Journal of Response to Writing

Volume 11 | Issue 1 Article 3

April 2025

EFL Essay Writing: Grammatical Accuracy and Productivity

Yoshimasa Ogawa
Showa Women's University

EFL Essay Writing: Grammatical Accuracy and Productivity

Yoshimasa Ogawa
Showa Women's University

Abstract

This study explored a way to help Japanese university students write longer essays while maintaining grammatical accuracy. Participants were three groups of students enrolled in a one-year EFL course in different academic years (N = 111), and the number of words they wrote in 30 minutes and the number of errors made per 100 words were compared. To improve the participants' grammatical accuracy, comprehensive coded feedback (e.g., Bonilla, et al., 2018, 2021; Hartshorn, et al., 2010) and selective metalinguistic explanation (e.g., Bitchener & Knoch, 2010; Sheen, 2007) were provided on the 12 paragraphs/essays they submitted. The first, sixth, and last essays were analyzed to assess their *verb tense* and *mechanical* errors. Regarding the length of writing, the first group kept writing about 150 words, the second group was encouraged to increase the length of writing at their own discretion, and the third group was systematically guided to write longer essays by following a prescribed guideline. The ANOVA results showed that the two groups that wrote longer essays significantly outperformed the short-essay group in the length of writing without sacrificing grammatical accuracy. The correlation analyses produced evidence against a possible trade-off between accuracy and fluency (Lambert & Kormos, 2014; Skehan, 2009).

Key Words: written corrective feedback, coded feedback, metalinguistic explanation, grammatical accuracy, length of writing (productivity)

Introduction

English as a second or foreign language (ESL/EFL) writers tend to make many basic grammatical errors in their compositions, and providing form-focused feedback is one of L2 writing teachers' major responsibilities. Historically, the role of written corrective feedback (WCF) in language acquisition has been challenged for the reasons that: (a) syntactic/lexical exemplars acquired through reading would help improve writing skills more than explicit grammar teaching (Krashen, 1984); (b) ESL writing teachers cannot provide appropriate corrective feedback consistently (Zamel, 1985); or (c) WCF imposes a heavy burden on the teacher and deters students from producing complex sentences (Truscott,1996). However, there is a growing consensus among L2 researchers about WCF's positive effects on writing accuracy (e.g., Bitchener & Knoch, 2008, 2010; Bonilla et al., 2018, 2021; Brown et al., 2023; Ferris, 1999). Feedback also helps learners acquire autonomous editing skills (e.g., Bitchener, 2008; Chandler, 2003; Lalande, 1982). Currently, the proponents of WCF provision are concerned more about how to combine different types of feedback to benefit specific learner groups (Bitchener & Storch, 2016; Bonilla, et al., 2018, 2021; Ferris, 2011). Another challenging issue is how to improve L2 learners' grammatical accuracy and fluency at the same time (Skehan, 1998, 2009; Skehan & Foster, 2001).

The present study is an action research study (Burns, 2005; Wallace, 1998). The primary purpose is to improve the quality of my own introductory writing course at a Japanese university on a year-on-year basis, although it followed a standard procedure for quantitative data analysis. In the past several years, I have been utilizing a combination of coded feedback and metalinguistic explanation to improve Japanese students' writing accuracy (Ogawa 2021a, 2021b). Coded feedback involves the teacher

indicating the error types by using abbreviations or acronyms (e.g., *Art* for article problems or *NF* for noun form errors). It has been effectively used in many teaching contexts (e.g., Bonilla et al, 2018, 2021; Ferris, 2006; Hartshorn et al., 2010; Lalande, 1982). Metalinguistic explanation means that the teacher explains a target grammatical rule, often referring to specific errors in learners' writings (e.g., Bitchener & Knoch, 2010; Sheen, 2007; Shintani & Ellis, 2013). In my studies, coded WCF has been provided comprehensively on all errors, and metalinguistic explanation focused on a set of frequent error types. Focused, or selective, feedback may be more effective because learners' attention is drawn to a limited number of target forms (Brown et al., 2023; Sheen et al., 2009), whereas unfocused, or comprehensive, WCF enables L2 learners to use all grammatical forms correctly (Bonilla et al., 2018, 2021; Ferris, 2011; Liu & Brown, 2015; Van Beuningen et al., 2012). The learner writing outcome in my past studies showed that the combined WCF contributed to the participants' writing accuracy, but the problem is that more intensive feedback did not necessarily result in increasingly greater accuracy (Ogawa 2021a, 2021b). Thus, it may be more practical to help students write longer essays while trying to maintain an acceptable level of grammatical accuracy.

The present study quantitatively evaluated the effects of combined WCF on the writing accuracy and fluency (or productivity) of three groups of Japanese university students enrolled in the same EFL writing course in three different academic years. The three groups received the same EFL writing training and corrective feedback but were required to produce different lengths of writing. Complexity, accuracy, and fluency (CAF) are often referred to as the three principal dimensions of linguistic performance, and there can be a trade-off between L2 learners' performances in these dimensions

(Skehan, 1998; Skehan & Foster, 2001, 2012). Leaving aside the issue of complexity, which entails assessment difficulties (Bulté & Housen, 2020; Kuiken, 2023; Norris & Ortega, 2009), this study explored a feasible instructional plan to make EFL learners write incrementally longer essays without sacrificing grammatical accuracy, which is a challenging task in L2 writing (Brown et al.'s, 2023).

The immediate goal of this study is to improve the quality of my own EFL writing course. However, the results may be shared with other L2 writing teachers teaching similar introductory writing courses where grammatical accuracy is emphasized. Longitudinal studies are believed to generate more solid evidence than cross-sectional studies, and the findings might be compared with other classroom studies to develop efficient teaching strategies in broader L2 writing contexts.

