〔論　　文〕

Balancing Corrective Feedback and Writing Practice in an EFL Writing Course

Yoshimasa Ogawa

Abstract

This is part of a series of investigation into the effects of multiple written corrective feedback (WCF) on Japanese EFL students' grammatical accuracy in writing. The multiple WCF involved comprehensive coded feedback and selective metalinguistic explanation. The study endeavored to determine whether there was a trade-off between repeated feedback and increased opportunities for writing practice. Statistical analyses also evaluated the coded WCF's effects on treatable and untreatable error types. The results supported the position that the feedback facilitates the students' ability to use verb tense accurately but does not help reduce the number of noun form or article errors (Ferris, 2006; Ogawa, 2021). More research is needed to clarify the treatability of noun form and article errors. The repeated-feedback treatment had a stronger effect on verb tense than the more-writing-practice treatment. The study also indicated that coded WCF alone had a positive effect on mechanical errors (i.e., treatable form) but had no effects on errors related to sentence structures, wrong words, or missing words (i.e., untreatable form). The implication was that thorough corrective feedback is crucial for form-focused writing instruction, regardless of the number of writing tasks assigned, and that the coded WCF, when used by itself, should also be focused on treatable errors.

Key words: written corrective feedback (WCF), error code system, metalinguistic explanation, comprehensive feedback, selective feedback

Introduction

Many ESL/EFL writing teachers considers the provision of written corrective feedback (WCF) on students' essays as one of their major responsibilities. In the past four decades, its effects on L2 writers' grammatical accuracy have been evaluated in various teaching contexts. Some researchers have expressed their reservations about the provision of WCF per se. Krashen (1984) emphasized the importance of comprehensive input and argued that reading, through which learners subconsciously acquire lexical and structural exemplars, facilitates L2 learners' writing abilities more than explicit instruction of grammatical forms. Zamel (1985) pointed out that ESL writing teachers are often unable to provide precise, clear-cut, and consistent corrective feedback and might fail to provide meaningful content-based comments, preoccupied with surface-level grammatical errors. Then, Truscott (1996, 1999) proposed that WCF be completely dispensed with because it does not contribute to language acquisition, imposes a heavy burden on the teacher, and hinders students' effort to produce complex sentences.

However, there is recently a growing consensus that form-focused corrective feedback is a useful instrument of L2 writing instruction. Ferris (1999), Lalande (1982), Bitchener (2008), Bitchener and Knoch (2010), for example, reported that L2 learners who received corrective feedback learned to use grammatical forms more accurately in new pieces of writing. Furthermore, Bitchener (2008), Chandler (2003), Ferris (2010) and Lalande (1982) pointed out that WCF engages L2 learners in deeper cognitive processing and motivates them to learn a self-editing strategy, an indispensable precondition for long-term language acquisition. There has also been a report that many writing teachers, particularly those with long experience or high-level expertise, provide WCF as part of their regular teaching practice (Evans, Hartshorn, & Tuioti, 2010). Overall, the positive effects of corrective feedback on language acquisition seem to outweigh its potential drawbacks.

Now that the importance of corrective feedback provision per se is generally acknowledged, L2 writing teachers are faced with the challenge of selecting the most appropriate feedback type, or combining several different types, for specific learner groups or instructional goals (Bitchener & Storch, 2016; Ferris, 2011; Ferris & Roberts, 2001). The present study, conducted in the form of an action research study (Burns, 2005; Wallace, 1998), evaluated the effects of multiple WCF on learner writing outcome over two years. Two groups of Japanese university students, enrolled in the same EFL course in two consecutive years, received two different versions of EFL writing instruction involving several writing assignments and multiple WCF, and their gains in grammatical accuracy were compared based on the errors that they made before and after treatment. The multiple WCF used in this study comprised comprehensive coded WCF and selective metalinguistic explanation. Both are indirect types of feedback, as opposed to direct error correction which has been utilized effectively in various studies (e.g., Ferris, 1999; Ferris & Roberts, 2001; Van Beuningen, De Jong, & Kuiken, 2012). Coded WCF involves the teacher indicating the error types by using an error code system (e.g., *Art* for article problems or *NF* for noun form errors). Many studies have provided evidence for the positive effects of coded WCF (e.g., Ferris, 2006; Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause & Anderson, 2010; Lalande, 1982). Metalinguistic explanation is another common form of indirect feedback: the teacher explains a target grammatical rule, often referring to sample error instances, and helps learners fix their own errors (e.g., Bitchener & Knoch, 2010; Sheen, 2007; Shintani & Ellis, 2013). In addition to the feedback types, L2 writing teachers also need to decide the scope of feedback. Unfocused, or comprehensive, feedback is provided on all errors that learners make, and focused, or selective, feedback is provided on a few chosen grammatical forms. Selective feedback has been found effective because learners' attention will be focused on the target forms (Sheen, Wright, & Moldawa, 2009), whereas comprehensive WCF has greater ecological validity in that it helps L2 to use all grammatical forms correctly (Ferris, 2010; Ferris, 2011; Liu & Brown, 2015). Again, the most important point is that L2 writing teachers can always utilize a combination of different types, or scopes, of WCF, tailored to a specific

EFL/ESL group (Bitchener & Storch, 2016; Ferris, 1999; Ferris, 2011; Hendrickson, 1980), so that the strengths of one approach will compensate for the weaknesses of the other(s). From a practicality perspective, however, such combined WCF must be administered in the way that the teacher can provide both frequent opportunities for writing and thorough feedback within the limited class time. Although the immediate goal of the present study is to improve the quality of instruction in my own EFL writing course, it is hoped that optimal balancing between feedback provision and writing practice will help improve similar L2 writing courses as well.

When utilizing any type of indirect WCF, it is also important to remember that there are treatable errors and untreatable errors (Ferris, 1999, 2011; see also Brown, 2012; Geiller, 2014). Treatable syntactic or lexical forms are clearly defined forms, and the rules for understanding and producing such forms are self-explanatory. On the other hand, learners may not easily correct untreatable errors by themselves based on the teacher's indirect feedback unless they have already acquired a large set of rules and exemplars and have experience to use them in real time. It is generally more effective to provide corrective feedback on treatable errors. However, there remains a possibility that competent and attentive students will repair their own errors on untreatable forms by utilizing their own experience and outside resources (Geiller, 2014). Therefore, it is worthwhile to analyze their response to corrective feedback on both treatable and untreatable errors.

In the past several years, I have been experimenting with several different versions of combined WCF involving comprehensive coded feedback and selective metalinguistic explanation. The analyses of the past groups' writings have produced evidence for the combined feedback's positive effects on their grammatical accuracy in new pieces of writing (Ogawa, 2018; Ogawa, 2021). However, three problems have surfaced in the process of instructional treatment and data analysis.

First, although some students closely attended to teacher feedback, others failed to do so, or their attention became weaker over time. Not surprisingly, the inattentive students did not seem to benefit from the combined WCF even when the provision of metalinguistic explanation was intensified. The existence of attentive students and those who fail to pay attention have been pointed out by earlier studies (Evans, et al., 2010; Radecki & Swales, 1988), and it is difficult to help those who refuse to utilize the feedback. From a research perspective, an accurate evaluation of multiple WCF's efficacy is difficult unless the data are drawn only from attentive students' compositions.

Second, in my earlier studies, I, as teacher, had required three drafts of each assigned paper, which is standard procedure for ESL and EFL writing courses, and offered form-focused feedback on all three drafts in the hope that they would polish their paragraphs or short essays through repeated revision. However, the problems were that (a) inattentive students, as above mentioned, did not benefit from corrective feedback, regardless of the type or intensity of feedback and (b) some of the attentive students often perfected their paragraphs at second draft,

and there was little room for further improvement in grammatical accuracy. This tendency might be typical of introductory EFL writing courses, as opposed to more advanced essay writing courses or disciplined-based writing tasks on which the teacher provides detailed content-based feedback and students had to respond to the repeated comments. The implication is that it may be more practical to make students submit two drafts for each writing assignment, instead of three, and assign a greater number of writing tasks so that they can practice using target grammatical forms by writing on new topics.

