



Examining the Effects of Peer Review in EFL Writing: A Coh-Metrix Analysis

Philip Yoongoo Jung*
Dongguk University

Jeong-Ah Shin
Dongguk University

ARTICLE INFO

Received: 10 September 2021

Revised: 24 October 2021

Accepted: 05 November 2021

Examples in: English
Applicable Languages: English
Applicable Levels: Tertiary

KEYWORDS

*Coh-Metrix analysis/
EFL writing/feedback/
corpus analysis*

코메트릭스 분석/외국어 쓰기/
피드백/코퍼스 분석

ABSTRACT

Jung, Philip Yoongoo, & Shin, Jeong-Ah. (2021). Examining the effects of peer review in EFL writing: A Coh-Metrix analysis. *Modern English Education*, 22(4), 60-69.

This study investigated the effectiveness of peer feedback in EFL writing classes. A total of 285 original editorials by 103 Korean learners of English were compared with revised drafts reviewed by classmates. Coh-Metrix, an automated tool offering a wide range of measures on cohesion and language was used for analyses. Fourteen indices of the Coh-Metrix system were selected to observe variances in lexical (i.e., TTR, frequency, age of acquisition, concreteness, familiarity, meaningfulness, polysemy), syntactic (i.e., number of words before the main verb, NP/VP density), and discourse (i.e., argument/content word overlap, LSA) aspects between the two sets of writings. There were significant improvements in the lexical aspect as positive modifications were noted in revised drafts. In addition, selective enhancements were observed in the syntactic aspect. Although the number of words before the main verb increased in the revised drafts, NP/VP density resulted in no difference between the two corpora. For the discourse aspect, no significant difference was found. Overall, the implementation of peer review led to positive outcomes in specific areas. However, results also revealed overt limitations of the practice in EFL settings. These findings suggest some educational implications in EFL writing pedagogy.

I. INTRODUCTION

Numerous studies have viewed peer feedback as a key technique for improving students' writing as it provides social, cognitive, affective, and methodological benefits (Mendonca & Johnson, 1994; Villamil & de Guerrero, 1996), and particularly, a well-oriented feedback guides

students to better acknowledge their subject area with clear guidance on improving their learning output (Orsmond et al., 2013). Nevertheless, there are still doubts about the practice of peer feedback in ESL/EFL classes as first, teachers may question its value within their particular teaching context and second, students may be uncertain about its purpose and advantages (Rollinson, 2005). Spe-

* First author: Philip Yoongoo Jung, Corresponding author: Jeong-Ah Shin

Philip Yoongoo Jung (Graduate student)
Department of English Language and Literature, Dongguk University, 30, Pildong-ro 1-gil, Jung-gu, Seoul, 04620, Korea
Email: yoongoo.jung@gmail.com / ISNI: 0000 0005 0408 7980

Jeong-Ah Shin (Professor)
Department of English Language and Literature, Dongguk University, 30, Pildong-ro 1-gil, Jung-gu, Seoul, 04620, Korea
Email: jashin@dongguk.edu / ISNI: 0000 0004 7517 411X

cifically, L2 learners often feel that only a better writer or a native speaker is qualified to provide writing critiques and consequently, they may view peer feedback by classmates whose English proficiency is similar as theirs as an inferior alternative to the “real thing” (Rollinson, 2005, p. 23). Such negative perceptions in ESL/EFL contexts may rise to be a reason for the reluctant adoption of peer feedback in teaching/learning writing, despite the many general advantages reported toward its practice thus far. This quantitative study intends to validate the advantages and limitations in the practice of peer review in an EFL context. We hypothesize that L2 learners would demonstrate some peculiar behaviors/patterns in providing English critiques, compared to L1 peer feedback. Thus, we present the following research question: Are there any differences in the lexical, syntactic, and discourse aspects between the original newspaper editorials written by advanced Korean learners of English and the revised drafts, peer reviewed by classmates? Two sets of writings (i.e., original and revised) are compared using Coh-Metrix, an automated analysis tool that provides quantitative measures on cohesion and language.

II. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

1. (Peer) Feedback Performance and Quality

Feedback is defined as “information provided by an agent regarding some aspect of one’s task performance” (Hattie & Timperely, 2007, p. 102), and “all post-response information that is provided to a learner to inform the learner on his or her actual state of learning or performance” (Narciss, 2008, p. 127). Feedback is known to produce strong positive effects on learning (see Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996), but such effects can be absent or even cause negative outcomes depending on the instructional conditions. Not only limited to if, but how feedback supports learning has been rationalized by Mory (2004) in four perspectives. Feedback can be considered as: incentive (for increasing response rate and/or accuracy), reinforcer (for connecting response to prior stimuli), information (to validate or change previous responses), and provision of scaffolds (to analyze learning processes). Contrast to the more formal method of feedback provision by the teacher, peer feedback has also gained interest by many educationalists. The key difference between the teacher feedback and peer feedback is that peers are not domain experts, as opposed to teachers (Gielen et al., 2010). Thus, it is quite projectable to generalize that the accuracy via peer feedback fluctuates. Tsui and Ng (2000) noted in their study that when students receive both teacher and peer feedback, teacher feedback was more frequently incorporated in revisions than peer feedback. No analysis toward the quality of revised final drafts was examined but the study reports that students found comments by teacher more useful. This study outlines some general

problems of peer feedback, such as depth of the feedback, accuracy, and credibility.