Literature Review

Multiple WCF

Multiple WCF, the major instructional instrument in the present study, has been used in various studies to improve L2 learners' writing accuracy. Hendrickson (1980) and Ferris (1999) combined direct correction and indirect WCF (e.g., underlining and codes). Indirect feedback encouraged L2 learners to notice and correct their own errors, while direct correction provided target forms to those with limited linguistic proficiencies. Sheen (2007) combined direct correction and metalinguistic explanation to facilitate the acquisition of the English articles. Bonilla et al. (2021) comprehensively provided direct correction and metalinguistic codes to improve EFL learners' use of both grammatical and non-grammatical forms (e.g., capitalization and punctuation). Bitchener (2008) and Bitchener and Knoch (2008, 2010) combined direct correction and

written and oral metalinguistic explanations, and Bitchener et al. (2005) combined explicit written directions for correcting errors with student-teacher individual conferences. All these studies showed that multiple WCF was more effective than the no-feedback treatment, but combining various WCF approaches was not necessarily more effective than using a single type of WCF (Bitchener, 2008; Bitchener & Knoch, 2008, 2010). Although Brown et al.'s meta-analysis (2023) showed that the combination of direct correction and metalinguistic explanation tends to be more effective than individual feedback, more research is needed to clarify which type of multiple WCF is more effective in each specific teaching context.

Likewise, repeated provision of corrective feedback has its own limitations. It is important to require L2 writers to submit a follow-up revision based upon teacher feedback because it makes them attend more closely to feedback (Ferris & Hedgcock, 2014). However, what has yet to be clarified is whether writing accuracy improves in proportion to the number of drafts to submit, particularly in introductory L2 courses as opposed to advanced writing tasks that are longer and require repeated content-based and form-focused feedback. For example, Hartshorn, et al, (2010), Evans et al. (2011), Hartshorn and Evans (2015), and Kurzer (2017) engaged L2 learners in short paragraph writing at every class session and provided coded WCF continually until their drafts of each paper became error-free. The learners' accuracy improved significantly, but there is still no evidence that L2 learners' grammatical accuracy in new writings increases in proportion to the frequency of WCF provision. Furthermore, repeated provision of feedback is extremely time-consuming, and regulating the number of drafts to check might be a practical alternative for greater manageability (Messenger et al., 2020).

WCF's Effects on Different Grammatical Forms

Regardless of the intensity of WCF, corrective feedback is not equally effective for all grammatical forms. One way to understand this issue is Ferris's (1999, 2011) categorization between treatable errors and untreatable errors (also see Bichener et al., 2005; Brown, 2012; Brown et al., 2023). Treatable forms (e.g., verb tense, subject-verb agreement, noun forms) are governed by clearly defined syntactic rules; therefore, learners can correct their own errors in response to indirect feedback, but they cannot do so with untreatable errors (e.g., word choice, idiomatic phrases or sentence structures). It must be noticed, however, that certain grammatical forms comprise both treatable syntactic rules and untreatable idiomatic usages: e.g., the English articles (Master, 1990, 1994). Furthermore, treatability is more a matter of degree than dichotomous distinction, and the rule-based standard should not be the only criterion for categorization (Bonilla et al., 2021; Shintani et al., 2014).

Another major factor that influences WCF's effectiveness is structural complexity. Even if a target form is categorized as treatable, learners may not be able to correct it in response to indirect feedback if the structure is very complex and their linguistic proficiencies are limited (Ferris, 1999; Shintani et al., 2014). Yet another criterion for fathoming the efficacy of WCF is the distinction between global errors and local errors (Bates et al., 1993; Burt, 1975; Burt & Kiparsky, 1972; Hendrickson, 1978). Learners are generally motivated to attend more closely to global errors that hinder communication of a message than local errors which do not affect the semantic meanings. To sum up, the availability of clearly defined syntactic rules, learners' ability to use complex forms, and their motivation to attend to corrective feedback combine to make corrective feedback more, or less, effective on different grammatical forms.

Complexity, Accuracy, and Fluency

L2 learners' overall writing (and speaking) proficiencies are often measured in complexity, accuracy, and fluency, and learners' performance in the three dimensions may not positively correlate. L2 learners' linguistic abilities are limited, and their attentional resources tend to be divided between the three dimensions. Hartshorn and Evans (2015) showed that coded WCF had a positive effect on ESL students' writing accuracy but not on fluency. Skehan (2009) argued that L2 learners tend to improve accuracy at the expense of fluency, or vice versa, in oral language tasks. Gunnarsson (2012), who examined Swedish students' essays in French, reported that "It is quite clear that some learners focus on accuracy at the expense of fluency, and to some extent complexity" (p. 273). Furthermore, learners' performance in each dimension can be affected by task type (Skehan, 2009; Skehan & Foster, 1997) or task condition (Chastain, 1990; Skehan & Foster, 1997, 1999). Thus, writing teachers/evaluators need to be aware of the complicated interrelationships between these three dimensions, although it is acknowledged the mutual influence is not always negative but can be positive (Larsen-Freeman, 2009; Robinson, 1995, 2001).

Another caveat concerning CAF is that the assessment of L2 writers' linguistic performance in each dimension entails varying degrees of difficulty. The evaluation of accuracy can be implemented comparatively straightforwardly by analyzing the types and numbers of grammatical errors: i.e., either by counting the number of errors for the target form in obligatory contexts (e.g., Doughty & Varela, 1998; Frantzen, 1995) or dividing the number of errors in an essay by the total number of words, T-unit, or clauses (e.g., Bonilla et al., 2021; Chastain, 1990; Kurzer, 2017; Polio, 1997). For more accurate measurement, linguistic units that contain errors can be categorized and

weighted depending on the degree to which the errors might impede or compromise comprehensibility (Foster & Wigglesworth, 2016; Kuiken & Vedder, 2012). Another alternative is to determine the maximum length of error-free clauses that L2 writers can produce (Skehan & Foster, 2012). It must be acknowledged, however, that dichotomous decisions between correct and incorrect language use are not necessarily easy (Lambert & Kormos, 2014; Polio, 1997) and that it is difficult to determine the extent to which L2 learners' use of linguistic structures deviates from the native speakers' (Housen & Kuiken, 2009; Housen et al., 2012). The assessment approaches to fluency are also established rather firmly. It is primarily related to the speed of writing, although oral fluency may also be measured by the speaker's ability to control for pauses and reformulation (Housen & Kuiken, 2009).