Third, although the target forms were all chosen from the common error types in L2 writing instruction and research (e.g., Ashwell, 2000; Ferris, 2006), the participants in my last study (Ogawa, 2021) happened to make very few errors on some of the selected forms, making it difficult to precisely determine the effects of corrective feedback on certain grammatical forms. Consequently, restricting statistical analyses only to the frequent error types might generate more valid information.

Based upon the above findings, the present study compared the two instructional treatments involving the combined WCF, which required either (a) submission of three drafts for eight writing assignments or (b) submission of two drafts for 12 assignments. The analysis was focused on attentive students' use of correct forms, carefully eliminating those who did not attend to teacher feedback. Then, only the frequent error types were used as variable for statistical analysis. In action research studies, educational purposes take precedence over the data collection and analysis, and this study experienced some procedural limitations. However, the findings can be compared, or combined, with the results of different WCF studies to find commonalities, which might help develop an efficient WCF approach in similar introductory L2 writing courses.

Literature Review

This section reviews earlier studies on (a) the efficacy of multiple WCF, which is the major instructional treatment used in the present project, (b) positive, and potentially negative, aspects of repeated feedback provision, (c) L2 students' attention to teacher feedback, which can influence the efficacy of any type of corrective feedback, and (d) the effects of WCF on different grammatical forms, which have not been fully clarified yet.

Multiple WCF

Each type of corrective feedback has its own purposes and functions, and it might be natural for L2 writing teachers to assume that combining several different types of feedback was likely to enhance L2 writers' grammatical accuracy. Indeed, some of the earlier studies have provided evidence for the multiple WCF's positive effects, but not all have. One frequent form of combination involves direct correction and indirect corrective feedback. Hendrickson (1980) combined various forms of direct correction (e.g., providing correct forms, crossing out unnecessary words, and writing hints for correction) and indirect WCF (e.g., underlining,

circling, using symbols and codes) to improve intermediate Spanish learners' accuracy in writing. Indirect corrective feedback was intended to train L2 learners to notice and correct their own errors, whereas direct correction aided learners with limited linguistic proficiencies who could not self-edit their texts. Likewise, Ferris (1999) proposed provision of indirect WCF for treatable errors (e.g., subject-verb agreement, run-on sentences, and verb form errors) so that students would learn to self-correct their own errors and offered direct correction for untreatable errors (e.g., word choice, idiomatic phrase or sentence structures, missing words, and unnecessary words) as input for acquisition of target forms. It is noteworthy that Hendrickson's (1980) and Ferris's (1999) approaches were both based on learners' proficiencies and the complexity of target structures. Afterward, Bitchener, Young, and Cameron (2005) reported that a combination of explicit written directions for correcting errors and student-teacher individual conferences significantly improved migrant ESL students' accurate use of the definite article and the past tense long-term. Furthermore, Sheen (2007) demonstrated that a combination of direct correction and metalinguistic explanation contributed more to ESL writers' accuracy in writing than direct correction only.

However, multiple WCF may not necessarily be more effective than simpler feedback in all L2 learning contexts. Bitchener (2008) compared the long-term effects of (a) written and oral meta-linguistic explanation, (b) written meta-linguistic explanation only, and (c) error circling on intermediate ESL students' ability to use the English definite and indefinite articles. Although all three types of feedback were more effective than a no-feedback treatment, there was no significant difference between the combined WCF and the individual feedback types. Likewise, Bitchener and Knoch (2008) compared the effects of (a) direct error correction combined with written and oral meta-cognitive explanation, (b) direct correction and written meta-linguistic explanation, and (c) only direct correction on international visa students' and migrant students' ability to use the English articles. All three feedback groups maintained their accurate use of the target forms long-term, but there was no significant difference between combined WCF and direct correction only. Bitchener and Knoch (2010) further compared the effects of (a) written and oral meta-linguistic explanation, (b) written meta-linguistic explanation, and (c) error circling on advanced ESL learners' ability to use the English definite and indefinite articles. The groups that received combined WCF or only written metalinguistic explanation both maintained their accurate use of the articles long-term, implying that written metalinguistic explanation only would suffice for advanced ESL learners and that a combined WCF was not necessary. Consequently, the components of multiple WCF must be carefully selected based upon the students' proficiencies or language-learning experiences and instructional goals.

Repeated Provision of WCF

Related to the advantages or disadvantages of combined WCF is the efficacy of repeated

provision of feedback. It must be first acknowledged that WCF becomes more useful when L2 writers are required to submit a new draft based upon it (Ferris & Hedgcock, 1998, 2014). The requirement of a follow-up revision will urge learners to attend more closely to teacher feedback. However, what has yet to be clarified is the optimal number of drafts to submit. Some studies have encouraged the provision of form-focused feedback on multiple drafts (Ashwell, 2000; Ferris, 1995). Particularly, the practitioners of dynamic WCF, or DWCF, (Evans, Hartshorn, & Strong-Krause, 2011; Hartshorn, et al., 2010; Kurzer, 2017) emphasized the importance of repeated feedback. They engaged L2 learners in short paragraph writing at every class session and continued to provide indirect coded WCF until their texts became error-free. However, there is still no evidence regarding whether L2 learners' grammatical accuracy will increase in proportion to the frequency of WCF provision. Repeated feedback also entails an extra burden on teachers or feedback providers. Messenger, Evans, and Hartshorn (2020), as DWCF advocates, acknowledged that regulating the number of drafts to check is a practical alternative to enhance the manageability of multi-draft writing instruction. In most teaching contexts, if the teacher collects more drafts of each paper, s/he tends to end up assigning fewer writing tasks; thus, one will have to weigh the effects of WCF provision against writing practice itself.

The issue of a possible trade-off between intensive feedback and frequent writing practice has not been systematically investigated, and the results of a few studies that touched on this issue did not concur. Sheen et al. (2009) showed that the ESL students who practiced writing without receiving any WCF achieved a greater gain in grammatical accuracy than those who received comprehensive WCF. On the other hand, Van Beuningen et al. (2012) reported that students who received corrective feedback wrote more accurately in new writings than those who had more opportunities for writing. As mentioned in the introduction, my own study (Ogawa, 2021), which provided multiple WCF in an introductory writing course at a Japanese university, showed that some of the higher-proficiency students did not benefit from the submission of three drafts. Pedagogically, L2 teachers should not waste the limited class time by providing corrective feedback excessively at the expense of writing practice itself. The optimal intensity of feedback needs to be determined for each specific learner group.

Students' Attention

Another factor that can influence WCF's effects on language acquisition is the students' attention to teacher feedback. Surveys of students' perceptions of teacher feedback have indicated that, overall, L2 writers appreciate corrective feedback from their teachers (Cohen, 1987; Ferris, 1995; Hyland, 2011). Hedgcock and Lefkowitz (1994, 1996) reported that foreign language students (learning French, Spanish, or German) were more dependent on, and interested in, form-focused feedback than ESL students who tended to be linguistically more competent and concerned more about the discourse construction or ideational content. However, Radecki and Swales (1988) reported that there were learners who were willing to accept the

teacher's feedback, those who were partially attentive, and those who resisted. Ferris and Hedgcock (1998) stated that "there is tremendous variability in students' ability to benefit from grammar instruction and feedback and to learn self-correction techniques" (p. 201) and pointed out that the degree to which they benefit from such treatment depends on their language learning attitude as well as other individual differences. Likewise, Ferris (2006), who reported coded feedback's positive effects on ESL students' grammatical accuracy, acknowledged that there was a large standard deviation evidencing considerable individual differences. Evans et al. (2010) also expressed their concern about the learners who were not attentive by observing that "[WCF] may be ineffective if the students are not motivated enough to take adequate advantage of the WCF they receive" (p. 64). Thus, it must be remembered that students who fail to pay attention do not benefit from any type of feedback, which, in turn, interferes with researchers' efforts for accurate assessment.