Nevertheless, having been practiced in many areas of writing, peer feedback still offers social, cognitive, affective, and methodological benefits (Mendonca & Johnson, 1994; Villamil & de Guerrero, 1996). Particularly, the practice of peer-evaluation in L1 has reported positive outcomes in previous literature; students are known to prioritize comments by peers over teacher-judgments (K. Cho & Schunn, 2007; Pierson, 1967) as well as receive more social support in peer responses than teacher feedback (Elbow, 1973). In addition, there are severe criticisms about the contrasting instructor-feedback. Teachers can be careless, exhausted, or insensitive as their comments are often unhelpful and confusing (Purves, 1984). Teachers’ intentions are forced and often involuntarily reflected in students’ writing (Brannon & Knoblauch, 1982). Some scholars even view the method of feedback in a writing class as a political struggle for power and status (Berlin, 1988; Fox, 1990; Perdue, 1984, cited in Santos, 1992). In a similar vein, Topping (1998) argued that peer feedback is more effective than teacher feedback due to the absence of “knowledge authority” (e.g., the teacher). As briefly introduced, advocates of peer feedback present a variety of pedagogical, psychological, or ideological grounds for its justification in education.

There are some opposing views toward peer review in L2 contexts. Nelson and Murphy (1993) claimed that the general problems of peer feedback (e.g., depth of feedback, accuracy, credibility, etc.) tend to be more apparent in L2 environments, compared to first language classes. General skeptical attitude of L2 writers is also reported as learners tend to view the practice of peer review by classmates as a poor alternative to peer reviews by native speakers (Rollinson, 2005). Yet, it is noted that L1 and L2 writing are practically identical (Zhang, 1995) as arguably, L2 writing theories have derived from L1 writing theories (Santos, 1992; Silva, 1993). Other works of literature further posit the similarities L1 and L2 writing share (Arapoff, 1969; Edelsky, 1982; Jacob, 1982), agreeing upon the transcendence of language factors in writing process. Chaudron (1984) thereby rationalized the benefits of peer feedback in L2 writing as follows: peer feedback is more informative than teacher’s feedback, open to a wider audience for broader perspectives, socially supportive, and a gateway to constructive learning effects of writing.

Considering the “quality” aspect of peer feedback, a few perspectives are signified. One perspective views peer feedback quality in terms of accuracy, consistency across assessors and/or concordance with teacher feedback (see Van Steendam et al., 2010). This approach originated from a summative view on peer assessment which emphasizes the scoring validity and reliability as primary concepts (Gielen et al., 2010). Another perspective views peer feedback quality in terms of content and/or style characteristics. Gielen et al. (2010) noted that these characteristics are not domain- and/or task-specific, pointing to a clear advantage of this approach. Therefore, training students

to concentrate on content and style characteristics results in generic skill transferable to other settings. Some studies based on this approach on content and/or style are as follows.

M. Kim's (2005) study, which bases this approach (on content and/or style), analyzed the relationship between feedback composition and performance. She noted that no increase in performance was reported for those who received constructive peer feedback. She further argued that this might have been due to the limited variance in peer feedback quality and/or writers' skepticism on their peers' capability as assessors.

Sluijsmans et al. (2002) also conducted a study based on the content and style perspective. They extracted and defined characteristics of constructive feedback from expert assessment reports and rationalized seven key criteria of feedback (see Sluijsmans et al., 2002). The study claims that by adopting an interventional perspective, students can be instructed to apply the key characteristics in giving feedback.

2. Editorials by EFL Students as Pedagogical Research Materials

Newspaper editorials are considered a type of argumentative writing as they exhibit professional stance or subjective opinions of the newspaper company. Indeed, it is quite common to encounter dissimilar interpretations by different newspapers toward a mutual event. In contrast, newspaper articles are meant to contain objective facts. Thus, editorials are often written for the purpose of persuasion to affect the readers' views/perspectives toward controversial issues (Connor, 1996). Consequently, editorials may comprise more "flexible" and diverse expressions. Most studies in editorial analyses are limited to discourse features. For more (psycho)linguistically detailed analyses, J. Lee (2020), with Coh-Metrix, compared editorials written by Korean, American, and British students. Her findings concluded that, among the three groups of students, Korean EFL learners used the easiest words, less complex syntactic patterns, and notably utilized connectives as cohesive devices. She also noted that while many studies exploited professional editorials in major newspapers (see Bolgun & Mangla, 2017; J. Choi, 2015; E. Y. Jang, 2014; C. K. Kim, 2009; Ljung, 1997), few did editorials by students, thereby suggesting a need for their adoption as a research material. The present study posits the suggestion and extends the analyses of editorials written by L2 learners of English by applying the pedagogical practice of L2 peer review. Generally, successful acquisition of a second language is presumed on how native-like her linguistic comprehension and output are. As Coh-Metrix studies have empirically shown the improvements in quality in L1 writing compared to L2 (see Coh-Metrix analyses between L1 and L2 production in S. Ahn, 2018; Crossley & McNamara, 2009, 2014; J. Lee, 2018, 2020), we postulate that any linguistic enhancement (computed by Coh-Metrix) reflected on an L2 output is a positive modification, bringing the nature of the L2 output