On the other hand, Kuiken (2023) stated that "complexity is a multilayered, multifaced and multidimensional in nature" (p. 84). Grammatical complexity is often computed based on the proportion of subordinate clauses (Wolfe-Quintero, Inagaki & Kim, 1998). However, there has been no clear-cut agreement about the most appropriate complexity indices (Bulté & Housen, 2020; Kuiken, 2023; Wolfe-Quintero et al., 1998). Furthermore, for advanced learners, morphological complexity and phraseological complexity are more important criteria than the proportion of subordinate clauses (Ortega, 2003; Norris & Ortega, 2009). Another major problem concerning complexity is that placing excessive pressures on beginning or intermediate L2 writers to produce complex sentences can result in less comprehensible sentences. Advanced writers/speakers often use succinct expressions to communicate a message effectively; therefore, it is not reasonable to force novice writers to use complex sentences (Lamber & Kormos, 2014; Norris & Ortega, 2009; Pallotti, 2020).

To recapitulate, among the three dimensions of L2 writers' performance, accuracy and fluency (or productivity) are relevant to the present study. Urging learners to produce increasingly longer essays while maintaining an acceptable level of grammatical accuracy is of crucial importance. In terms of accurate assessment, some grammatical forms are amenable to WCF while others are not: learners' errors can be accurately analyzed only in the former category. Thus, this study strove to evaluate the students' ability to write longer essays in an EFL course without losing their grammatical accuracy, and the following research questions were proposed.

RQ1: To what extent does WCF improve the writing accuracy of each of the three EFL groups that are required to produce different lengths of writing?

RQ2: Can the two groups that are encouraged or systematically guided to write longer essays produce longer writings without losing grammatical accuracy?

RQ3: Is there a negative correlation between the learners' grammatical accuracy and productivity in writing?

Method

Participants

The student participants included 139 first-year English majors enrolled in an EFL course at a private Japanese university. The course was a requirement for all English majors. The first-year students in the department were divided into several classes of about 20 to 25, and I taught two classes every year. The 2018 group included 42 students (28 men and 14 women), the 2019 group included 44 students (22 men and 22 women), and the 2021 group included 53 students (31 men and 22 women).

However, the numbers of participants in the three groups were reduced to 34, 41, and 36 respectively in data analysis. They all authorized the use of their paragraphs/essays for the present study by signing a written informed consent form.

Unfortunately, the data for the students' English proficiencies were not available on a unified scale. However, 76 out of the 111 students reported their standardized English proficiency test scores at my request: e.g., Test of English as a Foreign Language (TOEFL), The Test of English for International Communication (TOEIC). Three students' English proficiencies are equivalent to C1, 35 students' to B2, and 38 students' to B1 on the CEFR scale.[1] All the participants had received intensive, grammar-focused English education back in Japanese high schools, which enabled them to respond to the provided error-code feedback or metalinguistic explanation.

Instructional Treatment

The pertinent writing course ran for an entire academic year (or two semesters). The class met for a 90-minute session every week; the total number of class sessions was 30. In 2018 and 2019, class sessions were held in a computer laboratory, where students could type and submit their drafts during the class. Because of the coronavirus pandemic, in 2021 (and 2020), all the class sessions were run on Zoom, a video-conferencing system. Students finished all writing assignments during the Zoom sessions under the teacher's supervision and immediately uploaded their Word files onto an online learning platform called Moodle. They received feedback through the same class management system by the next class session. Every effort was made to make the

[1] The Common European Framework of Reference (https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale) C2 is the second most advanced level; B1-B2 levels indicate intermediate proficiency.

task/test conditions for in-class and Zoom groups as equal as possible, regulating writing time, procedure for submitting assignments, and tools to assist writing. Furthermore, the group enrolled in 2020, i.e., the transition period from face-to-face instruction to online instruction, was excluded from statistical analysis. Instruction was basically conducted in English, but important explanations for complicated grammatical rules were repeated in Japanese, which was the teacher's and students' first language.

The pertinent EFL course was designed to teach both oral and written English skills, but a special emphasis was placed on learning to write paragraphs or short essays. During the first half of each session, students read an English article or watched an English film clip and discussed the depicted issues in English. Then, the remaining class time was dedicated to writing training. The assigned textbook was *Get Your Message Across: Writing Communicative Paragraphs* (Jimbo et al., 2008). The rhetorical structures covered were: time order, space order, process/direction, cause/effect, exemplification, definition, classification, and comparison/contrast. After studying example paragraphs and finishing controlled writing exercises, the students wrote paragraphs or short essays. During the odd-numbered class sessions, they wrote their first drafts of an English paragraph/essay within the timeframe of 30 minutes. During the even-numbered sessions, they wrote the second drafts using part of the class time and submitted them within the same day.

Written Feedback

Error-code WCF was provided comprehensively on all grammatical errors in the first and second drafts of every paragraph/essay. In order to be consistent, I checked each student's draft at least twice and double-checked my own corrective feedback before proceeding to the next student's draft. Table 1 displays all error types for which

error-code feedback was provided; the codes for error types are shown in parentheses. Short comments on ideational issues or discourse structures were provided on either the first or second drafts whenever necessary. Holistic grades on a five-point scale were given on both drafts in terms of syntactic accuracy, vocabulary, content, discourse construction, and sociolinguistic style.

Table 1

Error Types for Which Coded Feedback Was Provided

Error category	Error type
Treatable errors	word order (WO), subject-verb agreement (SV Agr), pronoun agreement (Pro Agr), verb tense (VT), noun form (NF), word form (WF), run-on sentence (Run-on), fragment (Frag), voice (Vo), mechanics (Mec), spelling (Sp), article (Art)
Untreatable errors	wrong word (WW), word missing (WM), unnecessary word (UnW), sentence structure (S/Str), idiomatic expression (Id)
Stylistic errors	informal usage (Inf: e.g., contraction, sentence-final interjection, colloquial expression), sentence-initial conjunction (In-Conj), redundancy (Red), ambiguity (Amb), awkwardness (Awk)

Note. Adapted from Ogawa 2021a. Target forms for metalinguistic explanation are underlined.

