs should be more organized," with no more detail. Reviewing a different paper in Japanese, the reviewer makes a relatable observation, but this time expands into significantly more detail, "[Paragraph order should be reorganized according to your 2 categories: communication and information. Now, the body paragraphs are not connected well]."<sup>2</sup> This kind of specificity and detailed critique allows students to be much more confident in the accuracy and validity of peer commentary, thus possibly explaining why Japanese holistic comments were so much more widely implemented into the essay drafts.

Table 3  
*Numbers of Instances of Feedback by Type and Language*

| Type of feedback  | Number of instances of English feedback | Number of instances of Japanese feedback |
|---|---|--|
| 1) Spelling & Grammar   | 13                                      | 23                                       |
| 2) Intro / thesis   | 16                                      | 18                                       |
| 3) Topic statements   | 17                                      | 21                                       |
| 4) Body paragraph structure                                       | 14                                      | 19                                       |
| 5) Concluding paragraph structure                                 | 12                                      | 17                                       |
| TOTAL: Word/Sentence/Paragraph-level Analysis (categories 1-5)    | 72                                      | 98                                       |
| 6) Phrasing issues  | 12                                      | 15                                       |
| 7) Logic issues   | 8                                       | 9  |
| 8) Persuasiveness issues  | 4                                       | 6  |
| 9) Global comments  | 6                                       | 4  |
| Total: Global Issues (Coherence/Appropriateness) (categories 6-9) | 30                                      | 34                                       |
| 10) Social Commentary   | 0                                       | 4  |

The topic of personal comments also bears further discussion. While the use of peer review activities for personal interaction falls outside the scope of this study, and the category was simply used to classify commentary that was social in nature and therefore did not respond directly to the paper being reviewed, I still found it rather curious that there were no such comments made in L2 English. While it was somewhat predictable that L1 would lend itself more to socializing (assuming that students would normally speak to one another in Japanese outside of class), it must be noted that these students are all extremely proficient in English by Japanese academic standards, and make social use of English on a daily basis at the school (where roughly 20% of the student body is composed of international students). The absence of any sort of personal comments made during English-language peer reviews suggests that the students were subconsciously treating English as a transactional medium, and interaction was reserved for Japanese. While the four examples garnered are hardly enough to base any firm conclusions on, the use of L1 and L2 for interactive and social purposes in written peer review activities would be an issue well worth devoting attention to.

### **Research Question 2: Acceptance of Comments for Integration in Later Revisions vs. Language of Feedback in Individual Feedback Categories**

In analyzing comments to determine whether or not the essay authors incorporated them into their subsequent essay drafts, more differences start to emerge. We can see the raw numbers for feedback integration according to whether feedback was delivered in L1 or L2 in Table 4. Direct analysis of the degree of association between L1 or L2 use in peer review and the subsequent degree of integration of peer review comments into the final paper for each of the individual subtypes of feedback (via isolating each horizontal line on Table 4 and using a 2x2 Chi-square test) showed several interesting trends. Comments regarding topic statements were significantly more likely to be implemented when feedback was delivered in English:  $X^2(1, N = 38) = 3.75, p = 0.05 (\Phi=0.31,$

medium effect size); as were comments about essay conclusions:  $X^2(1, N = 29) = 5.15, p = 0.02$  ( $\Phi=0.42$ , medium effect size). Comments regarding body paragraph structure delivered in English approached significance:  $X^2(1, N = 33) = 2.95, p < 0.09$  ( $\Phi=0.29$ , small effect size). Peer feedback delivered in Japanese, however, yielded rates of incorporation in subsequent drafts approaching significant difference when comments were about phrasing issues:  $X^2(1, N = 27) = 2.79, p < 0.10$  ( $\Phi=0.32$ , medium effect size); and logic issues:  $X^2(1, N = 17) = 2.84, p = 0.09$  ( $\Phi=0.40$ , medium effect size). The analysis of personal comments was suspended as 3 of the 4 comments were social in nature, and did not require any sort of decision which would yield a measureable change in the paper. Other categories failed to produce any significant difference between English and Japanese feedback (Spelling & Grammar:  $X^2(1, N = 36) = 1.00, p < 0.32$ ; Intro/thesis:  $X^2(1, N = 34) = 0.17, p < 0.68$ ; Persuasiveness:  $X^2(1, N = 10) = 0.27, p < 0.60$ ; Global:  $X^2(1, N = 10) = 1.67, p < 0.20$ ).