closer toward L1 output-like. Upon this rationale, the present study intends to seek how feedback from a non-native speaker may elicit such constructive modifications, hypothesizing that, if feedback turns out to be effective, this quantitative comparison is similarly applicable/interpretable to those between L1 and L2 production.

3. The Coh-Metrix System

Coh-Metrix is an automated tool developed by the Institute for Intelligent Systems at the University of Memphis for the purpose of analyzing the linguistic and psycholinguistic features of English text with a wide range of measures on cohesion and language. The Coh-Metrix system has served as a reliable textual assessment tool in many areas (Crossley & McNamara, 2014; McNamara et al., 2010). Its applications range not only from comparing simple writing output by different groups to evaluating formal publications such as English exams, academic papers, news articles/editorials, textbooks/teacher-guidebooks (B. Ahn & Y. Ma, 2015; J. Y. Chang, 2018; M. Jeon, 2011, 2014, 2015; M. Jeon & I. Lim, 2009; P. Y. Jung & J. A. Shin, in press; J. Kim & J. Yang, 2002; S. Kim & M. Jeon, 2016; J. Lee, 2018, 2020; S. H. Lee, 2013; J. Ryu & M. Jeon, 2020; I. Sung, 2014). Among the 106 indices offered by the tool, those selected for the present study are briefly explained below by sub-categories.

1) Lexical Diversity

The Coh-Metrix system computes the diversity of lexical items in a text segment through the type-token ratio (TTR). TTR is calculated by dividing the number of unique words (i.e., types) by the total number of words (i.e., tokens). Thus, the maximum output value is 1 and the closer the output is to 1, the more diverse lexical items present in the target segment.

2) Word Frequency

Word frequency in Coh-Metrix is computed using CELEX corpus, database from the Dutch Centre for Lexical Information (Baayen et al., 1995). Word frequency represents how regularly particular words are included in a text segment and such measure is argued to be an important degree to determine the difficulty of a text (Graesser et al., 2004; Graesser et al., 2007). It is generally expected that the higher word frequency points to inclusion of more familiar and easier words in the text segment.

3) Word Features

The Coh-Metrix system provides word features using the MRC (Medical Research Council) Psycholinguistic Database (Coltheart, 1981). Indices, *age of acquisition*, *concreteness*, *familiarity*, *meaningfulness*, and *polysemy* are suggested to be crucial factors in determining the difficulty of texts (Graessers et al., 2004; Jeon, 2015; Ryu &

Jeon, 2020). *Age of acquisition* refers to the age-of-acquisition norms by Gilhooly and Logie (1980), based on the assumption that certain words appear in learners' language use earlier than others. *Concreteness* is to verify how concrete a word is; "pencil" is more concrete than "policy." *Familiarity* is a rating of how familiar a word is to adults. *Meaningfulness* denotes to how meaningful a word is based on a large corpus developed by Toglia and Battig (1978). *Polysemy* computes the number of senses of a word. A word with multiple senses (e.g., *bank*) may be ambiguous, which may hinder processing for less-skilled or low-knowledge readers (Gernsbacher, 1990; Just & Carpenter, 1987; McNamara & McDaniel, 2004).

4) Syntactic Complexity

McNamara et al. (2014) suggest that the syntax of a text segment tends to be easier to comprehend with shorter sentences, fewer words before the main verb of the main clause. In addition, density measurements of phrase types are vital donors in determining the syntactic complexity of a text segment. Specifically, incidences of noun phrases (NP) and verb phrases (VP) affect processing difficulty of text as higher incidences indicate that the text is more informationally dense with complex syntax (McNamara et al., 2014).

5) Referential Cohesion

A text segment with high referential cohesion includes words and ideas that are overt in the whole text, providing explicit grounds that connect the text (McNamara et al., 2014). Argument overlap refers to the overlap of nouns, pronouns, or noun phrases. Content word overlap considers the sharing of content words (e.g., nouns, verbs, adjectives, adverbs). Higher reference cohesion signifies increased in the comprehension, potentially suggesting the decrease in its difficulty (Graesser et al., 2007).

6) Semantic Cohesion

The Coh-Metrix system measures the semantic cohesion through the applications of the Latent Semantic Analysis (LSA), a mathematical and statistical technique based on a large corpus for representing world knowledge. LSA computes the semantic similarities between words, sentences, and paragraphs (Landauer & Dumais, 1997; Landauer et al., 2007). Two separate LSA measures are provided: between adjacent sentences and for all sentences in a text.