WCF's Effects on Different Grammatical Forms

Another related concern involves what categories of grammatical forms are more amenable to either direct or indirect WCF. Regarding individual target forms, Ferris (2006) reported that comprehensive coded WCF helped to decrease the number of verb errors and lexical errors significantly in ESL students' essays whereas the errors related to nouns, articles, and sentence structures did not decrease significantly. In terms of error categories, one major difference may be found between system learning and item learning. As above mentioned, Bitchener et al. (2005) showed that in-text metalinguistic explanations combined with student-teacher conferences facilitated the acquisition of rule-governed syntactic forms (i.e., the definite article and the past tense) but did not help learners acquire lexical items (i.e., prepositions). On the other hand, Van Beuningen et al. (2012) reported that direct error correction facilitated secondary-school Dutch-as-a-Second-Language learners' acquisition of syntactic forms (e.g., articles, inflections, word order) while coded WCF helped their acquisition of lexical items. Intuitively, one might be tempted to assume that indirect feedback that encourages learners to reflect on their own errors is more effective for system learning, but more evidence is needed to verify such an interpretation.

Another mediating factor may be the complexity of grammatical forms or the learners' ability to deal with them. Ferris (1999) proposed that direct correction can be useful as feedback on complex or idiosyncratic forms that learners are unable to repair in response to indirect WCF. Shintani, Ellis, and Suzuki (2014) also showed that direct correction was more effective for Japanese students' acquisition of hypothetical conditionals (i.e., complex grammatical form) than indefinite article (i.e., simple form), whereas Shintani and Ellis (2013) showed that metalinguistic explanation was more effective on ESL learners' accurate use of English indefinite articles than direct correction. However, Lalande's (1982) study, involving German-as-a-Foreign-Language students, showed that coded WCF was more effective for the acquisition of *case*, one of the most difficult grammatical rules to teach, than direct correction. Thus, although

the complexity of a target grammatical form seems to have influence on the efficacy of WCF, more studies are needed to determine precisely which feedback type is the best for which error category in each specific learning context.

To sum up, although multiple WCF is likely to facilitate L2 students' grammatical accuracy in new writings, the types of feedback to combine must be carefully chosen based on the students' abilities or language learning experiences and the target forms to acquire. Furthermore, the more repetitive provision does not always enhance the feedback's positive effect on L2 learners' accurate use of grammatical forms; it is necessary to determine an optimal balance between intensive feedback and frequent opportunities for writing practice in each instructional setup. In the process of data analysis, it must be remembered that no feedback will be effective if students do not pay attention, and, therefore, the data should be drawn from the attentive students' writing sample. It is also worthwhile to evaluate the effects of corrective feedback on treatable and untreatable errors. Thus, the following research questions were proposed.

Research Question 1: To what extent does the combination of comprehensive coded WCF and selective metalinguistic explanation facilitate attentive students' ability to use treatable grammatical forms in an introductory EFL writing course?

Research Question 2: Is there a trade-off between the feedback intensity and the number of opportunities for writing concerning EFL students' grammatical accuracy in new writings?

Research Question 3: To what extent does coded WCF influence EFL writers' accurate use of treatable and untreatable grammatical forms?

Method

Participants

This study was conducted with two groups of first-year Japanese EFL students enrolled at a prestigious private Japanese university in either 2017 or 2018. They were taking the same EFL course, a requirement at the English department, in two consecutive years. Most participants were English majors, and many of them were preparing to be EFL teachers at Japanese junior-high or high schools. The English majors were divided into six classes of about 25 to take this course, and I taught one class in 2017 and two classes in 2018. One of my 2018 classes included five primary education majors trying to obtain EFL teaching credentials. Their EFL learning backgrounds were approximately the same as the English majors at the beginning of the freshman year, so I decided to include them as participants. The class in 2017, which submitted more revisions, is hereafter referred to as Intensive Feedback Group (see "Instructional Treatment" below). The classes in 2018, which engaged in a greater number of writing tasks, is referred to as Intensive Practice Group. Intensive Feedback Group included 26 students (13 men and 13 women). Intensive Practice Group included 42 students (28 men and 14 women); five of the women were primary education majors. All participants authorized the use of their writings for

the present research study by individually signing a written informed consent form. However, the number of students in Intensive Feedback Group was reduced to 21, and the number of students in Intensive Practice Group was reduced to 29 in the process of data analysis.

Unfortunately, the precise data for the students' English proficiencies were not available. Intensive Practice Group (2018) voluntarily reported their past scores in various standardized English proficiency tests, and those scores were approximately equivalent to B1 or B2 on the CEFR scale. No language proficiency data were available for Intensive Feedback Group (2017), but their level was believed to be about the same as Intensive Practice Group.

The students in both groups had received intensive grammar-focused English education in Japanese high schools. Therefore, their familiarity with explicit grammatical rules and terminology enabled them to respond to the provided error-code feedback and understand metalinguistic explanation. On the other hand, their EFL writing practice in high school had been limited to sentence-level composition, and the majority had virtually no experience of writing in paragraph form before entering the university. Consequently, the influence of prior paragraph or essay writing experience was mostly controlled for, if not entirely.

Instructional Treatment

The course in which the participants were enrolled was an overall EFL course to teach both oral and written English skills. However, it had been structured, and labeled, as a writing course until a few years before, and, thus, a special emphasis was placed on learning to write paragraphs or short essays. During the first half, or one-third, of each class session, they engaged in reading, listening, and speaking practice based upon a piece of written material or a film clip provided by the teacher. Then, the remaining class time was dedicated to writing training. The pertinent EFL course continued for two semesters, extending over an entire academic year. The class met once a week for a 90-minute session, and the total number of class sessions was 30. Class sessions were held in a computer laboratory where students could type and submit their drafts at once when instructed to do so.

In the writing practice section, the class first reviewed a few example paragraphs for the target rhetorical pattern. The textbook was *Get Your Message Across: Writing Communicative Paragraphs* (Jimbo, Elwood, Morita, Watanabe, Yamada, & Yoffe, 2008). The rhetorical patterns covered during the course were: time order, space order, process/direction, cause/effect, exemplification, definition, classification, and comparison/contrast. Then, in order to learn about the sequence markers and other characteristic conventions for each rhetorical structure, they engaged in controlled writing exercises in the textbook. At every second or third session, they wrote the first drafts of an assigned paragraph or essay in place of such exercises. At the following class session (i.e., a week later), I returned the drafts with coded WCF and, subsequently, held a mini-lesson to provide metalinguistic explanation on one major error type per session (see the "Written Corrective Feedback" section below). The class sessions were

conducted in English, but the explanations for important grammatical rules were repeated in Japanese, which was the teacher's and students' first language.

The syllabus, the materials, and the type and scope of corrective feedback for the two groups were the same except for the number of drafts to submit for each assignment and the number of writing tasks assigned. As shown in Table 1, the students in Intensive Feedback Group (i.e., in 2017) were assigned eight free writing tasks during the year (or four per semester), which focused on the eight rhetorical structures above mentioned, and submitted three typed drafts of each assigned paper (hereafter referred to as *intensive feedback*). They wrote, and submitted, the first drafts in the classroom within the timeframe of 30 minutes and finished the second and third drafts at home.