Table 4

*Acceptance/Integration of Feedback into Subsequent Draft by Type and Language*

| Type of feedback                  | English feedback:<br>Integrated | English feedback:<br>NOT Integrated | Japanese feedback:<br>Integrated | Japanese feedback:<br>NOT Integrated |
|-----------------------------------|---------------------------------|-------------------------------------|----------------------------------|--------------------------------------|
| 1) Spelling & Grammar             | 11                              | 2                                   | 16                               | 7                                    |
| 2) Intro / Thesis                 | 10                              | 6                                   | 10                               | 8                                    |
| 3) Topic statements               | 14*                             | 3                                   | 11                               | 10                                   |
| 4) Body paragraph structure       | 12**                            | 2                                   | 11                               | 8                                    |
| 5) Concluding paragraph structure | 10*                             | 2                                   | 7                                | 10                                   |
| 6) Phrasing issues                | 7                               | 5                                   | 13**                             | 2                                    |
| 7) Logic issues                   | 3                               | 5                                   | 7**                              | 2                                    |
| 8) Persuasiveness issues          | 2                               | 2                                   | 4                                | 2                                    |
| 9) Global comments                | 2                               | 4                                   | 3                                | 1                                    |
| 10) Social Commentary             | 0                               | 0                                   | 1 (**+3)                         | 0                                    |

*Notes.*

\*statistically significant difference

\*\*approaching significance

\*\*\*3 of the 4 personal comments did not involve any decisions on the paper itself

**Research Question 3: Relationship Between Language of Peer Review and Types of Commentary Integrated into Subsequent Revisions**

By simplifying the feedback categories—word/sentence/paragraph-level analyses and global issues, more patterns emerge. Table 5 provides the total number of integrated and non-integrated comments in each category according to the language of peer review. Direct comparison between language of peer review feedback and degree of integration into the final paper across the board (i.e., adding together comments from both categories vs uptake analyzed according to language of peer review) yielded no significant results:  $X^2(1, N = 234) = 1.41, p = 0.49$ . However, an analysis of the degree of uptake according to type of comments (i.e., mechanical vs rhetorical) in English peer reviews revealed that comments and recommendations about mechanical issues were significantly more likely to be integrated into the subsequent version of the paper than were comments regarding

rhetoric:  $X^2(1, N = 102) = 10.63, p = 0.005$  ( $\Phi=0.32$ , medium effect size). Japanese peer review yielded a significantly higher proportion of uptake of comments concerning rhetorical issues:  $X^2(1, N = 132) = 5.82, p = 0.05$  ( $\Phi=0.21$ , small effect size). Direct comparison of the degree of uptake of comments concerning mechanical issues according to the language of peer review revealed a significantly higher degree of uptake when the comments were delivered in English:  $X^2(1, N = 170) = 9.8, p = 0.007$  ( $\Phi=0.24$ , small effect size), whereas comparison of uptake of comments about rhetorical issues showed a significant advantage for comments made in Japanese:  $X^2(1, N = 64) = 7.43, p = 0.02$  ( $\Phi=0.34$ , medium effect size).

Table 5

*Acceptance/Integration of Feedback into Subsequent Draft by Simplified Type and Language*

| Type of feedback  | English feedback:<br>Integrated | English feedback:<br>NOT Integrated | Japanese feedback:<br>Integrated | Japanese feedback:<br>NOT Integrated |
|---|---------------------------------|-------------------------------------|----------------------------------|--------------------------------------|
| Word/Sentence/Paragraph-level Analysis (categories 1-5)       | 57                              | 15                                  | 55                               | 43                                   |
| Global Issues (Coherence/<br>Appropriateness (categories 6-9) | 14                              | 16                                  | 27                               | 7                                    |

### Discussion

There are some useful discoveries from the experiment which may help elucidate how students conduct and respond to written peer feedback according to whether they use L1 or L2. Comparing the instances of feedback, one does find some support for Huang's (1996) study of oral feedback, as feedback delivered in the students' L1 did produce more comments regarding word, sentence, and paragraph-level issues, which could be classified in Huang's terminology as "language issues." This type of focus can also be used to explain the near-significant (i.e.,  $p < 0.10$ ) difference in the number of comments regarding the structure of the conclusion paragraph.

However, it is when we start comparing the relative willingness of students to incorporate peer feedback that we start to see a definite trend. Peer feedback delivered in L2 regarding word, sentence, and paragraph-level issues seemed more likely to be accepted and implemented into the next draft, as evidenced by the significantly higher rate of comment incorporation for comments in L2-English regarding topic statements, body paragraph structure, and conclusion paragraphs. By contrast, comments regarding phrasing issues or logic issues were significantly more likely to be accepted if delivered in L1-Japanese—even though the number of comments made in English and Japanese were quite similar. Thus, the evidence suggests that in written peer review activities, much like Huang (1996) found in oral feedback sessions, L1 feedback will focus more on issues of language (i.e., at the word/sentence/paragraph level); however, commentary of this type is much more likely to be implemented when delivered in the L2. While no real differences in frequency of comments regarding issues of language use, rhetoric, and reasoning were found between L1 and L2 use, students were more likely to incorporate such suggestions when delivered in their L1.