III. METHOD

1. Participants and Data Collection

For two semesters in 2019, writing works from 103 students (mean age: 20.6; *SD*: 0.7) of Seoul National Univer-

sity of Education with a TOEIC score of 700 (out of 990) or above were collected for the current study (from nine different class sessions). Though TOEIC does not specifically test writing ability of a test taker, a relatively high score would arguably serve as an indication of advanced state among English learners; these 103 students, for the general purpose of this study, are classified as advanced learners of English. All participants were pre-service elementary school teachers with diverse sub-majors (e.g., English education, Science education, Music education, etc.). The English writing course requires each student to complete three 250-300 words writing assignments in a semester. Students were informed of this research, and all provided a consent for the use of their writing works. They were not compensated as no extra work or attention was obligatory out of the boundary of the course syllabus. For each assignment, students were required to find and select any picture which they believed impacted the world. Sample images provided by the instructor included those visually demonstrating important historical events such as the collapse of the Berlin Wall in 1989, Apollo 11 landing of the moon in 1969, the Hindenburg disaster in 1937, etc. Such iconic photos and their relevant information can easily be found on the internet. For each photo, students were to write a news editorial which explains the event in detail along with their implications and personal/professional opinions. Shortly after the submission, each student was given an editorial written by another student in the same class session to conduct a detailed peer review by providing comments, corrections, and suggestions. These to-be-reviewed, original papers were randomly distributed with the writers' personal information (i.e., name, student number) removed. Then, the peer reviewed versions were collected and returned to the initial writers with reviewers' personal information removed; the writers were to conduct the last round of self-revision based on the feedback received prior to the final submission of the assignment. The writers were specifically instructed to minimize any addition of new content to the draft but to make revisions mainly based on the feedback received from the reviewers. Each student was given a week to complete their own assignment for the initial submission, a week to complete a peer review of another student's work, followed by another week to revise their own work that was once reviewed by a classmate for the final submission. This process was repeated three times (for three assignments). Any original drafts that did not receive peer feedback (as some irregularly failed to provide peer feedback) were excluded from the corpus package; out of 300 original drafts written and collected, 15 were removed from analyses. In sum, two sets of corpora were constructed: a set of 285 originally submitted writing assignments and another of their revised, peer reviewed versions.

2. Coh-Metrix Indices

Using 14 indices of the Coh-Metrix 3 system, the purpose of this study is to compare the lexical, syntactic, and

discourse aspects of news editorials written by advanced Korean learners of English and their respectively peer reviewed versions. For the lexical aspects, lexical diversity (i.e., type-token ratio) and word information data (i.e., frequency of content words, age of acquisition, concreteness, familiarity, meaningfulness, and polysemy) were stored. For the syntactic complexity analyses, the number of words before the main verb, NP density, and VP density data were computed. To analyze the discourse aspects, referential cohesion (i.e., argument overlap and content word overlap) and semantic cohesion (LSA between adjacent and all sentences) were selected as variables. One might question the exclusion of the descriptive comparison between the two corpora (e.g., word/sentence/paragraph count) in this research. It was done so considering the nature of the task given to students (i.e., peer review). In contrast to comparing two fully varied writings (e.g., L1 vs. L2), we presume that differences in the descriptive statistics between pre- and post-peer-reviewed writings are likely minimal to none (as writers were instructed not to add entirely new content but to conduct the revision mainly based on feedback received). Table 1 shows the details of the indices selected.

TABLE 1
The Coh-Metrix Indices for the Present Study

Aspect	Variable	Measurement	Index
Lexical	Lexical diversity	Type-token ratio (all words)	LDTTRa
		Frequency (content words)	WRDFRQmc
	Word information	Age of acquisition	WRDAOAc
		Concreteness	WRDCNCc
		Familiarity	WRDFAMc
		Meaningfulness	WRDMEAc
Syntactic	Syntactic complexity	Words before the main verb	SYNLE
		NP density	DRNP
		VP density	DRVP
Discourse	Referential cohesion	Argument overlap (global)	CRFAOa
		Content word overlap (global)	CRFCWOa
	Semantic cohesion	LSA (adjacent sentences)	LSASS1
		LSA (global)	LSASS1d

3. Statistical Analysis

Paired samples *t*-tests were conducted to compare the two corpora constructed. The stored results computed by Coh-Metrix were analyzed using SPSS 25.0 (Statistical Package for the Social Sciences). Independent variables were the gathered corpus packages, and the dependent variables were the 14 Coh-Metrix indices selected (see

Table 1 for details). Each analysis was performed at a significance level of 5% ($p = .05$).