To enhance the effects of coded WCF, metalinguistic explanation was provided as a group in the form of a mini-lecture, which took about 10–15 minutes. The target forms (underlined in Table 1) were selected from the frequent error types in the previous student groups' writing samples (i.e., before the year 2018). The class studied one target form per week. All groups received metalinguistic explanation on all the selected forms in the first semester and reviewed them in the second semester. First, I distributed a handout presenting a set of sample sentences that contained the target error type for the

week and instructed the students to work in small groups to identify the errors. Then, I displayed one example sentence at a time on a large screen and called on students randomly to correct the ungrammatical parts. In the end, I orally explained the target grammatical rule, displaying its written summary on the large screen. It must be noticed that *verb tense* received metalinguistic explanation, but *mechanics* did not (Table 1).

Target Length of Writing

All groups were instructed to start with the targeted length of about 150 words. The group in 2018 continued to write about 150 words throughout the year, concentrating on grammatical accuracy, although they were free to write longer paragraphs of their own accord. This study is an action research study, in which no group had originally been intended to be a control group. However, because all groups received the same writing training and corrective feedback—except for the length of writing required—I refer to this first group as Control Group. The 2019 group, hereafter referred to as Encouraged Group, was encouraged, even during the first semester, to increase the length of writing to the best of their ability. They were informed that they might earn a better holistic grade for writing a longer essay if they maintained grammatical accuracy. They were free to switch from a single paragraph to multiple paragraphs or an essay when their writings became long. For the 2021 group, hereafter as Guided Group, the targeted number of words was specified for each assignment. Like Control Group, they were instructed to concentrate on grammatical accuracy in the first semester and, in the second semester, produced incrementally longer compositions, switching from a paragraph to a five-paragraph essay at the prescribed stages (Table 2). All groups were informed of their respective year-long writing plans at the beginning of the academic year in April.

Table 2

Target Length of Writing for Guided Group (2021)

Semester	Assigned Task	Targeted # of Words	Format
First (spring)	Writings 1-6	150 words	Single paragraph
Second (fall)	Writing 7	150-200 words	Single paragraph
Second	Writing 8	200-250 words	Multiple paragraphs
Second	Writing 9	200-300 words	Multiple paragraph
Second	Writing 10	250-300 words	Multiple paragraphs
Second	Writing 11	300-350	Essay
Second	Writing 12	350-400 words	Essay

Analysis

A quasi-experimental pretest-posttest design was used to quantitatively compare the three groups' improvement in writing accuracy and productivity. In this paper, the term *productivity* is used, instead of *fluency*, because only the students' ability to produce longer paragraphs/essays under time restriction was measured. As shown in Table 3, their first drafts for three major writing assignments served as pretest, mid-test, and posttest.

Table 3

Drafts Used as Writing Tests

Test	Draft	Timing
Pretest	Draft 1 of Writing Task 2	At the beginning of the first semester
Mid-test	Draft 1 of Writing Task 6	At the end of the first semester
Posttest	Draft 1 of Writing Task 12	At the end of the second semester

The pretest prompted the students to write what they would do if they were perfectly balanced bilinguals, and the posttest asked them to write what they would do if they were the head of a country who would never be voted out of office. Both issues were believed to be familiar, and understandable, to Japanese students, who could produce stories by using their common knowledge or personal experience. Writing Task 6—one of the weekly assignments based on a target rhetorical structure (cause and effect)—was used as the mid-term writing test.

Writing accuracy was operationalized as the students' error scores at the three tests. The target forms for analysis were *verb tense* and *mechanical convention* (or *mechanics*). The latter included punctuation, capitalization, italicization, quotation, and other mechanical rules; spelling errors were not included because most word processers provide autocorrections. *Verb tense* is a treatable error (Ogawa, 2021a, 2021b; Brown et al, 2023; Ferris, 2006); so is *mechanics* (Ogawa, 2021b; Bonilla et al., 2021). The total numbers of errors that the 111 students made in the three tests are displayed in Table 4. *Noun form*, *articles*, and *subject-verb agreement* were among the other frequent error types and were covered in mini-lectures, but no student groups in the past had learned to use those forms more accurately after receiving both coded WCF and metalinguistic explanation (Ogawa, 2021a, 2021b). For statistical analyses, the numbers of errors per 100 words were calculated, and those normed frequencies were used as error rates (Biber, Conrad, & Reppen, 1998; also see Kurzer, 2018). In this study, the students were allowed to use any expressions and structures in their essay writing, and it seemed appropriate to measure the average frequency of each error type, instead of counting errors in obligatory contexts. In order to avoid procedural complexity, I chose not to categorize and weight error types depending on the seriousness of error. Writing

productivity was operationalized as the average number of words each group produced at each writing test; Microsoft Word's word-count function was utilized for this purpose.

Table 4
The Total Numbers of Errors Made at the Three Writing Tests (N = 111)

Forms receiving multiple WCF	# of Errors	Forms receiving only coded WCF	# of Errors
Noun form	491	Mechanics	495
Article	184	Sentence Structure	350
Verb tense	173	Wrong Word	230
Subject-verb	153	Word Missing	169
Word form	110	Ambiguous	103
Sentence-initial conjunction	71	Awkward	82
Informal	55	Singular/plural	80
Voice	40	Unnecessary Word	36
Run-on	39	Spelling	28
Pronoun Agreement	35	Word order	22
Fragment	24	Redundant	22

I counted the number of learner errors for each target form before returning the drafts to students. Before the first check, I prepared three sets of photocopies of each student's drafts and checked the unmarked drafts for the second and the third time after the semester was over. Error types, instead of tokens, were counted. A Pearson's product-moment correlation coefficient test indicated that the intra-rater reliability between the first and third checks was $r = .98$, $p = .001$. Additionally, another Japanese EFL instructor with an MA in applied linguistics and EFL teaching experiences checked and counted the verb-tense and mechanical errors in 15% of the analyzed essays as a co-rater, and the inter-rater reliability was $r = .85$, $p = .001$

Three two-way AVOVAs were performed to statistically analyze the three groups'

improvement in *verb-tense* accuracy, *mechanics* accuracy, and productivity over the three tests. The within-subjects factor was *test*, and the between-subjects factor was *group*; the dependent variables were the three groups' error or production means at the three tests. Partial eta squared effect sizes were calculated for ANOVAs (small > .01; medium > .06; large > .14), and Cohen's *d* effect sizes were calculated for post hoc *t* tests (small > .2; medium > .5; large > .8). 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 in all post hoc tests. Then, Pearson product-moment correlation coefficient tests were conducted to assess the correlations between the students' accuracy gains and productivity gains. The criteria were: weak > .2; medium > .4; large > .7.