Table 1
Writing Assignments and Tests for 2017 and 2018 Groups

Semester	Intensive Feedback (2017)	Intensive Practice (2018)	Test
Spring		Personal Experience 1 (Task 1)	
	Time Order (Task 1)	Personal Opinion 1 (Task 2)	＝Pretest*
	Space Order (Task 2)	Time Order (Task 3)	
		Space Order (Task 4)	
	Process/Direction (Task 3)	Process/Direction (Task 5)	
	Cause/Effect (Task 4)	Cause/Effect (Task 6)	＝2nd Test
Fall		Personal Experience 2 (Task 7)	
	Exemplification (Task 5)	Exemplification (Task 8)	
	Definition (Task 6)	Definition (Task 9)	
		Classification (Task 10)	
	Classification (Task 7)	Comparison/Contrast (Task 11)	
	Comparison/Contrast (Task 8)	Personal Opinion 2 (Task 12)	＝Posttest

Note. *2018 Group's first writing task on personal experience (self-introduction) was not used as Pretest because it was extremely basic.

Intensive Practice Group was assigned 12 writing tasks (Table 1) and submitted two drafts of each paper (hereafter as *intensive practice*). In addition to the eight rhetorical patterns, they wrote four more paragraphs or short essays about their personal experiences or opinions. Just as Intensive Feedback Group did, Intensive Practice Group wrote the first drafts in the classroom and the second drafts at home.

The major limitation in this study was that the writing topics (or rhetorical patterns) for the pretest and the posttest differed between the two groups. However, this was an inevitable compromise as I decided to increase the number of writing assignments in 2018 in order to expedite the students' overall writing training.

Written Corrective Feedback

Comprehensive error-code WCF. Both groups received error-code WCF on all grammatical errors they made (Appendix A). The only difference was that Intensive Feedback Group received feedback on three drafts of each paper, and Intensive Practice Group received feedback on two

drafts. The error types were indicated by codes written at the error location. Table 2 displays all error types on which error-code feedback was provided; the codes for error types are shown in parentheses. The treatable error types on which metalinguistic explanation was provided in both 2017 and 2018 are underlined. Additionally, short content-based comments were provided on any of the submitted drafts whenever it seemed necessary or appropriate. Then, holistic grades (based on the students' performance in grammar, vocabulary, content, discourse construction, and style) were given on the final drafts of each paper.

Table 2
Error Types on Which WCF Was Provided

Error Category	Error Types
Treatable Syntactic Errors	subject-verb agreement (SV Agr), verb tense (VT), noun form (NF), articles (Art), word form (WF), mechanics (Mec), number-related errors (#), run-on sentence (Run-on), fragment (Frag), pronoun agreement (Pro Agr), word order (WO), voice (Vo), spelling (Sp)
Treatable Stylistic Errors	informal usage (Inf), sentence-initial conjunction (In-Conj), redundant (Red), ambiguous (Amb), awkward (Awk)
Untreatable Errors	wrong word (WW), word missing (WM), sentence structure (S/Str), idiomatic expression (Id), unnecessary word (UnW)

Note. Adapted from Ogawa (2021, p. 104). *Informal usage* included contractions, sentence-final interjections, colloquial expressions, etc. *Mechanics* included punctuation, capitalization errors, etc.

Metalinguistic explanation. To enhance the effects of comprehensive coded WCF, mini-lectures were held during class sessions to provide metalinguistic explanation on some of the major treatable error types (Appendix B). One mini-lecture took about 10–15 minutes. In 2017, metalinguistic explanations were focused on six syntactic error types and two stylistically undesirable forms (i.e., the underlined items in Table 2). Metalinguistic explanation was provided on all the eight forms in the spring, and each target form was reviewed two more times in the fall.

In 2018, metalinguistic explanation was provided on the eight forms above and three additional forms (pronoun agreement, voice, and word form); one mini-lecture was administered for each form in the spring; all the 11 forms were reviewed one more time in the fall of 2018. It must be noted that the three frequent treatable forms that were analyzed for the effects of combined WCF (i.e., *verb tense, noun form*, and *article*) received metalinguistic explanation in both 2017 and 2018.

A mini-lecture was conducted using a handout and PowerPoint slides. First, a handout showing several example sentences that contained the target error was distributed, and the students were instructed to work in small groups to identify the errors. Then, the example sentences were individually displayed on a large screen one after another, and students were randomly asked to correct the problematic parts. In the end, I orally explained the target grammatical rule with a summarized written explanation displayed on the large screen.

Analysis

Selection of attentive students. Students who did not attend to teacher feedback were

eliminated through a multi-tiered process. First, the students who did not submit at least three-fourths of the required drafts were deleted from the sample because they had neither experienced enough writing practice nor received enough feedback. Second, the students who kept making paragraph-structure or formatting errors (e.g., stand-alone one-sentence paragraphs, indentation problems, the wrong font size and type) three times despite the explicit instruction at the beginning of the first semester, and my repeated written comments, against them were eliminated. Third, the students' undesirable behavior in the classroom were checked and recorded (e.g., incessant private conversations during a class) as evidence of lack of attention. As it turned out, all the students who were removed from the data set violated more than one of the three conditions. Although these cut-off standards were somewhat arbitrary, I decided that these objective criteria were more valid than, for example, a questionnaire survey requesting students to self-declare the degree of their attention to teacher feedback.

Comparison between Intensive Feedback and Intensive Practice Groups. A quasi-experimental design was used to evaluate the two treatment groups' improvement in writing accuracy. The target grammatical forms for analysis were selected based on the frequencies of errors. The total numbers of errors found in the six sets of drafts (i.e., three sets for each group) were studied, and a clear discrepancy was found between the seven chosen forms and all the others as shown in Table 3. Although the numbers of *ambiguous* and *awkward* errors were large, these two were excluded from the analysis because they involved various types of structural, ideational, and rhetorical problems that could not be classified into any of the clearly defined types.

The statistical analyses consisted of two parts. The first part evaluated the combined WCF's effects on the three most frequent treatable forms. The second part evaluated the coded WCF's effects on the most frequent treatable error type (*mechanics*) and the three most frequent untreatable error types (*wrong word, word missing*, and *sentence structure*).

Table 3
The Numbers of Errors for Treatable and Untreatable Grammatical Forms

Treatable Forms	# of Errors	Errors/100 Words	Untreatable Forms	# of Errors	Errors/100 Words
Noun form	144	102.8	Mechanics	192	137.2
Article	86	62.0	Sentence Structure	160	115.4
Verb tense	66	46.2	Wrong Word	136	94.6
Subject-Verb	51	35.1	Word Missing	87	61.9
Informal	37	25.3	Ambiguous	62	45.4
Initial Conjunction	23	17.7	Awkward	62	39.0
Run-on	15	11.5	Number	43	28.5
Fragment	4	2.5	Word Form	38	27.8
			Pronoun Agreement	27	21.7
			Unnecessary Word	17	11.1
			Spelling	16	11.1
			Word Order	12	9.1
			Voice	12	8.5
			Redundant	6	4.0
			Equal Form	3	2.1

The data were drawn from three sets of students' writing sample (See Table 1). The two groups' first drafts of: (1) the first major paper assigned, serving as pretest, (2) the last paper in the first semester (i.e., the fourth assignment for Intensive Feedback Group and the sixth assignment for Intensive Practice Group) serving as the second test; and (3) the last paper serving as posttest. The students' performance at the second test was believed to reflect their improvement after writing and receiving WCF for one semester. Then, the participants' gains (i.e., fewer errors), or losses (i.e., more errors), in error scores over the one-year period were meant to reflect their overall improvement or regression in their accurate use of the target forms as a function of either of *intensive feedback* or *intensive practice* treatment.

The numbers of learner errors on each target form in these drafts were counted before returning them and were checked again at the end of each semester. Error types, instead of tokens, were counted. For every case of discrepancy, the types, and the numbers, of errors were reexamined for a third time, and the data after the final checking were used for statistical analysis.

The length of a paragraph differed from person to person and from assignment to assignment. Therefore, the numbers of errors were normalized (Biber, Conrad, & Reppen, 1998), and the *normed* frequencies were used as error scores. Biber et al. proposed that the number of errors be divided by the average number of words per essay, but this study employed a simplified procedure of computing the number of errors per 100 words. Table 4 shows the average length of students' writing for each writing test.