IV. RESULTS AND DISCUSSION

1. Lexical Aspect

1) Lexical Diversity

As presented in Table 2, statistically significant differences between the two independent variables were found for TTR ($t(285) = -11.441, p < .001$). Crossly and McNamara (2011) argued that lexical diversity is highly associated with the quality of English writing. Revised, peer reviewed versions revealed a significant increase in the output, indicating the improvement of the writings' richness in terms of lexical diversity. The TTR results indicate that through a detailed peer review, interestingly, though done by (non-native) classmates in the same class, the quality of writing (in terms of lexical diversity) indeed improved. This further suggests that with sufficient and consistent utilization of peer feedback in writing classes, learners can certainly produce writings with more diverse lexical items. The findings seem to be a potential application to previously reported concerns that Korean learners' need to learn to exploit diverse English expressions in their writings (M. K. Jeong & N. Kim, 2014).

TABLE 2
Results of Lexical Diversity

Variable	Original ($n = 285$) <i>M</i> (<i>SD</i>)	Revised ($n = 285$) <i>M</i> (<i>SD</i>)	<i>t</i>	<i>p</i>
TTR	.471 (.078)	.564 (.126)	-11.441	.000***

*** $p < .001$

2) Word Frequency (Content Words)

In this study, the frequency of content words was analyzed. Content words (e.g., nouns, verbs, adjectives, adverbs) are generally known to be rarer in writings compared to other words (e.g., function words such as prepositions, determiners, pronouns) and are also considered semantically richer. Frequent use of content words is associated with richer bodies of world knowledge (Beck et al., 2002; Haberlandt & Graesser, 1985; Perfetti, 2007), hypothetically signifying an increase in the writing's general quality. As can be seen in Table 3, word frequency measures of content words in students' news editorials significantly decreased in the revised drafts ($t(285) = 7.53, p < .001$). The results advocate that the peer reviewed drafts included richer content words, highly denoting that substantial improvements were made. The findings empirically support the benefits of peer feedback in L2 (Chaudron, 1984) as the writers seem to have developed broader perspectives (i.e., world knowledge) with the use of more variety

of content words in their writings based on the peer feedback received. The results, thus further proposes the prominence applications of peer review in teaching/learning English writing.

TABLE 3
Results of Word Frequency (Content Words)

Variable	Original (n = 285) M (SD)	Revised (n = 285) M (SD)	t	p
Word frequency	1.28 (.314)	1.017 (.513)	7.53	.000***

*** p < .001

3) Word Features

Results of word features are outlined in Table 4. For *age of acquisition*, revised, peer reviewed versions computed higher output value compared to the original drafts, indicating the inclusion of lexical items that are relatively for older learners ($t(285) = -9.225, p < .001$). No statistically significant differences were detected for *concreteness*. For *familiarity*, the output value decreased from the original to peer reviewed versions ($t(285) = 11.569, p < .001$), implying the inclusion of more less familiar words (i.e., difficult words) in the revised drafts. For *meaningfulness*, the output value was detected to be significantly lower in the revised drafts, suggesting the inclusion of more relatively difficult lexical items ($t(285) = 4.528, p < .001$). A significant difference was also detected for *polysemy* ($t(285) = 5.437, p < .001$); the value decreased in the revised drafts, implying the inclusion of fewer higher frequency words. Overall, the results of the word features showed that a greater variety of relatively difficult words were included in the revised, peer reviewed drafts. Thus, as a results, it indicates that the writers generated higher quality editorials in terms of the lexical applications, based on peer feedback received.

TABLE 4
Results of Word Features

Variable	Original (n = 285) M (SD)	Revised (n = 285) M (SD)	t	p
Age of acquisition	332.494 (40.644)	364.278 (52.276)	-9.225	.000***
Concreteness	381.722 (21.93)	381.955 (22.974)	-.13	.897
Familiarity	577.135 (7.983)	567.883 (12.241)	11.569	.000***
Meaningfulness	437.729 (13.665)	432.27 (16.194)	4.528	.000***
Polysemy	4.087 (.348)	3.924 (.39)	5.437	.000***

*** p < .001

2. Syntactic Aspect

Table 5 reports the results of syntactic complexity of the original and the revised, peer reviewed editorials. Crossley and McNamara (2011) proposed that syntactically complex texts lead to higher quality of writing. In comparing the syntactic aspect of the two sets of editorials, the number of words before the main verb and NP/VP density data were stored. First, statistically significant difference was detected comparing the number of words before the main verbs between the original and revised editorials ($t(285) = -4.288, p < .001$). The results show that the number of words before the main verb increased in the revised drafts. The findings are applicable with M. K. Jeong's (2015) study that the native speakers' argumentative essays contain more words before the main verb compared to the non-native speakers', indicating that the revised drafts were more native-like compared to the original drafts. Further note that the genres are identical to an extent as editorials are a part of argumentative writing. Second, no difference was detected for either phrasal density. The improvements, therefore, were revealed to be selective: and this may be due to the limitations (e.g., language proficiency) the writers/reviewers possess. As can easily be assumed, peer feedback in classroom contexts have some limitations such as limited depth of feedback, accuracy, and credibility. Such lacking characteristics tend to be more apparent in L2 contexts, compared to any first language classes (Nelson & Murphy, 1993).