Results

The descriptive statistics for accuracy and productivity means at the three tests are shown in Table 5. The greater numbers of words indicate students' improvement in *productivity*. The smaller *verb-tense* or *mechanics* error rates indicate greater grammatical accuracy. Overall, all three groups' *verb-tense* accuracy improved from pretest to the mid-test and, then, regressed to some extent; the changes are graphically displayed in Figure 1. All groups' *mechanics* accuracy and *productivity* kept improving steadily from pretest to the mid-test and to posttest.

Table 5

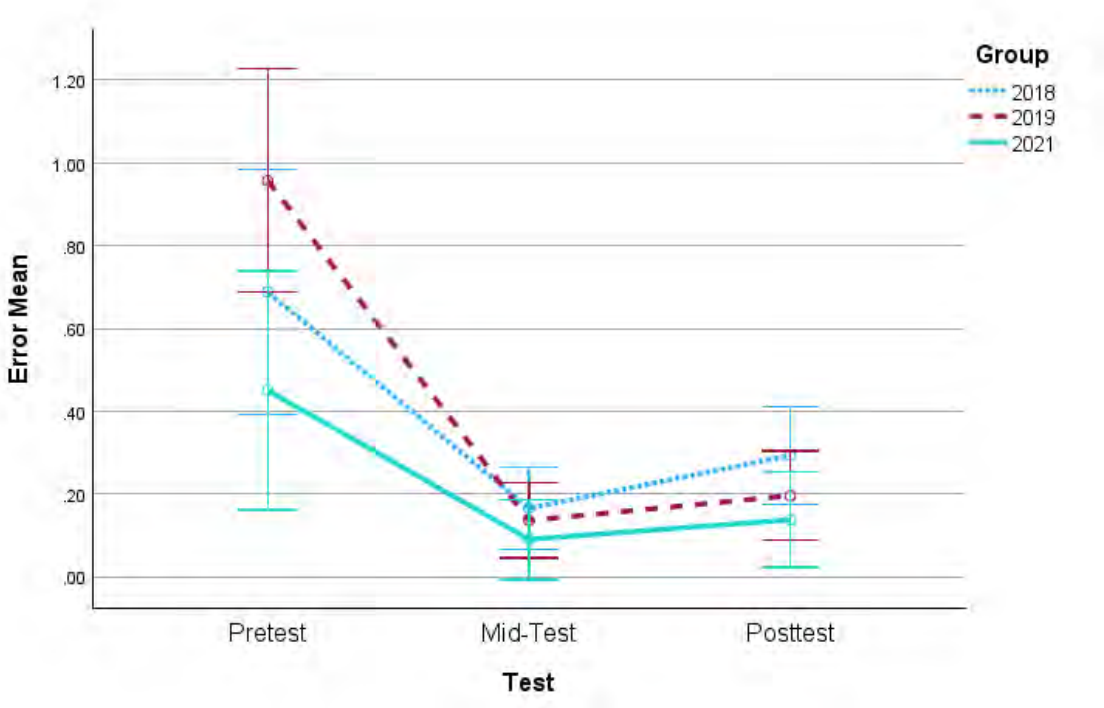
Verb Tense/Mechanics Error Means & Number of Words Written in 30 Minutes

Dimension	Group	Pretest (*SD*)	Mid-Test (*SD*)	Posttest (*SD*)
Verb Tense	Control	0.69 (1.00)	0.17 (0.34)	0.29 (0.47)
	Encouraged	0.96 (0.97)	0.14 (0.30)	0.20 (0.32)
	Guided	0.45 (0.59)	0.09 (0.23)	0.14 (0.22)
Mechanics	Control	1.42 (1.04)	1.13 (1.11)	0.77 (0.74)
	Encouraged	1.44 (1.33)	0.97 (1.70)	0.87 (0.74)
	Guided	1.08 (0.80)	0.61 (0.64)	0.35 (0.30)
# of Words	Control	126.21 (26.06)	130.59 (27.40)	152.47 (27.69)
	Encouraged	125.59 (33.15)	186.32 (42.80)	209.00 (27.86)
	Guided	151.86 (30.85)	155.03 (42.06)	324.61 (76.84)

Note. Control Group, $n = 34$; Encouraged, $n = 41$, Guided, $n = 36$.

Figure 1

Changes in the Verb-Tense Accuracy Means



Prior to the two-way ANOVAs, the three groups' pretest means for *verb tense*,

mechanics, and *productivity* were compared. The results indicated significant differences between: Guided and Encouraged Groups' *verb-tense* rates, $p = .008$; Guided and Control Groups' *productivity* rates, $p = .001$; and Guided and Encouraged Groups' *productivity* rates, $p = .001$. Consequently, the *gains* from test to test were compared in all the post-hoc between-groups comparisons. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for all the three ANOVAs, so the Greenhouse-Geisser statistics were used for correction.

The two-way ANOVA for *verb tense* showed that the *test* main effect was significant, and the effect size was large, $F(1.37, 147.37) = 34.52$, $p = .001$, $\eta_p^2 = 0.24$. The *group* main effect was also significant, and the effect size was medium, $F(2, 108) = 3.50$, $p = .03$, $\eta_p^2 = .06$. To follow up the *test* main effect, the means of the three groups' error means ($N = 111$) were compared over the three tests. In Table 6, larger negative mean-differences in the Pre-to-Mid or Pre-to-Post period represent greater improvement in accuracy. In the Mid-to-Post period, during which the error rates increased on the rebound, the small positive number reflects minor regression. Overall, the whole group's accuracy increased from pretest to the mid-test significantly and then decreased slightly at posttest. To follow up the significant *group* main effect, a one-way repeated measures ANOVA was performed to evaluate the *gain* differences between groups; the result was significant, $F(1.23, 135.31) = 39.18$, $p = .001$, $\eta_p^2 = .26$. The pairwise tests, comparing the three groups' gains during each period (Table 7), indicated that Encouraged Group outperformed Guided Group during the Pre-Mid and Pre-Post periods to a nearly significant degree; this may be partly because Guided Group's error mean at pretest was very low, resulting in a ceiling effect. Overall, all groups followed the same pattern of a major improvement followed by an insignificant setback.