Table 4
The Length of Students' Writing

	Pretest		2nd Test		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*
Intensive Feedback	144.62	45.53	170.62	42.34	152.81	25.23
Intensive Practice	122.45	25.78	131.76	26.68	151.66	29.96
Two Groups Combined	131.76	36.72	148.08	38.91	152.14	27.80

Note. Intensive Feedback Group, $n=21$; Intensive Practice Group, $n=29$.

It was also confirmed that there was no trade-off between accuracy and productivity. Not only that, but an ANOVA indicated that there was significant improvement in the length of students' writing between the test means, $F(2, 96)=5.73$, $p=.004$, $\eta^2=.11$; there was a significant gain from pretest to the second test, $t(49)=2.53$, $p=.02$, $d=.43$ (small) and from pretest to posttest, $t(49)=3.33$, $p=.002$, $d=.63$ (medium). Individually, Intensive Feedback Group's productivity increased significantly from pretest to the second test, $t(20)=2.41$, $p=.026$, $d=.59$ (medium); Intensive Practice Group's productivity increased significantly from pretest to posttest, $t(28)=4.40$, $p=.001$, $d=.83$ (large), and from the second test to posttest, $t(28)=2.70$, $p=.01$, $d=.70$ (medium).

Then, a two-way ANOVA was performed to evaluate the effects of *intensive feedback* and *intensive practice* on the attentive students' ability to use the three forms (verb tense, noun form, and article) accurately in new writings. The between-subjects factor was *treatment* with

two levels (*intensive feedback* and *intensive practice*), and the within-subjects factor was *test* with three levels (pretest, second test, and posttest). The dependent variables were the two groups' error scores at the three tests. When the results of two-way ANOVAs indicated a significant *test* main effect, one-way ANOVAs were conducted for each group, which were then followed by paired-samples *t*-tests between writing tests. When there was a significant interaction effect, independent-samples *t*-tests were performed to determine the difference between the two treatment groups at each test. Partial eta squared effect sizes were calculated for ANOVAs, and Cohen's *d* effect sizes were calculated for *t*-tests.

Through a similar procedure, a three-way ANOVA was conducted to evaluate the effects of coded WCF on the most frequent *treatable* error type (mechanics) and *untreatable* error types (wrong word, word missing, and sentence structure) at the three writing tests. The between-subjects factor was *treatment*, the first within-subjects factors was *test*, and the second within-subjects factor was *form* with two levels (treatable and untreatable).

The level of significance was set at $\alpha = .05$ for all statistical analyses. Holm's sequential Bonferroni adjustment was used to control for Type I errors (which occur when the statistical test results indicate that there is a genuine effect when, in fact, there is not) in all post hoc tests.

Results

The Effects of Combined WCF on Three Treatable Errors

Table 5 displays the means and standard deviations of the students' error scores for three frequent treatable error types that received combined WCF (*verb tense, noun form*, and *article*). Intensive Feedback Group's and Intensive Practice Group's respective error means at each test (i.e., *pretest*, the *second test*, and *posttest*) are shown individually. An error-mean decrease over the tests means an improvement in the students' accurate use of the target form.

Two-Way ANOVAs were performed to evaluate the extent to which the combined WCF facilitated the two treatment groups' accurate use of the three forms. Before conducting the ANOVAs, I checked the data set to ensure that it met the needed requirement for the tests. There was no significant difference between the two groups' error means at the pretest for any of the three forms. Regarding *verb tense*, Mauchly's sphericity test was not passed, and, therefore, the Greenhouse-Geisser adjustment was applied to the interpretation of the results.

Table 5
The Number of Students' Errors per 100 words (Verb Tense, Noun Form, Article)

Group	Form	Pretest		2nd Test		Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*
2017	Verb Tense	0.79	0.81	0.00	0.00	0.10	0.32
	Noun Form	0.40	0.51	0.52	0.54	0.84	1.14
	Article	0.37	0.58	0.33	0.48	0.50	0.66
2018	Verb Tense	0.46	0.81	0.17	0.34	0.32	0.53
	Noun Form	0.62	1.02	1.16	1.71	0.50	0.74
	Article	0.56	0.66	0.33	0.41	0.38	0.48

Note. Intensive Feedback Group (2017), $n = 21$; Intensive Practice Group (2018), $n = 29$.

The ANOVA results (Table 6) showed that combined WCF had a significantly positive *test* effect on *verb tense*. The test means (i.e., the average of the two groups' means) at the three writing tests were computed respectively, and paired-samples *t*-tests were performed. The results revealed that there was a significant improvement (i.e., decrease in the mean) from the pretest (M=.60, SD=.82) to the second test (M=.10, SD=.27), $t(49)$=-4.26, p=.001 (α=.017 after Holm's sequential Bonferroni adjustment), d=.82 (large) and from the pretest to the posttest (M=0.23, SD=.46), $t(49)$=-2.54, p=.014 (α=.025), d=.57 (medium). The ANOVA also indicated that the interaction effect was almost significant, p=.051; consequently, paired-samples *t*-tests were conducted for the two groups individually. Intensive Feedback Group improved significantly from the pretest (M=.79, SD=.81) to the second test (M=.00, SD=.00), $t(20)$=-4.48, p=.001 (α=.017), d=1.38 (large) and from the pretest to the posttest (M=.10, SD=.32), $t(20)$=-3.33, p=.003 (α=.025), d=1.12 (large). On the other hand, there was no significant change over the tests for Intensive Practice Group, although the improvement from the pretest (M=.46, SD=.81) to the second test (M=.17, SD=.34) approached significance, $t(28)$=-1.98, p=.058, d=.47 (small). That is, *intensive feedback* contributed to the acquisition of *verb tense* more greatly than *intensive practice*, whereas *intensive practice* itself had a nearly significant positive effect during the first semester.

Regarding *noun form*, Intensive Feedback Group's error mean gradually *rose* from the pretest to the second test to the posttest, while Intensive Practice Group's error mean rose from the pretest to the second test and, then, fell from the second test to the posttest. However, neither the *test* or *treatment* main effect nor the interaction effect was significant. As for the *article* errors, both treatment groups' error means decreased from the pretest to the second test and, then, increased from the second test to the posttest, but there was no significant main or interaction effect for this form, either. That is, regardless of the instructional plans (i.e., emphasizing either corrective feedback or opportunities for writing), the corrective feedback did not significantly help reduce the number of *noun form* or *article* errors.

Table 6
ANOVA Results for the Combined WCF's Effects on Three Treatable Error

Form	Factor	df	SS	MS	F	p	η^2
Verb tense	Test	1.37	7.84	5.73	12.24	0.001	0.20
	Test x Treatment	1.37	2.26	1.65	3.54	0.051	0.07
	Residual	65.69	30.73	0.47			
	Treatment	1	0.02	0.02	0.06	0.80	0.00
	Residual	48	14.04	0.29			
Noun form	Test	2	2.72	1.36	1.23	0.30	0.03
	Test x Treatment	2	5.94	2.97	2.68	0.07	0.03
	Residual	96	106.23	1.11			
	Treatment	1	1.09	1.09	0.92	0.34	0.02
	Residual	48	56.62	1.18			
Article	Test	2	0.52	0.26	0.89	0.41	0.02
	Test x Treatment	2	0.60	0.30	1.03	0.36	0.02
	Residual	96	27.91	0.29			
	Treatment	1	0.01	0.01	0.04	0.85	0.00
	Residual	48	15.76	0.33			

The Effects of Coded WCF on Four Forms

Table 7 displays the descriptive statistics for four forms on which only coded WCF was provided. The four target forms included one treatable error type (*mechanics*) and three untreatable error types (*sentence structure, wrong word*, and *word missing*). Each group's mean and standard deviation at the pretest, the second test, and the posttest were calculated. The same table also contains the means and standard deviations for the set of three untreatable errors averaged as one variable.

