Effects of Written Feedback and Revision on Learners' Accuracy in Using Two English Grammatical Structures

Natsuko Shintani

Nanyang Technological University

Rod Ellis

University of Auckland

&

Wataru Suzuki

Miyagi University of Education

Abstract

The study compared the effects of two types of form-focused written feedback – direct corrective feedback (DCF) and metalinguistic explanation (ME) given to the whole class – on Japanese university students' accuracy of use of two grammatical structures: indefinite article and the hypothetical conditional. Both types of feedback were given with and without an opportunity to rewrite. Accuracy of use was measured in new pieces of writing. The feedback led to increased accuracy for the hypothetical conditional but not for the indefinite article. The effectiveness of the DCF proved longer lasting than the ME. Also, providing opportunity for revision enhanced the effect of the feedback. Overall, DCF followed by revision proved the most effective type of feedback. The results suggest that when form-focused written feedback is directed at two features that vary in saliency and complexity, learners are likely to focus on the structure that contributes more to the "global" meaning of the text. The results also indicate that directly correcting the errors learners make with respect to a complex syntactical structure is more beneficial than giving them a metalinguistic explanation.

Keywords: form-focused instruction, written feedback, L2 acquisition, implicit/explicit L2 knowledge, grammar

Introduction

While the efficacy of form-focused written feedback remains a matter of debate, there is growing empirical evidence that it can result in increased grammatical accuracy in new pieces of writing. The purpose of this article is to take the research on form-focused feedback on writing a step further in three principal ways. The first is to explore the effects of two different types of written feedback—direct corrective feedback and metalinguistic explanation.[1] To date, researchers have investigated the effects of various types of written corrective feedback (WCF). However, WCF is time-consuming as it necessitates teachers providing corrections on individual learners' texts. In contrast, metalinguistic explanation (ME) is more economical of teachers' time as it involves providing explicit instruction on

errors that have been observed in learners' writing to the whole class. Second, relatively few studies have investigated whether asking learners to rewrite following feedback enhances the effect of the feedback. We investigate this issue by comparing groups that were required to revise with groups that only received feedback. Finally, this article investigates whether the effectiveness of form-focused written feedback (i.e., WCF and ME) varies according to the grammatical target. Much of the recent research has focused on the indefinite and definite English articles or verb forms but it is important to know whether written feedback is effective for complex subordinate constructions. To date, no study of written feedback has investigated this. These three issues—all crucial to understanding the role of form-focused feedback—will be the focus of the present study.

Review of the Literature

Is Written Feedback Effective?

Much of the recent research on WCF has been directed at addressing Truscott's (1996, 1998, 2004, 2007, 2010) claim there was neither any empirical or theoretical justification for correcting students' written errors (see also Krashen, 1982; Schwartz, 1993). In response, theoretical arguments in favor of WCF were provided in Ferris' (1999, 2006) and Bruton's (2009, 2010) rebuttal of Truscott's claim. Empirical evidence demonstrating that WCF is effective in improving grammatical accuracy in new pieces of writing has also been forthcoming (e.g., see Bitchener, 2008; Bitchener & Knoch, 2010; Ellis, Sheen, Murakami, & Takashima, 2008; Van Beuningen, De Jong, & Kuiken, 2012).

However, there are a number of issues outstanding. Much of the research to date (e.g., Bitchener, 2008; Sheen, 2007; Frear, 2012) has focused intensively on a narrow set of grammatical features (e.g., English articles or past tense -*ed*). Thus, there is no evidence that WCF leads to greater accuracy in the use of syntactic features. Other researchers (e.g., Van Beuningen et al., 2012) have investigated unfocused written CF (i.e., CF directed a number of

different grammatical features) and have shown that it results in greater overall accuracy in new writing but have not shown a direct connection between correcting errors in specific grammatical structures and improved accuracy in those structures. Finally, few studies have investigated the effect of asking students to rewrite following correction on accuracy in new pieces of writing (but see Chandler, 2003; Frear, 2012).

Types of Feedback

We use the term "feedback" to refer to an attempt to draw learners' attention to problems in their writing. Our concern here is only with form-focused feedback (i.e., we will not consider feedback directed at organizational or content problems). We distinguish two types of form-focused written feedback: WCF and ME. The former consists of corrections made on individual students' texts and constitutes the focus of most of the research to date. The latter consists of explicit descriptions of grammatical features shown to be problematic in learners' writing. In this case, no specific corrections are made; the ME is provided in the form of post-writing instruction, for example, as in the handout given to the students in the current study.[2]

WCF comes in a variety of forms (Bitchener & Ferris, 2012). In our study, we have elected to investigate direct WCF on the grounds that this has been shown to be more effective especially for learners of limited second language (L2) proficiency (Van Beuningen et al., 2012). Indirect feedback affords opportunities for "guided-learning and problem solving" (Lalande, 1982, p.143) but direct feedback is considered preferable if learners are unable to correct their own errors. While some studies have reported no difference in the effect of direct and indirect WCF when rewriting (e.g., Ferris, 2006) or in new pieces of writing (Semke,1984; Robb, Ross, & Shortreed, 1986; Vyatkina, 2010) other studies have found that direct WCF is more effective than indirect WCF (Bitchener & Knoch, 2010;

Hashemnezhad & Mohammadnejad, 2012; Frear, 2012). These studies varied enormously in four regards at least: how these two types of feedback were operationalized, whether the feedback was focused or unfocused, in the nature of the writing tasks investigated, and also in the kinds of learners they investigated. Several of the studies also suffered from various design problems (as discussed by Ferris, 2004; and Truscott, 2007), in particular the absence of a control group. Thus, we place greater store on Van Beuningen et al.'s (2012) carefully designed study, which reported a larger effect for direct WCF on grammatical accuracy.