Table 6

Pairwise Comparisons Between Tests: Verb-Tense Errors

Tests	Mean Difference	*t*	*p*	*d*
Mid_Pre	-0.57	-6.74	**0.001	0.88
Post_Mid	0.08	1.85	0.063	0.25
Post_Pre	-0.50	-5.76	**0.001	0.35

Note. $N = 111$. **Statistically significant after Holm's Sequential Bonferroni adjustment.

Table 7

Gain-Differences Between Groups During Each Period: Verb Tense

Period	Groups	Gain Difference	*t*	*p*	*d*
Pre to Mid	Encouraged_Control	-0.30	-1.29	0.20	0.99
	Guided_Control	0.16	0.78	0.44	0.86
	Guided_Encouraged	0.46	2.52	*0.014	0.82
Mid to Post	Encouraged_Control	-0.07	-0.59	0.56	0.50
	Guided_Control	-0.08	-0.80	0.43	0.42
	Guided_Encouraged	-0.01	-0.14	0.89	0.40
Pre to Post	Encouraged_Control	-0.37	-1.52	0.13	1.02
	Guided_Control	0.08	0.38	0.71	0.90
	Guided_Encouraged	0.45	2.53	*0.016	0.80

Note. *Approached statistical significance. The *negative* mean-differences indicate that the group on the left side performed better.

The two-way ANOVA for *mechanic*s showed that the *test* main effect was significant, and the effect size was large, $F(1.83, 197.98) = 16.81$, $p = 0.001$, $\eta_p^2 = 0.14$; the *group* main effect was significant, $F(2, 108) = 5.64$, $p = 0.005$, $\eta_p^2 = 0.10$, and the effect size was medium. The follow-up pairwise tests for the *test* main effect (Table 8) indicated that the whole group's accuracy ($N = 111$) improved from pretest through posttest steadily, and the changes were significant. On the other hand, the one-way ANOVA following up the *group* main effect showed that the gains were not

significantly different between groups, $F(2, 108) = 0.20$, $p = 0.82$, $\eta_p^2 = 0.01$.

Table 8

Pairwise Comparisons Between Tests: Mechanics Errors

Tests	Mean Difference	*t*	*p*	*d*
Mid_Pre	-0.41	3.24	**0.002	1.34
Post_Mid	-0.23	2.28	**0.025	1.07
Post_Pre	-0.64	6.16	**0.001	1.10

Note. $N = 111$. **Statistically significant.

The two-way ANOVA for *productivity* indicated that the *test* main effect was significant, $F(1.78, 191.72) = 216.94$, $p = 0.001$, $\eta_p^2 = 0.67$; so was the *group* main effect, $F(2, 108) = 60.36$, $p = 0.001$, $\eta_p^2 = 0.53$. The interaction effect was also significant, and the effect size was large, $F(3.55, 191.72) = 67.59$, $p = 0.001$, $\eta_p^2 = 0.56$. Following up the interaction effect, first, each individual group's gains in the number of words were compared over the three tests. As shown in Table 9, Encouraged Group's productivity increased significantly from pretest to the mid-test and, then, to posttest. Although Control Group and Guided Group did not improve significantly between pretest and the mid-test, their productivity increased significantly from the mid-test to posttest and from pretest test posttest. Next, the three groups' gains during each period were compared. A post-hoc one-way ANOVA indicated a significant *gain* effect, $F(2, 108) = 26.50$, $p = 0.001$, $\eta_p^2 = 0.33$, and the pairwise comparison results (Table 10) showed that both Encouraged and Guided Groups outperformed Control Group between pretest and posttest and between the mid-test and posttest; Encouraged Group outperformed Control Group and Guided Group between pretest and the mid test. Another noteworthy finding is that Guided Group outperformed Encouraged Group

significantly between pretest and posttest.

Table 9

Pairwise Comparisons Between Tests Within Each Group: Productivity

Group	Tests	Mean Difference	*t*	*p*	*d*
Control	Mid_Pre	4.38	0.67	0.51	0.16
	Post_Mid	21.88	4.19	**0.001	0.72
	Post_Pre	26.27	4.67	**0.001	0.80
Encouraged	Mid_Pre	60.73	9.43	**0.001	1.47
	Post_Mid	22.68	3.22	**0.003	0.50
	Post_Pre	83.42	13.38	**0.001	2.09
Guided	Mid_Pre	3.17	0.48	0.64	0.08
	Post_Mid	169.58	13.23	**0.001	2.20
	Post_Pre	172.75	13.53	**0.001	2.26

Table 10

Pairwise Comparisons Between Groups During Each Period: Productivity

Period	Groups	Gain Difference	*t*	*p*	*d*
Pre-to-Mid	Encouraged_Control	56.35	6.11	**0.001	1.42
	Guided_Control	-1.22	-0.13	0.90	0.03
	Guided_Encouraged	-57.57	-6.21	**0.001	1.42
Mid-to-Post	Encouraged_Control	0.80	0.09	0.95	0.02
	Guided_Control	147.70	10.67	**0.001	2.50
	Guided_Encouraged	146.90	10.04	**0.001	2.37
Pre-to-Post	Encouraged_Control	57.15	6.68	**0.001	1.55
	Guided_Control	146.49	10.5	**0.001	2.46
	Guided_Encouraged	89.34	6.29	**0.001	1.49

Note. **Statistically significant. The *positive* mean-differences indicate that the group on the left side performed better.

Finally, Pearson product-moment correlation coefficient tests were conducted to

evaluate the correlations between *productivity* and *verb-tense* or *mechanics* accuracy. Table 11 shows that there was no significant negative—or positive—correlation between productivity gains and either *verb-tense* or *mechanics* accuracy gains during any period. The correlation coefficients were also computed within each group, but there was no significant correlation between *productivity* and either form in any period for any group.