Table 7
The Number of Students' Errors per 100 words (Mechanics, Sentence Structure, Wrong Word, Word Missing)

Group	Form	Pretest		2nd Test		Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*
2017	Mechanics	1.35	0.57	0.39	0.45	0.84	0.81
	Sentence Structure	0.38	0.67	0.69	0.72	0.40	0.65
	Wrong Word	0.82	0.78	0.55	0.53	0.55	0.20
	Word Missing	0.45	0.55	0.14	0.31	0.40	0.63
	SS＋WW＋WM/3	0.55	0.38	0.46	0.28	0.45	0.43
2018	Mechanics	1.20	1.13	0.80	0.81	0.86	0.87
	Sentence Structure	0.77	0.83	1.29	1.14	0.85	0.97
	Wrong Word	0.40	0.67	0.75	0.92	0.73	0.77
	Word Missing	0.56	0.71	0.38	0.56	0.48	0.58
	SS＋WW＋WM/3	0.57	0.35	0.61	0.45	0.69	0.45

Note. Intensive Feedback Group (2017), $n=21$; Intensive Practice Group (2018), $n=29$.

Two-way ANOVAs were conducted to evaluate the effects of coded WCF on Intensive Feedback and Intensive Practice Groups' accurate use of the four individual forms. Prior to the ANOVAs, the two groups' pretest means and the Sphericity test results were checked. There was no significant difference between their means at the pretest for any of the four forms. Mauchly's sphericity test was not passed for *mechanics*; thus, the Greenhouse-Geisser adjustment was applied to the interpretation of the results for this form. The ANOVA results are displayed in Table 8.

The main *test* effect for *mechanics* was significant, and the effect size was large. To follow it up, the test means (i.e., the average of the two groups' means) at the three tests were computed, and paired-samples t-tests were conducted between the three writing tests. The second-test mean ($M=.63, SD=.71$) was significantly smaller than the pretest mean ($M=1.26, SD=.93$), $t(49)=-3.84$, $p=.001$ ($\alpha=.017$), $d=.76$ (medium), indicating a significant improvement. The test mean decreased from the pretest to the posttest ($M=.85, SD=.84$), although the difference narrowly missed the significant level after Holm's sequential Bonferroni adjustment, $t(49)=-2.30$, $p=.026$ ($\alpha=.025$), $d=.46$ (small). There was no significant difference between the second test and the posttest. That is, a positive effect of coded WCF was recognized in the first semester, but it was not maintained during the second semester. The treatment effect was not significant, showing that the choice between intensive feedback or intensive practice did not affect the efficacy of coded WCF.

Table 8

ANOVA Results for the Coded WCF's Effects on One Treatable Error and Three Untreatable Errors

Form	Factor	df	SS	MS	F	p	η²
Mechanics	Test	1.73	11.53	6.68	9.27	0.001	0.16
	Test x treatment	1.73	1.95	1.13	1.56	0.22	0.03
	Residual	82.78	59.69	0.72			
	Treatment	1	0.31	0.31	0.38	0.54	0.01
	Residual	48	39.48	0.82			
Sentence Structure	Test	2	5.14	2.57	3.23	0.04	0.06
	Test x treatment	2	0.28	0.14	0.18	0.84	0.01
	Residual	96	76.51	0.80			
	Treatment	1	8.39	8.39	11.94	0.001	0.20
	Residual	48	33.71	0.70			
Wrong Word	Test	2	0.05	0.03	0.05	0.95	0.001
	Test x treatment	2	3.04	1.52	2.84	0.06	0.001
	Residual	96	51.42	0.54			
	Treatment	1	0.01	0.01	0.01	0.91	0.001
	Residual	48	26.89	0.56			
Word Missing	Test	2	1.58	0.79	2.36	0.10	0.05
	Test x treatment	2	0.19	0.09	0.28	0.76	0.01
	Residual	96	32.28	0.34			
	Treatment	1	0.74	0.74	2.22	0.14	0.04
	Residual	48	15.93	0.33			

The ANOVA results for *sentence structure* indicated that the *treatment* main effect was significant, evidence that intensive feedback was more effective for the students' accurate use of this form than intensive practice. The test main effect was also significant; however, the post hoc tests showed that the second-test mean (M=1.04, SD=1.02) was *higher* (i.e., a sign of regression) than the pretest mean (M=.61, SD=.79), narrowly missing the significant level, $t(49)$=2.29, p= .026 (α=.017), d=.47 (small). On the other hand, the posttest mean (M=.66, SD=.87) was lower than the second-test mean, which was also almost significant, $t(49)$=$-$2.28, p=.027 (α=.025), d= .40 (small). The ANOVA for *wrong word* and *word missing* showed that there was no significant main or interaction effect for either of these forms. Overall, the statistical test results supported the view that coded WCF did not facilitate the students' accurate use of untreatable errors, regardless of the intensity of feedback or the number of opportunities for new writings.

Finally, a three-way ANOVA was conducted to compare the effects of coded WCF on *mechanics* (treatable form) and the *sentence structure, wrong word*, and *word missing* combined (i.e., the three untreatable forms averaged and used as one variable; hereafter referred to *combined*). The results (Table 9) indicated that the *form* main effect, the *test* main effect, and the *form* x *test* interaction were significant.

Table 9
Three-Way ANOVA Results (Mechanics vs. the Untreatable Errors Combined)

Factor	df	SS	MS	F	p	η^2
Form	1	7.45	7.45	15.73	0.001	0.25
Form x Treatment	1	0.22	0.22	0.47	0.50	0.01
Residual	48	22.73	0.47			
Test	1.60	4.71	2.95	5.36	0.01	0.10
Test x Treatment	1.60	2.34	1.46	2.66	0.09	0.05
Residual	76.73	42.19	0.55			
Form x Test	2	6.97	3.49	10.38	0.001	0.18
Form x Test x Treatment	2	0.26	0.13	0.38	0.68	0.01
Residual	96	32.25	0.34			
Treatment	1	1.58	1.58	3.04	0.09	0.06
Residual	48	24.99	0.52			

To follow up the significant *form* x *test* interaction effect, a one-way ANOVA was conducted on each level of the *form* factor (*mechanics* and *combined*), which was followed by pairwise post hoc tests. Then, the test means (i.e., the average of the two groups' error means at each test) were compared between the two forms. Prior to the statistical analyses, the data set was checked to ensure that it met the needed requirements for the test. The *mechanics* error mean ($M=1.26$, $SD=.93$) and the *combined* error mean ($M=.56$, $SD=.36$) were significantly different at the pretest, $t(49)=5.12$, $p=.001$, $d=.99$ (large); consequently, the difference means from test to test were used as variables. For both forms, Mauchly's sphericity was not passed, and, therefore, the Greenhouse-Geisser adjustment was applied. Table 10 displays the difference means between tests for the two forms.

Table 10
Difference Means Between Tests

Tests	M	SD
Mechanics Test 1—Mechanics Test 2	−0.64	1.17
Combined Test 1—Combined Test 2	0.10	0.59
Mechanics Test 1—Mechanics Test 3	−0.41	1.27
Combined Test 1—Combined Test 3	0.03	0.57
Mechanics Test 2—Mechanics Test 3	0.22	0.89
Combined Test 2—Combined Test 3	−0.07	0.53

Note. N=50.

The ANOVA results for *mechanics* showed that there was a significant *difference mean* main effect, $F(1.34, 65.55)=12.19$, $p=.001$, $\eta^2=.20$. The post hoc tests revealed that the improvement (i.e., a decrease in the error scores) from the pretest to the second test ($M=−.64$, $SD=1.17$) was greater than the change (which was actually a regression) from the second test to the posttest ($M=.22$, $SD=.89$), $t(49)=−3.68$, $p=.001$, $d=.83$ (large). The gain in accuracy from the pretest to the posttest ($M=−.41$, $SD=1.27$) was significantly greater than the change from the second test to the posttest (which was also a regression), $t(49)=−3.84$, $p=.001$, $d=.58$ (medium). The one-way ANOVA for the *combined* error types showed that there was no

significant difference between the difference means, $F(1.23, 60.03)=1.44$, $p=.24$, $\eta^2=.03$.