TABLE 5
Results of Syntactic Complexity

Variable	Original (n = 285) M (SD)	Revised (n = 285) M (SD)	t	p
Words before the main verb	4.141 (1.441)	4.732 (2.093)	-4.288	.000***
NP density	358.181 (22.919)	364.975 (34.632)	-2.81	.084
VP density	229.006 (33.049)	218.688 (37.533)	3.852	.914

*** p < .001

3. Discourse Aspect

Table 6 shows the analyzed results of the discourse aspects. No statistically significant differences between the two versions of editorials were observed in any index computed. The results, therefore, indicate no specific improvements in terms of the discourse aspects, further revealing the limitations demonstrated by the writers/peer reviewers. Discourse, highly associated with pragmatics differs from the lexical/syntactic aspects of a language. Unlike grammar or technical vocabulary, it is widely accepted that pragmatical knowledge is something we acquire along the acquisition of

L1. However, acquisition of pragmatic fluency in L2 is argued to require intensive instruction (see Bouton, 1994; Cohen, 1996; House, 1996). Schmidt (1993) emphasized that learners must notice and focus attention on relevant input for any foreign language pragmatic acquisition. Bialystok (1993) argued the importance of possessing certain strategies toward the intended interpretations and acknowledge the social and contextual needs of the communicative situation. Providing feedback toward the lexical/syntactic aspects can be complemented with some external tools (e.g., dictionary, thesaurus, automated paraphrasing tool, etc.). Nevertheless, it is assumable that L2 learners may in general experience a great deal of difficulty in natural discourse/pragmatics.

TABLE 6
Results of Discourse Aspects

Variable	Original (<i>n</i> = 285) <i>M</i> (<i>SD</i>)	Revised (<i>n</i> = 285) <i>M</i> (<i>SD</i>)	<i>t</i>	<i>p</i>
Argument overlap	.433 (.152)	.458 (.173)	-1.77	.078
Content word overlap	.084 (.031)	.078 (.033)	1.979	.169
LSA (all sentences)	.196 (.076)	.209 (.078)	-2.29	.433
LSA (adj. sentences)	.169 (.041)	.153 (.051)	4.228	.063

V. CONCLUSION AND IMPLICATIONS

The present study compared the differences in the lexical, syntactic, and discourse aspects of original news editorials written by advanced Korean learners of English and their revised drafts, revised based on feedback provided by classmates. Two sets of writings (i.e., original and revised) were collected for each of the three writing tasks to construct two corpora. Using Coh-Metrix, an automated tool to analyze linguistic and psycholinguistic features of English texts, 14 indices were selected to compare the two sets writings. The results, hope to offer a handle in establishing a base for writing pedagogy and toward its systematic and quantitative assessment methods, specifically in applying the practice of peer review in EFL writing classes, despite the overt limitations (e.g., limited proficiency) students (i.e., both writers and reviewers) share as non-native speakers of English. In addition, writers of this research received no professional or official training on how to provide critiques for English writing. This would also imply that, if any significant improvements are observed in the post-review drafts, two implications can be made; first, the implementation of the practice of peer review is defensible in teaching/learning EFL writing, and second, it is easily practicable as even untrained, non-native speakers could provide meaningful L2 peer feedback.

The findings and their educational implications are as follows.

For the lexical aspects, data of lexical diversity, word frequency of content words, and various word features (i.e., age of acquisition, concreteness, familiarity, meaningfulness, polysemy) were stored. The TTR results revealed an increase in inclusion of more diverse lexical items in the revised (peer reviewed) drafts, indicating the positive effects of applying the practice of peer reviews in teaching/learning EFL writing. The reviews/feedback given by other students led to broadening the use of more diverse English expressions by the writers in their revised drafts, as pointed out to be an important necessity for Korean students to acquire (M. K. Jeong & N. Kim, 2014). Word frequency of content words were also computed to observe whether the quality of writing is improved with peer reviews. The use of content words is suggested to be a significant measure in determining the presence of richer bodies of world knowledge (Beck et al., 2002; Haberlandt & Graesser, 1985; Perfetti, 2007). The two sets writings show significant differences; namely, the revised drafts contained richer content words. Analyses of word features also exhibited positive reflections in the revised drafts as word repetitions were reduced with inclusion of more difficult and distinct lexical items, additionally proposing the encouraging applications of peer reviews in teaching/learning writing.

In analyzing the syntactic complexity of editorials, data for the number of words before the main verb, NP density, and VP density were collected. Number of words before the main verb is one norm to determine the difficulty of comprehension, as argued that syntactically complex (e.g., longer sentences) texts tend to produce higher quality in general (Crossley & McNamara, 2011). The number of words before the main verb increased from the original to the revised drafts, indicating that a complexity has increased after the peer review. As native speakers' editorials are reported to contain more words before the main verb compared to L2 editorials (M. K. Jeong, 2015), it is projected that the practice of peer reviews in EFL writing leads to more natural, native-like writing output in terms of sentence complexity. However, data of NP/VP density were insignificant. Overall, the results of syntactic aspects demonstrated selective improvements in the revised drafts. Particularly, such inconsistent outcomes may be due to the limitations/tendencies of the students, potentially both the writers and the reviewers, as being non-native speakers of English. We project that with sufficient exposures to authentic writing samples along with appropriate training measures toward writing and conducting peer reviews, positive outcomes can be expected. Regarding the detailed syntactic tendencies of how L2 writers conduct peer reviews, we leave it for future research.