We also chose to investigate relatively focused (i.e., the feedback was directed at just two grammatical structures) rather than unfocused (i.e., comprehensive) feedback. Van Beuningen et al. (2012) argued that unfocused WCF corresponds to actual teaching practice and thus has high ecological validity. However, focused feedback may have a better chance of enabling learners to restructure their interlanguage systems as learners receive repeated evidence of how to correct the same error. Also, as Shintani and Ellis (2013) noted, a range of different types of errors can be addressed over time if teachers switch the focus of the correction in different pieces of writing. There are only a few studies that have investigated focused versus unfocused WCF. Ellis et al. (2008) reported no difference in the effect of these two types of WCF on new pieces of writing either in the short or the long term. However, they noted that their unfocused condition was, in fact, quite focused as it provided multiple correction of the target error. Sheen, Wright, and Moldawa (2009) set out to make a clearer distinction between the two types. They concluded that unfocused WCF was of limited value whereas focused WCF led to clear gains in grammatical accuracy in new pieces of writing. Farrokhi and Sittapour (2012) also found focused WCF more effective. There is also clear evidence from other studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2010; Frear, 2012) that focused WCF leads to statistically significant gains in grammatical accuracy in new writing—at least in the case of English articles and regular past tense.

An alternative to WCF is metalinguistic explanation (ME). This type of feedback has been little investigated. In Shintani and Ellis (2013), ME took the form of a written handout explaining the use of English definite and indefinite articles. They reported that the learners who received direct WCF in their study failed to develop awareness of the rule whereas those receiving the explicit grammatical explanation demonstrated a much better understanding, were able to use it when rewriting their original text, and demonstrated greater accuracy in a new piece of writing. In the study reported below the ME took the same form as in Shintani and Ellis.

Direct WCF and ME are likely to involve different types of processing by learners. In the case of WCF, learners are provided with input (positive evidence) and can carry out a cognitive comparison of the incorrect and correct forms. However, they may or may not attempt to develop an understanding of the underlying rule that the corrections illustrate. In the case of ME, learners need to identify the specific errors they have made themselves and also work out how to correct them. In this respect, it requires a similar kind of guided problem solving to indirect WCF but differs from it in that specific errors in the learners' writing have not been identified. Arguably, therefore it requires a deeper level of processing.

The Effect of Feedback Followed by Rewriting on Accuracy in New Writing

A number of studies (e.g., Fathman & Whalley, 1990; Ferris & Roberts, 2001) have shown that WCF enables learners to correct their errors when they rewrite their original text. Truscott (1996) rightly pointed out that this does not constitute evidence of learning as learners merely have to copy the corrections in the case of direct WCF. However, it is possible that the opportunity to rewrite enhances the effect of the WCF and thus increases the chance of learning taking place as evidenced in new writing. In the case of ME, rewriting requires learners to apply the information provided to identify and correct their errors.

A number of studies have investigated whether revision leads to increased accuracy in new pieces of writing. Chandler (2003) investigated the effect of direct WCF plus revision. One group revised immediately after correction while another group revised weeks after receiving the feedback. Chandler reported the learners who revised after each piece of writing improved in accuracy from the first to the fifth assignment but there was no improvement in the other group. Hartshorn et al. (2010) examined the effects of what they called "dynamic corrective feedback" (p. 87). This involved a feedback cycle where learners were required to repeatedly revise the same text following indirect (coded) WCF until their writing was error free. They were also asked to record the types of errors in an error log. This resulted in significant improvements in accuracy compared to learners who wrote multidraft papers without any WCF. Frear (2012) compared the effect of writing plus revision (with no WCF) and writing with both WCF and the opportunity to revise on gains in accuracy of two grammatical structures (regular and irregular past tense) by Taiwanese college students. Simply asking students to revise without the help of WCF had no effect on the accuracy of either structure. Those learners who received direct WCF and then revised improved their accuracy in one of the structures (the regular past tense) but not on the other (irregular past tense). Van Beuningen et al. (2012), in the same study referred to above, carried out a similar comparison to Frear but also included a group that just had writing practice (i.e., no revision and no WCF). They reported that direct WCF plus revision resulted in greater short-term and long-term grammatical accuracy than both writing practice alone and writing practice that included the opportunity to revise.