Subsequently, independent-samples t-tests were performed to compare the difference means between the two forms during each period. The *mechanics'* gain ($M=-.64$, $SD=1.17$) was significantly greater than the *combined*'s gain ($M=.10$, $SD=.59$) between the pretest and the second test, $t(49)=-4.63$, $p=.001$, $d=.80$ (large); likewise, the *mechanics'* gain ($M=-.41$, $SD=1.27$) was greater than the *combined*'s gain ($M=.03$, $SD=.57$) between the pretest and the posttest, $t(49)=-2.52$, $p=.015$, $d=.45$ (small). The overall test results confirmed that treatable errors are more amenable to corrective feedback than untreatable errors.

Discussion

The first research question addressed the extent to which the combination of comprehensive coded WCF and selective metalinguistic explanation facilitated Japanese EFL students' ability to use three treatable grammatical forms. The present study served as a replication study to support Ogawa (2021), restricting the participants to attentive students. The results generated additional evidence that the feedback helped the students use verb tense accurately and that it did not have a positive effect on the use of noun forms and articles which have also been commonly classified as treatable forms. There are two possible explanations for this seemingly contradictory results. First, as stated in Ogawa (2021), the students might not have been motivated to attend to local errors which do not impede communicating a message. Second, the proper use of noun forms requires difficult decisions about the characteristics of nouns (Ferris & Hedgcock, 2014), including the distinction between countable and non-countable nouns or between concrete and abstract nouns. The correct noun forms are also interrelated with the use of definite and indefinite articles. Coincidentally, the mechanical errors are local errors, but the analysis of the efficacy of coded WCF showed that the students' accurate use of mechanical conventions improved significantly. Consequently, among the two interpretations, the latter might be the more likely cause as far as the pertinent student group is concerned.

Regarding the English definite and indefinite articles, they serve diverse grammatical functions including not only anaphora but generic reference (e.g., "the Germans are good musicians," "the computer is one of the greatest inventions in the 21st century"), collective units (e.g., "the United States," "the Rocky Mountains"), the one and only object in the world (e.g., "the sun," "the universe"), and immediately identifiable objects (e.g., "I'm going to the gym"). There are also a countless number of idiomatic usages (e.g., "Do you have the time?" as against "Do you have time?"; "I just caught a cold; It wasn't the flu"). Although the studies that investigated the issue of anaphora alone reported that the articles were treatable (e.g., Bitchener, 2008; Bitchener & Knoch, 2008, 2010), the present study analyzed all error instances involving the articles, and, thus, the results indicating that article errors were not amenable to multiple WCF were not surprising, another point which was indicated in Ogawa (2021). The effects of corrective feedback may possibly be enhanced if more detailed metalinguistic explanations are provided, at

least, on the major functions of articles, if not on all idiomatic usages. Furthermore, there is a strong possibility that comprehensive coded WCF, as opposed to metalinguistic explanation, will play a major role because the subcategories of article errors can be individually attended to. Nonetheless, the overall implication is that the classification between treatable and untreatable forms ought to be carefully defined, or interpreted, in such contexts as Japanese EFL courses where the learners' first language does not have the articles and plural suffixes.

The second research question was related to the issue of whether there is a trade-off between feedback intensity and opportunities for writing concerning the combined WCF's effects on the acquisition of treatable forms. One notable finding is that intensive feedback had a stronger effect on the students' ability to use verb tense than intensive practice, producing some evidence of trade-off. That is, in L2 writing classes where grammatical accuracy is required, repeated feedback may be an indispensable condition. It is also noteworthy that Intensive Practice Group's error mean for verb tense decreased on the second test and, then, increased to some extent on the posttest. The same tendency was observed about the coded WCF's effects on mechanical errors. When students engage in more practice, their attention may be drawn to other aspects of their writing, rather than the teacher's previous explanation. Consequently, as the earlier studies (Ashwell, 2000; Evans et al., 2011; Ferris, 1995; Ferris & Hedgcock, 2014; Hartshorn et al., 2010; Kurzer, 2017) suggested, it is very important to remind students of the major rules repeatedly throughout the entire course in one way or another even if the teacher decides to reduce the number of drafts to check.

The third research question concerned the extent to which coded WCF influences EFL writers' accurate use of treatable and untreatable grammatical forms. The coded WCF had a positive effect on mechanical errors (i.e., a treatable error type), but not on incorrect sentence structures, wrong words, and missing words (i.e., untreatable errors). Although mechanical errors can be categorized as typical local errors which do not interfere with the communication of a message, the number of error instances was by far the greatest in the writing sample collected in this project. Thus, the fact that coded WCF had a positive effect on both groups' accurate use of this form is encouraging news to teachers who spend a large amount of time providing feedback. On the other hand, the category labeled as *sentence structure* was a composite of many different grammatical rules. Consequently, as is the case of the articles, future studies should be designed to focus on several specific sentence-structure rules (e.g., adverbial clauses, relative or dependent clauses, hypothetical conditionals) and evaluate students' reaction to feedback on the selected discrete forms.

The coded WCF's strong efficacy itself is meaningful. As above mentioned, coded feedback was provided comprehensively on every draft, and it was, in a practical sense, tailored to individual students' needs, whereas metalinguistic explanation had to focus on a limited number of grammatical rules. Furthermore, the comprehensive coded feedback was provided equally to the two groups (i.e., the total number of drafts on which they received feedback was the same),

and Intensive Practice Group did not suffer any disadvantage in this respect. Multiple WCF tends to be useful, but the roles of individual feedback types should also be recognized and fully utilized to serve different purposes in L2 writing instruction.

Conclusion

The present study strove to find additional evidence for multiple WCF's positive effects on EFL students' use of treatable grammatical forms and, further, probed into a possible trade-off between the two treatments emphasizing either intensive feedback or more opportunities for writing. The results showed that the multiple WCF had a positive effect on verb tense, a treatable form, and the *intensive feedback* treatment was more effective than the *intensive practice* treatment. The feedback did not contribute to the acquisition of noun forms and articles regardless of the treatment groups. The evaluation of the effects of coded WCF alone on the students' ability to use treatable and untreatable forms showed that the feedback was effective for mechanical errors (i.e., treatable error type), but not for untreatable forms.

The pedagogical implications were as follows. First, although the categorization into treatable and untreatable forms is very useful, it might be better to subcategorize the treatable errors for Japanese EFL writing courses when dealing with the complex grammatical rules and idiomatic usages that are absent in the Japanese syntactic system. The article system is a typical example and can be treated as a highly untreatable form. For this form, the teacher may be advised to give multiple mini-lectures, focusing on one specific usage at a time and using several class sessions. If it is impossible to cover all the diverse usages in mini-lectures, comprehensive coded feedback on individual students' drafts might be fully utilized to attend to the idiosyncratic error types. Second, it seems necessary to formulate a new strategy to reinforce the form-focused feedback in the two-draft teaching plan. Now that the study confirmed the important role of repeated feedback, one solution might be to determine a perfect timing for repeating the important grammatical rules, instead of mechanically repeating metalinguistic explanation on the same form at regular intervals. Such a decision might be based on some cognitive psychology studies on studying and testing effects (e.g., Karpicke & Roediger's, 2007; Roediger & Karpicke, 2006). Third, it should be noticed that the group that was assigned more writing tasks in the present project learned to write an increasingly longer paragraph or essay from the pretest to the posttest, but the length of their writing at the final test were not longer than that of the group receiving feedback on three drafts. In addition to the number of writing tasks to be assigned, the teacher must systematically guide students to produce lengthier writings, which might enhance the effects of writing practice by enabling them to use the target forms more often in each paper.