For the discourse aspects, argument overlap, reference cohesion (i.e., argument overlap between adjacent

sentences and across the whole text) and semantic cohesion (i.e., LSA) were analyzed. The two types (i.e., original and revised) did not show any statistically significant differences. This may possibly indicate that peer reviewers were unable to minimize the repeated use of certain word or information. The practice of peer review did not result in any improvements, further indicating the overt linguistic limitations L2 learners possess. L2 discourse competence, as highly associated with pragmatics, requires intensive trainings and exposures (see Bouton, 1994; Cohen, 1996; House, 1996). Communication-based approach (e.g., *Communicative Language Teaching, Task-based Instruction*) with sufficient exposure to authentic English is suggested to reinforce L2 writers' discourse competence.

Lastly, some limitations of the present study and suggestions toward the future ones can be addressed. First, in the writing classes of the current study, no systematic trainings on providing peer reviews/feedback have been conducted. This was purposeful as to observe how relatively untrained non-native speakers' feedback can affect the writing output. The results, as explained above confirmed partial/selective improvements in writing, but of course, this study surely awaits further analyses with the involvement of trained L2 peer reviewers. Second, the content/topic of editorials may have produced unique outcomes in terms of linguistic and psycholinguistic features, as suggested by C. K. Kim (2009), linguistic features could be quite different depending on the topics of the texts. Third, this study could not eliminate the variable of self-revision completely. Though writers were instructed not to add any entirely new content, but to conduct revision mainly based on the feedback received, we acknowledge that the revised drafts analyzed likely have been affected by both peer feedback and self-revision.

In sum, the current study quantitatively analyzed how peer reviews by non-native speakers of English can affect writing output. Though, the results and interpretations were selective and inconsistent to an extent, they have demonstrated that significant differences indeed exist between the original and the improved revised, peer reviewed drafts. This study hopes to contribute to the field of English writing pedagogy, particularly in applying systematic peer reviews in EFL writing, which is generally known to be rare and difficult.

REFERENCES

- Ahn, Byung-kyoo, & Ma, Yumi. (2015). A Coh-Metrix analysis of elementary school English textbooks. *English21*, 28(3), 435-460.
- Ahn, Soojin. (2018). An analysis of Coh-Metrix on the differences in English expository and argumentative writing of Korean and native English university students. *New Korean Journal of English Language and Literature*, 60(3), 177-205.
- Arapoff, N. (1969). Discover and transform: A method of teaching writing to foreign students. *TESOL Quarterly*, 3(4), 297-304.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania.
- Bangert-Drowns, R., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.
- Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary development*. Guilford Press.
- Berlin, J. (1988). Rhetoric and ideology in the writing class. *College English*, 50(5), 477-494.
- Bialystok, E. (1993). Symbolic representation and attentional control in pragmatic competence. *Interlanguage Pragmatics*, 3(1), 43-57.
- Bolgun, M. A., & Mangla, A. (2017). A contrastive rhetoric analysis of English and Hindi editorials. *Journal of Pan-Pacific Association of Applied Linguistics*, 21(2), 15-39.
- Bouton, L. F. (1994). Conversational implicature in the second language: Learned slowly when not deliberately taught. *Journal of Pragmatics*, 22(2), 157-167.
- Chang, Ji-Yeon. (2018). A comparison of 2017-2018 CSAT reading passages via Coh-Metrix: Focusing on descriptive, readability, and easibility measures. *Foreign Languages Education*, 25(4), 81-106.
- Chaudron, C. (1984). Evaluating writing: Effects of feedback on revision. *RELC Journal*, 15(2), 1-14.
- Cho, Kwangsu, & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline. *Computers and Education*, 48(3), 409-426.
- Choi, Jungeun. (2015). A comparative discourse analysis of English and Korean editorials. *Korean Journal of English Language and Linguistics*, 15(2), 231-257.
- Cohen, A. (1996). Developing the ability to perform speech acts. *Studies in Second Language Acquisition*, 18(2), 253-267.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 497-505.
- Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second-language writing*. Cambridge University Press.
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119-135.
- Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21(2-3), 170-191.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79.