### Conclusion

These findings can be of some use for curricular planning because they suggest that students are unconsciously focusing on opposite areas of essay analysis depending upon the language of the peer review. If these results are representative, it would follow that teachers could strategically use L2 peer review to good effect for focus on language form issues, and that it would be more effective to allow peer reviews focusing primarily on issues of rhetorical effectiveness to be conducted in an L1.

It must be noted that the study described herein is limited by nature, being a relatively small test group from a single nation/language background, therefore it would be advisable to ascertain the applicability of such

results to the larger East Asian or global context via additional testing in other classroom contexts. These results cannot yet be considered as indicative of broader L1/L2 issues, but rather as local (i.e., Japanese/English, university-level, etc.) effects. While the results *might* fall within a larger trend, it is important not to generalize the results of 32 students in northern Japan without plenty of verification from a wide array of locales and L1/L2 combinations. Additionally, the small sample size introduces a power concern (seen herein by all significant effects only having small or medium effect size) which could only be eliminated through more expansive testing of this type. Furthermore, the institution where the study was administered, being a rather small English-medium university with a high national ranking, there is admittedly a high chance for selection bias to have influenced the results. In order to be able to generalize these results beyond their immediate context, it would be useful to try to replicate the study in other areas of the Asia Pacific region. Whether the results captured herein are specific to Japanese learners (or even a subsection thereof) or represent a general trend in L2 learners of essay writing, it would be invaluable to the field to amass more varied information on the interrelation of language and written peer review commentary. While national, regional, and/or linguistic differences may well emerge from such expanded study, this knowledge could then enable writing teachers to target peer review activities to best suit the individualized learning aims of classes.

### Notes

<sup>1</sup>Problems here are rare, but it is very important to make sure that peer reviews are not used as a forum for airing grievances or executing vendettas).

<sup>2</sup>Translation from Japanese by one of the graduate students involved in the initial sorting.

### References

- Allison, D., & Ng, P. (1992). Developing text revision abilities. In M. Lau and M. J. Murphy (Eds.), *Developing Writing: Purposes and Practice* (pp. 106-130). Hong Kong: Institute of Language in Education.
- Arndt, V. (1993). Response to writing: Using feedback to inform the writing process. In M.N. Brock and L. Walters (Eds.), *Teaching Composition Around the Pacific Rim: Politics and Pedagogy* (pp. 90-116). UK: Multilingual Matters.
- Beaven, M.H. (1977). Individualized goal setting, self-evaluation, and peer evaluation. In C.R. Cooper and L. Odell (Eds.), *Evaluating Writing: Describing, Measuring, Judging* (pp. 135-156). Urbana, IL: National Council of Teachers of English.
- Coniam, D., & Wong, R. (2004). Internet Relay Chat as a tool in the autonomous development of ESL learners' English language ability: An exploratory study. *System*, 32(3), 321-335.
- Crank, V. (2002). Asynchronous electronic peer response in a hybrid basic writing classroom. *Teaching English in the Two Year College*, 30(2), 145-155.
- Dearden, J. (2014). *English as a medium of instruction – a growing global phenomenon: Phase 1*. London: British Council.
- Ferris, D.R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Routledge.
- Gere, A.R., & Abbot, R. (1985). Talking about writing: The language of writing groups. *Research in the Teaching of Writing*, 19, 362-385.
- Huang, S.Y. (1996). L1 or L2 peer response sessions? Differences in verbal interaction between a writing group that communicates in Mandarin Chinese and one that uses English. *Tunghai Journal*, 38(1), 131-151.
- Jacobs, G.M., Curtis, A., Braine, G., & Huang, S.Y. (1998). Feedback on student writing: Taking the middle path. *Journal of Second Language Writing*, 7, 307-317.
- Keh, C.L. (1990). Feedback in the writing process: a model and methods for implementation. *ELT Journal* 44(4), 294-304.
- Liao, H.C., & Lo, Y.F. (2012). Peer review comments provided by high- and low-proficiency L2 learners: A comparative study. *International Journal of English Linguistics*, 2(5), 45-54.