Table 11
Correlation Test Results

Variable 1	Variable 2	*r*	*p*
Verb Pre-Mid	Productivity Pre-Mid	-0.05	0.58
Verb Mid-Post	Productivity Mid-Post	0.04	0.67
Verb Pre-Post	Productivity Pre-Post	-0.08	0.38
Mechanics Pre-Mid	Productivity Pre-Mid	0.16	0.09
Mechanics Mid-Post	Productivity Mid-Post	0.09	0.35
Mechanics Pre-Post	Productivity Pre-Post	-0.01	0.94

Discussion

The first research question concerned the extent to which WCF improved the writing accuracy of the three participant groups that produced different lengths of writing. Corrective feedback seemed to contribute to every group's accurate use of *verb tense* and *mechanics*. First, all groups improved on *verb tense* from pretest to the mid-test and, then, regressed slightly. This was slightly different from my previous study (Ogawa, 2021a) in which the students continued to improve until posttest although at a slower pace. However, the slight setback was not surprising because students' attention to feedback tends to become somewhat tenuous over time. More importantly,

Encouraged Group's and Guided Group's accuracy improved significantly from pretest to posttest.

The three groups achieved even greater improvement in mechanics accuracy. Whereas verb-tense errors received multiple WCF, mechanical errors received only coded feedback, providing additional evidence that corrective feedback was more effective for mechanics. Furthermore, although learners are commonly believed to pay closer attention to global errors than to local errors, in the present study, the students' mechanical errors (i.e., local error) decreased more sharply—and steadily—than their verb-tense errors (i.e., global error). A possible explanation is that the correct use of verb-tense is more difficult in that learners must attend to the contexts to decide which verb tense is appropriate and then conjugate the verb. Therefore, the structural complexity of target forms might have mediated the effects of WCF on the learners' accuracy (Ferris, 1999; Shintani et al. 2014) more strongly than the local/global hierarchy.

The second research question addressed the issue of whether the two groups that wrote longer essays improved in productivity without losing grammatical accuracy. The gains in productivity were significant for every group between the mid-test and posttest and between pretest and posttest. (It must be remembered that Control Group and Guided Group were instructed to keep writing about 150 words between pretest and the mid-test, concentrating on accuracy.) Then, Encouraged Group and Guided Group outperformed Control Group between pretest and posttest in productivity without suffering any setback in accuracy. In other words, some form of pressure for writing longer essays was beneficial even in an EFL writing course whose main purpose was grammatical accuracy. On the other hand, Encouraged Group's major progress in the

first semester was rather surprising and might be attributable to this individual group's extra effort for fluency (Gunnarsson, 2012). Overall, however, the most important point is that Guided Group, which was systematically guided to increase the number of words they wrote and organize their writings in paragraph or essay format at optimal timings, learned to write longer essays than Encouraged Group in the end.

The third research question probed a possible trade-off between the learners' accuracy and productivity in writing. The correlation test results indicated no significant negative correlations between productivity and the accuracy of either form during any period. Skehan (1998, 2009) and Skehan and Foster (2001) expressed their concern about a possible trade-off between the three dimensions of linguistic performance, and some researchers argued that learners' performance in the three dimensions can be affected by task type (Skehan, 2009) and task condition (Chastain, 1990). In this study, which engaged Japanese students in timed writing tasks, no negative correlations were found between accuracy and productivity. That is, urging introductory EFL students to write longer essays within a limited length of time—a condition that makes the task harder—did not deter their improvement in writing accuracy. More frequent opportunities to practice using the target forms might have enhanced the effects of corrective feedback—aided by the students' advanced explicit grammar knowledge—or, as Robinson (1995, 2001) theorized, a more demanding task might have resulted in keener attention to input and more finely modified output. Although the acquisition process has yet to be further investigated, it is safe to say that making EFL students produce longer essays over time does not hinder their writing accuracy when corrective feedback is constantly provided.

Conclusion

The study results demonstrated that all student groups, regardless of the required length of writing, learned to use the verb-tense form and mechanical conventions more accurately after receiving writing training and corrective feedback. The two groups that were either encouraged or systematically guided to write longer essays maintained their accuracy in time-restricted writing tasks. There was no significant negative correlation between production and accuracy of either target form, indicating the absence of a trade-off.

The educational implications are as follows. First, even in the introductory EFL writing course that emphasizes grammatical accuracy, urging the students to increase the length of their writing in one way or another is beneficial. Second, Encouraged Group that started writing longer essays in the first semester did so without sacrificing grammatical accuracy. That is, it is not necessary to wait until the second semester to start making students write longer essays, dedicating the first semester to the training for improving accuracy. Finally, an instructional plan to guide learners to write longer essays by prescribing the expected length and format of writing for each assignment is even more effective in facilitating productivity than simply telling them to try to produce longer compositions to the best of their ability.

One major methodological limitation was that, although I endeavored to make the task and testing conditions equal between the face-to-face and Zoom sessions, the learning environments might have differed slightly. For example, students in the classroom might have motivated each other to produce longer essays by sharing their products firsthand. Furthermore, in the classroom, I made students use the same desk-top computers to minimize the use of technological writing/editing tools and proctored

their writing. However, in Zoom sessions, I could only proclaim a rule not to use them and warn them that any violation of this policy, which might be indirectly evidenced in their writing products, would result in lower grades. Fortunately, the three groups followed the same improvement patterns in both accuracy and productivity. However, translation tools and AI-based writing software have now become more widely available than at the time of my data collection, and it will be a challenge in future writing assessments to control for students' use of such technological devices.

My future studies should also cover more diverse target learners, research dimensions, and grammatical forms. Since each student group has a different degree of motivation to write longer or more accurate essays, it is worthwhile to conduct replication studies with multiple groups in similar teaching contexts. In-depth interviews are needed to evaluate learners' psychological reaction to instructional treatments. Moreover, although this study focused on two forms clearly amenable to corrective feedback, it is worthwhile to closely analyze some of the frequent error types for which WCF's effects have not been recognized. For example, a detailed analysis of noun form errors, which are interrelated with certain functions of the articles, might be pedagogically beneficial. Finally, this study dealt with two of the three dimensions of linguistic performance—accuracy and fluency—but the issue of complexity must also be addressed when designing syllabi for more advanced writing courses.