- Edelsky, C. (1982). Writing in a bilingual program: The relation of L1 and L2 texts. *TESOL Quarterly*, 16(2), 211-228.
- Elbow, P. (1973). *Writing without teachers*. Oxford University Press.
- Fox, T. (1990). *The social uses of writing: Politics and Pedagogy*. Ablex.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Erlbaum.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304-315.
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12(4), 395-427.
- Graesser, A. C., Jeon, Moongee., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15(3), 199-213.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114(3), 357-374.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18(2), 225-252.
- Jacobs, S. E. (1982). *Composing and coherence: The writing of eleven pre-medical students. Linguistics and literacy series: 3*. Center for Applied Linguistics.
- Jang, Eun-Young. (2014). Comparison of the discourse function of English relative pronouns used in 'news report' and 'editorial' of newspaper. *Studies in English Language & Literature*, 56(3), 365-382.
- Jeon, Moongee. (2011). A corpus-based analysis of the continuity of the reading materials in middle school English 1 and 2 textbooks with Coh-Metrix. *The Journal of Linguistic Science*, 56(1), 201-218.
- Jeon, Monngee. (2014). An analysis of the continuity of the reading materials in high school English textbooks. *Journal of the Korean Data Analysis Society*, 16(2), 925-938.
- Jeon, Moongee. (2015). An analysis of the continuity among middle school English textbooks with an automated language analysis program. *Modern English Education*, 16(1), 195-218.
- Jeon, Moongee, & Lim, Injae. (2009). A corpus-based analysis of middle school English 1 textbooks with Coh-Metrix. *English Language Teaching*, 21(4), 265-292.
- Jeong, Mikyung. (2015). *A comparative study on the English argumentative essays of Korean and native English university students with Coh-Metrix: Focused on lexical, syntactic, and discourse features* [Unpublished doctoral dissertation]. Kangwon National University.
- Jeong, Mikyung, & Kim, Namgook. (2014). An analysis of the linguistic features on the corpus of Korean EFL learners and native English speakers with Coh-Metrix. *Studies in Linguistics*, 33, 373-395.
- Jung, Philip-Yoongoo, & Shin, Jeong-Ah. (in press). Comparing the difficulty of classroom English in teacher-guidebooks by levels using Coh-Metrix. *English Teaching*.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Allyn & Bacon.
- Kim, Chul-Kyu. (2009). A corpus-based comparison of meta-discourse in argumentative texts of a British newspaper and Korean university students' English argumentative texts. *Discourse and Cognition*, 16(3), 65-88.
- Kim, Jeong-ryeol & Yang, Jiyun. (2002). An analysis of the continuity of elementary and middle school English textbooks using Coh-Metrix. *English Teaching*, 67(2), 319-341.
- Kim, Minjeong. (2005). *The effects of the assessor and assessee's roles on preservice teachers' metacognitive awareness, performance, and attitude in a technology-related design task*. [Doctoral dissertation], Florida State University, Tallahassee.
- Kim, Sojung., & Jeon, Moongee. (2016). An analysis study of English writing of elementary school 6th grade English language learners using Coh-Metrix. *Modern English Education*, 17(3), 263-287.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum.
- Lee, Jiyoung. (2018). A Coh-Metrix analysis of lexical, syntactic and discourse aspects in the newspaper articles of Korean and British university students. *Modern English Education*, 19(4), 17-26.
- Lee, Jiyoung. (2020). A Coh-Metrix analysis of the English editorials of Korean, American, and British university students. *Modern English Education*, 21(3), 15-27.
- Lee, Seong-Hwan. (2013). A Corpus-based analysis of the difficulty of the reading materials in high school English 1 and 2 textbooks using Coh-Metrix. *Studies in Foreign Language Education*, 27(2), 131-148.
- Ljung, M. (1997). Text complexity in British and American newspapers. In J. Aarts, I. de Mönnink, & H. Wekker (Eds.), *Studies in English language and teaching* (pp. 75-83). Rodopi.

- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 465-182.
- Mendonca, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769.
- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Lawrence Erlbaum.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. *Handbook of Research on Educational Communications and Technology*, 3, 125-144.
- Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27(1), 135-142.
- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. C. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240-252.
- Perdue, V. A. (1984). *Writing as an act of power: Basic writing pedagogy as social practice (Composition, literacy, rhetoric)*. [Unpublished doctoral dissertation], University of Michigan.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383.
- Pierson, H. (1967). *Peer and teacher correction: A comparison of the effects of two methods of teaching composition in grade nine English classes*. [Unpublished doctoral dissertation], New York University.
- Purves, A. C. (1984). The teacher as reader: An anatomy. *College English*, 46(3), 259-265.
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT journal*, 59(1), 23-30.
- Ryu, Jisu, & Jeon, Moongee. (2020). An analysis of the text complexity of listening scripts of high school English textbooks using Coh-Metrix. *Korea Journal of English Language and Linguistics*, 20(1), 363-383.
- Santos, T. (1992). Ideology in composition: L1 and ESL. *Journal of Second Language Writing*, 1(1), 1-15.
- Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21-42). Oxford University Press.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-677.
- Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education*, 27(5), 443-454.
- Sung, Ilho. (2014). An Analysis of Korean High School English Textbooks through Readability Formulae and Coh-Metrix. *Studies in English Language & Literature*, 40(4), 299-320.
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Erlbaum.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147-170.
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327.
- Villamil, O. S., & De Guerrero, M. C. (1996). Peer revision in the L2 classroom: Social-cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing*, 5(1), 51-75.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209-222.