- Liu, J. & Sadler, R.W. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, 2, 193-227.
- LoCastro, V. (2000). Evidence of accommodation to L2 pragmatic norms in peer review tasks of Japanese learners of English. *JALT Journal*, 22(2), 245-270.
- Lockhart, C., & Ng, P. (1993). How useful is peer response? *Perspectives*, 5(1), 17-29.
- Mendonca, C.O., & Johnson, K.E. (1994). Peer review negotiations: revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769.
- Nelson, G.L., & Murphy, J.M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27, 135-142.
- Reed, J. (2002). The pedagogical challenges for Western ESL teachers in Asia. *TESL Ontario: Contact*, 28(4), 1-8.
- Salih, A.R.A. (2013). Peer response to L2 student writing: Patterns and expectations. *English Language Teaching*, 6(3), 42-50.
- Strenski, E., Feagin, C.O., & Singer, J.A. (2005). Email small group peer review revisited. *Computers and Composition*, 22(2), 191-208.
- Trend alert: English spreads as teaching language in universities worldwide (2012, July 11). *ICEF Monitor*, Retrieved from <http://monitor.icef.com/2012/07/trend-alert-english-spreads-as-teaching-language-in-universities-worldwide/>
- Tsui, A.B.M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147-170.
- Wu, W.S. (2006). The effect of blog peer review and teacher feedback on the revisions of EFL writers. *Journal of Education and Foreign Languages and Literature*, 3, 125-138.
- Xu, Y. (2007). Re-examining the effects and affects of electronic peer reviews in a first-year composition class. *The Reading Matrix*, 7(2), 1-21.
- Yu, S.L., & Lee, I. (2015). Understanding EFL students' participation in group peer feedback of L2 writing: A case study from an activity theory perspective. *Language Teaching Research*, 19(5), 572-593.
- Zhang, S. (1995). Re-examining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209-222.

## **Appendix 1** **English Peer Review Form**

### **Peer Review Worksheet:**

#### *Introduction Paragraph*

1. Write what you believe the thesis is.
2. Where is it located? Is it in the expected place (the very last sentence of the Introduction)?
3. How does the writer lead up to the thesis? Is the information useless, helpful, boring, interesting, alluring, off-putting? Is the information directly connected to the thesis? Explain your answer.

#### *Body Paragraphs*

4. List the individual topics for each paragraph.
5. Are there any paragraphs that do not deal with a specific topic?
6. How well do the topic sentences for each body paragraph represent what those paragraphs contain?
7. How well do the body paragraphs serve as evidence for the thesis? Do they directly connect to the thesis? Are any points unexplained?
8. Which paragraphs seem useless, aimless, or need reorganizing? Explain any problems.
9. How well does each paragraph transition to the next? Do the paragraphs seem disjointed or carefully arranged?

#### *Conclusion Paragraphs*

10. How long is the conclusion?
11. Is the thesis *restated* in some way in the Conclusion?
12. How alike/different is it from the thesis in the Introduction?
13. Where is the thesis located?

#### *Evaluation Summary*

14. What are three excellent aspects of this paper?
15. If you had to make three recommendations for change, what would they be? Name them in order of importance.
16. How much effort do you think went into this draft?

## Appendix 2

### Japanese Peer Review Form

#### 相互評価シート

##### 序文

1. この論文の主題を述べなさい
2. 主題はどこに書かれていますか。主題は序文の最後の部分にありましたか。
3. 筆者はどのように主題へと導いていますか。序文の導入部に書かれている情報は役立ちましたか。または（不要、退屈、興味深い、魅惑的、的外れ）ものでしたか。それらの情報は主題に直接的に関与していましたか。自分の意見を述べなさい。

##### 本文

4. 各段落の論題を述べなさい
5. それぞれの論題に対し直接関与していない段落はありましたか
6. 各段落においてそれぞれの論題は明確に示されていましたか
7. 本文はこの論文の主題を証明するものとしての役割を果たしていますか。主題に直接繋がるものですか。説明が不十分な部分はありましたか
8. 不必要、目的のない段落がある場合はそれらの問題を指摘しなさい
9. 各段落はそれぞれ次の段落に円滑に繋がっていましたか。各段落は分裂しているものでしたか。それとも注意深く繋がられていたものでしたか

##### 結論

10. 結論はどのくらいの長さでしたか
11. 主題と結論に関連性はありましたか
12. 序章の主題と結論を比較し、似ている部分、異なっている部分を述べなさい
13. 主題はどこに書かれていますか

##### 評価

14. この論文において優れた見地を3つ述べなさい
15. 校正すべき部分を3つ、最重要箇所から順に述べなさい
16. この論文の下書きはどの程度練られていたか自分の意見を述べなさい

---

#### About the Author

Clay Williams is an associate professor in the English Language Teaching graduate program at Akita International University. He is the author of *Teaching English in East Asia: A Teacher's Guide to Chinese, Japanese, and Korean Learners*, as well as other books and articles about English pedagogy in the East Asian region.