The methodological implications for future studies were as follows. First, noun forms and articles involve complex grammatical rules and can be regarded as untreatable forms in a practical sense; furthermore, the errors related to these two forms overlap. Consequently,

future studies might be conducted on individual sub-rules of either grammatical form. Second, as acknowledged in the method section, the major limitation of this study was that the writing topics for the pretest and the posttest did not match between the two treatments. Such discrepancy must be corrected in order to precisely assess the effects of any form of corrective feedback on L2 students' writing accuracy.

Although the design for data collection and assessment was not perfect in this action research study, the results have provided additional evidence for multiple WCF's and coded WCF's positive effects on EFL writers' grammatical accuracy. The study also demonstrated the importance of repeated feedback, which should be administered even though frequent opportunities for writing practice are necessary. Furthermore, it shed light on the next step of research for the betterment of my writing course: i.e., to explore ways to guide students to produce longer and more sophisticated essays without sacrificing grammatical accuracy. It is also hoped that the findings will be utilized by other L2 writing teachers who aim to facilitate their students' overall writing proficiencies involving accuracy and productivity.

References

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*(3), 227-257. https://doi.org/10.1016/S1060-3743(00)00027-8

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102-118. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*(3), 409-431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2009a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*(2), 193-214. https://doi.org/10.1093/applin/amp016

Bitchener, J., & Knoch, U. (2009b). The value of a focused approach to written corrective feedback. *ELT Journal, 63*(3), 204-211. https://doi.org/10.1093/elt/ccn043

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217. https://doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development.* Bristol, UK: Multilingual Matters.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191-205. https://doi.org/10.1016/j.jslw.2005.08.001

Brown, D. (2012). The written corrective feedback debate: Next steps for classroom teachers and practitioners. *TESOL Quarterly, 46*(4), 861-867. https://doi.org/10.1002/tesq.63

Burns, A. (2005). Action research: An evolving paradigm? *Language Teaching, 38*(2), 57-74.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and

fluency of L2 student writing. *Journal of Second Language Writing, 12*(3), 267–296. https://doi.org/10.1016/S1060-3743(03)00038-9

Cohen, A. D. (1987). Student processing of feedback on their composition. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 57–69). New York, NY: Prentice Hall.

Evans, N. W., Hartshorn, K. J., & Tuioti, E. A. (2010). Written corrective feedback: Practitioners' perspectives. *International Journal of English Studies, 10*(2), 47–77. https://doi.org/10.6018/ijes/2010/2/119191

Evans, N. W., Hartshorn, K. J., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System, 39*(2), 229–239. https://doi.org/10.1016/j.system.2011.04.012

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*(1), 33–53. https://doi.org/10.2307/3587804

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1–11. https://doi.org/10.1016/S1060-3743(99)80110-6

Ferris, D. R. (2006). Does error feedback help student writers?: New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). New York: Cambridge University Press.

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32*(2), 181–201. https://doi.org/10.1017/S0272263109990490

Ferris, D. R. (2011). *Treatment of error: In second language student writing* (2nd ed.). An Arbor: The University of Michigan Press.

Ferris, D. R., & Hedgcock, J. S. (1998). *Teaching ESL composition: Purpose, process, and practice*. Mahwah. NJ: Lawrence Erlbaum Associates.

Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice*. New York, NY: Routledge.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Geiller, L. (2014). How EFL students can use Google to correct their "untreatable" written errors. *The EUROCALL Review, 22*(2), 26–45.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44*(1), 84–109.

Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing, 3*(2), 141–163. https://doi.org/10.1016/1060-3743(94)90012-4

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal, 80*(3), 287–308. https://doi.org/10.1111/j.1540-4781.1996.tb01612.x

Hendrickson, J. M. (1980). The treatment of error in written work. *The Modern Language Journal, 64*(2), 216–221. https://doi.org/10.1111/j.1540-4781.1980.tb05188.x

Hyland, F. (2011). The language learning potential of form-focused feedback on writing: Students' and teachers' perceptions. In R. M. Manchon (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 159–179). Amsterdam: John Benjamins.

Jimbo, H., Elwood, K., Morita, A., Watanabe, Y., Yamada, S., & Yoffe, L. (2008). *Get your message across:*

Writing communicative paragraphs. Nan'un-do.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Krashen, S. D. (1984). *Writing: Research, theory and applications*. Torrance, CA: Laredo Publishing Co., Inc.

Kurzer, K. (2017). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly, 52*(1), 5–33. https://doi.org/10.1002/tesq.366

Lalande, J. F., II. (1982). Reducing composition errors: An experiment. *The Modern Language Journal, 66*(2), 140–149. https://doi.org/10.2307/326382

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing, 30*, 66–81. https://doi.org/10.1016/j.jslw.2015.08.011

Messenger, R. A., Evans, N. W., & Hartshorn, K. J. (2020). Managing dynamic written corrective feedback: Perceptions of experienced teachers. *Journal of Response to Writing, 6*(1), 108–138. https://journalrw.org/index.php/jrw/article/view/139

Ogawa, Y. (2018). The effects of unfocused error-code WCF and focused metalinguistic explanation on EFL writing accuracy. *Gakuen, 930*, 16–36.

Ogawa, Y. (2021). Written corrective feedback in EFL: Combining error codes and metalinguistic explanation. *Journal of Response to Writing, 7*(1), 94–127.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Radecki, P. M., & Swales, J. M. (1988). ESL student reaction to written comments on their written work. *System, 16*(3), 355–365. https://doi.org/10.1016/0346-251X(88)90078-4

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(4), 556–569. https://doi.org/10.1016/j.system.2009.09.002

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing, 22*(3), 286–306. https://doi.org/10.1016/j.jslw.2013.03.011

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning, 64*(1), 103–131. https://doi.org/10.1111/lang.12029

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Truscott, J. (1999). The case for "the case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*(2), 111–122. https://doi.org/10.1016/S1060-3743(99)80124-6

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19*(1), 79–102.

Appendix A
Unfocused Error-Code Feedback on a Sample Paragraph

In this paragraph, I'll (Inf) introduce (WW) my typical day. I get up at four in the morning to work as a part-time job (WW). After working at a convenience store for 4 (Mec) hours, I head to the university. I usually had (VT) two or three classes to attend. For example, on Monday, a French class in the first period and a PE class in the third period. (Frag) Between classes, I talk to my friends or eat lunch together, spending time with good friends is a good diversion (Run-on). After the classes, I plactice (Sp) *karate* from five (WM) seven p.m.

(*Note*. This sample was improvised, instead of using an actual student's writing draft, in order to protect the participants' privacy.)

Appendix B
A Sample Handout for Metalinguistic Explanation
(Target grammatical form: Noun form)

Instruction: Find and correct grammatical errors in the following sentences.

1. We will find opportunities to work in global situation [→ a global situation/global situations].
2. If you speak English, you can communicate with foreign diplomat or tourist [→ foreign diplomats or tourists].
3. You can enjoy literature in different language [→ languages].
4. Language skill is [→ language skills are] useful for widening our world knowledge.
5. It is said that a lack of sleep tends to result in lifestyle-related disease [→ a lifestyle-related disease/lifestyle-related diseases].
6. In order to be in good shape, I always use stairs at railway station and university [→ railway stations and universities], instead of riding in an elevator.
7. It is my hobby to grow tomato [→ tomatoes] in my private garden.
8. I belong to Japanese archery club [→ a Japanese archery club/the university's Japanese club].
9. You can see beautiful scene [→ a beautiful scene] when taking a walk.

(*Note*. The underlines, indicating the locations of errors, and the correct answers in parentheses were not shown in the handout distributed to the participants.)

Metalinguistic Explanation: A countable word or phrase must be either preceded by an indefinite article (*a/an*) or have the plural suffix (*-s*) on its end. The definite article (*the*) may be used, in place of an indefinite article, when the referent has been mentioned previously or when the readers or listeners can identify it in context.

(小川　喜正　　英語コミュニケーション学科)