Overall, however, the study results supported the proposed instructional treatment: i.e., providing multiple WCF for writing accuracy and gradually guiding students to write longer essays. The findings may provide useful information to teachers who teach similar form-focused L2 writing courses. The study has also shown that several factors, including the structural complexity of a target form, the degree of error

treatability, and the degree to which errors interfere with communication of a message, might interact to influence the effects of WCF on L2 learners' grammatical accuracy, shedding light on an interesting issue to explore in future research. I hope that these findings add to the general knowledge of L2 writing instruction and research.

References

Bates, L., Lane, J., & Lange, E. (1993). Writing clearly: Responding to ESL compositions. Boston, MA: Heinle & Heinle Publishers.

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. (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

Bonilla, López. M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The differential effects of comprehensive feedback forms in the L2 writing class.

Language Learning, *68*(3), 813-850. https://doi.org/10.1111/lang.12295

Bonilla, López. M., Van Steendam, E., Speelman, D., & Buyse, K. (2021). Comprehensive corrective feedback in foreign language writing: The response of individual error categories. *Journal of Writing Research*, *13*, 31-70. https://doi.org/10.17239/jowr-2021.13.01.02

Brown, D. (2012). The written corrective feedback debate: Next steps for classroom teachers and practitioners. *TESOL Quarterly, 46*(3), 861-867. https://doi.org/10.1002/tesq.63

Brown, D., Liu, Q., & Norouzian, R. (2023). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research*, 1-33. https://doi.org/10.1177/13621688221147374

Bulté, B., & Housen, A. (2020). A critical appraisal of the CDST approach to investigating linguistic complexity in L2 writing development. In G. G. Fogal and M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 207-240). Amsterdam: John Benjamins Publishing Company. [Chapter 9]

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

Burt, M. (1975). Error analysis in the adult EFL classroom. *TESOL Quarterly*, *9*(1), 53-63. https://doi.org/10.2307/3586012

Burt, M. & Kiparsky, C. (1972). The Cooficon: A repair manual for English. Rowley, Massachusetts: Newbury House Publishers.

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

Chastain, K. (1990). Characteristics of graded and ungraded compositions. *The Modern*

Language Journal, *74*(1), 10-14.

Doughty, C, & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). Cambridge: Cambridge University Press.

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

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

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: Contents and issues* (pp. 81-104). New York: Cambridge University Press.

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. (2014). *Teaching L2 composition: Purpose, process, and practice*. New York, NY: Routledge.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ration. *Annual Review of Applied Linguistics*, *36*, 98-116. https://doi.org/10.1017/S0267190515000082

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *The Modern Language Journal*, *79*(3), 329-344. https://doi.org/10.1111/j.1540-4781.1995.tb01108.x

Gunnarsson, C. (2012). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 247-276). John Benjamins.

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. https://doi.org/10.5054/tq.2010.213781

Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback: A 30-week study. *Journal of Response to Writing*, *1*(2), 6-34. https://scholarsarchive.byu.edu/journalrw/vol1/iss2/2

Hendrickson, J. M. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal*, *62*(8), 387-398.

Hendrickson, J. M. (1980). The treatment of error in written work. *The Modern Language Journal*, *64*, 216-221. https://doi.org/10.2307/325306

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A. House, F. Kuiken, & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy and*

fluency in SLS (pp. 1-20). Philadelphia: John Benjamins.

Jimbo, H., Elwood, K., Morita, A., Watanabe, Y., Yamada, S., & Yoffe, L. (2008). *Get your message across: Writing communicative paragraphs*. Tokyo: Nan'un-do.

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

Kuiken, F. (2023). Linguistic complexity in second language acquisition. *Linguistics Vanguard 9*(s1), 83-93. https://doi.org/10.1515/lingvan-2021-0112

Kuiken, & Vedder (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-169). John Benjamins.

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

Lambert, C, & Kormos, J. (2014) Complexity, accuracy, and fluency in tasked-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, *35*(5), 607-614. https://doi.org/10.1093/applin/amu047

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579-589. https://doi.org/10.1093/applin/amp043

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

Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly, 24*(3), 461-478.

Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). New York: Cambridge University Press.

Messenger, R., Evans, N., & Hartshorn, K. J. (2020). Managing Dynamic Written Corrective Feedback: Perceptions of experienced teachers. *Journal of Response to Writing, 6*(1). https://journalrw.org/index.php/jrw/article/view/139

Norris, J. M., & Ortega, L. (2009). Toward an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, *30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Ogawa (2021a). Written corrective feedback in EFL: Combining error codes and metalinguistic explanation. *Journal of Response to Writing, 7*(1), 94-125. https://scholarsarchive.byu.edu/journalrw/vol7/iss1/4

Ogawa (2021b). Balancing corrective feedback and writing practice in an EFL writing course. *Gakuen*, *966*, 12-36.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Pallotti, G. (2020). Measuring complexity, accuracy, and fluency (CAF). In P. Winke & T, Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 201-2109). New York, NY: Routledge.

https://doi.org/10.4324/9781351034784

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*(1), 101-143. https://doi.org/10.1111/0023-8333.31997003

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99-140. https://doi.org/10.1111/j.1467-1770.1995.tb00964.x

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring Interactions in a componential framework. *Applied Linguistics*, *22*(1), 27-57. https://doi.org/10.1093/applin/22.1.27

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). Different 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

Skehan, P. (1998). *A cognitive approach to language learning*. New York: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532. https://doi.org/10.1093/applin/amp047

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93-120. https://doi.org/10.1111/1467-9922.00071

Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge: Cambridge University Press.

Skehan, P., & Foster, p. (2012). Complexity, accuracy, fluency and lexis in task-based performance. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199-220). John Benjamins.

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

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.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. -Y. (1998). *Second language development in writing: measures of fluency, accuracy, and complexity*. Honolulu: University of Hawai'i Press.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, *19*, 79-102. https://doi.org/10.2307/3586773