For the purpose of our study, the crucial finding is that WCF plus revision is more effective than WCF alone (as in Frear's 2012 study). Irrespective of whether there were multiple opportunities to revise following WCF (as in Chandler, 2003, and Hartshorn et al., 2010) or only a single opportunity (as in Frear's and Van Beuningen et al.'s studies), revision

following WCF has been shown to lead to greater accuracy in new writing. The most obvious explanation is that WCF in conjunction with the opportunity to revise results in "pushed output" especially if the corrections are removed before they start to write. Swain (1985) hypothesized that pushed output contributes to the noticing of grammatical forms that might otherwise go unattended. The need to revise involves explicit attention to the initial error and its correction, which may promote storage of the target features in memory.

The Effect of Feedback on Different Grammatical Structures

Many of the studies that have investigated the effects of WCF have employed general measures of accuracy – for example, an error-free T-unit ratio as in Hartshorn et al. (2010) or error-words ratio as in Van Beuningen et al. (2012). A number of studies (e.g., Bitchener & Knoch, 2010; Sheen, 2007) have investigated English articles but did not distinguish the effect of WCF on the indefinite and definite articles independently although there is clear evidence that these structures vary in their learning difficulty (Huebner, 1983; Young, 1996). In fact, there is ample evidence in second language acquisition research to show that grammatical structures vary in their learnability and are acquired at different developmental stages (Pienemann, 1998). Thus, it is important to conduct fine-grained studies that focus on the effect of feedback on different grammatical structures. This allows for specific theoretical hypotheses to be investigated as in studies of oral corrective feedback (e.g., Yang & Lyster, 2010).

The relative difficulty of different grammatical structures potentially influences the extent to which they are treatable through feedback. Ferris (1999) defined treatable errors as errors that occur "in a patterned, rule-governed way" whereas errors that are untreatable are those where "there is no handbook or set of rules students can consult" (p. 6). As an example, errors in English regular past tense can be considered treatable whereas errors in irregular past tense are likely to be untreatable. Frear (2012) provided evidence to support this

hypothesis. He reported that focused direct WCF led to improvement in new pieces of writing in the case of regular past tense but not for irregular past tense. Bitchener, Young, and Cameron (2005) also suggested that WCF is more likely to be effective with "treatable" structures. They investigated the effect of different kinds of feedback on three structures (i.e., past tense, definite article, and prepositions). They reported gains in accuracy for past tense and definite article, both of which can be considered patterned and rule-governed, but not for prepositions, which is a much more idiosyncratic grammatical feature.

The distinction between treatable and untreatable errors is, however, an ad hoc one with no clear theoretical basis. It is problematic in a number of respects. First, rule-governed features can vary enormously in their complexity. For example, the rule for the use of the indefinite article for first mention can be considered relatively straightforward whereas the rule for the hypothetical conditional (one of the target structures in the current study) is more difficult as it involves complex verb forms and is also semantically challenging. Evidence in support of this can be found in Ellis (2006). In tests measuring two types of linguistic knowledge (implicit and explicit), which were administered to learners with mixed levels of proficiency, the scores for indefinite article were notably higher than those for hypothetical conditional.

There is clearly a need for studies that investigate the effects of WCF and ME on different structures that vary not just in terms of whether they are rule-based but also in their complexity. In this way, it may be possible to arrive at a fuller understanding of what constitutes a treatable error.

The Current Study

The study has both a theoretical and practical purpose. The two types of feedback differ in an important theoretical way. As noted above, direct corrective feedback (DCF) provides learners with the correct forms and thus is input-providing although it may also

involve output if learners are required to revise (as was the case in one the groups investigated in the study reported below). ME does not provide learners with corrections on their errors but instead prompts them to locate their errors and provide their own corrections. In this respect it can be considered output-prompting. Input-providing and output-prompting feedback draw on different theories of L2 acquisition (i.e., cognitive interactionist theories that emphasize the importance of input, and noticing and skill-learning theories that view output as assisting the proceduralization of declarative knowledge, respectively). This distinction has been investigated in studies of oral CF that compared the effects of input-providing feedback in the form of recasts with output-prompting feedback in the form of prompts such as clarification requests and elicitation (e.g., Lyster, 2004). Comparing the effects of DCF and ME provides a way of investigating in studies of written CF.

The distinction between DCF and ME is also significant pedagogically. DCF is time-consuming as teachers need to correct each student's errors but it has been shown to be effective. ME is arguably less onerous for teachers as once it has been prepared it can be used repeatedly when the same error is observed to re-occur in learners' writing. It also provides learners with explicit information about the rules they are breaking which they can use to monitor future writing for accuracy.

In line with the preceding review of the written feedback literature five research questions were devised. These were directed at investigating the effects of two types of feedback (DCF and ME), with or without the opportunity to rewrite, on accuracy in new pieces of writing, as examined through a dictogloss task:

1. Does DCF lead to more accurate use of the hypothetical conditional and the indefinite article than writing with no feedback?

2. Does DCF followed by rewriting lead to more accurate use of the hypothetical conditional and the indefinite article than DCF alone?

3. Does ME lead to more accurate use of the hypothetical conditional and the indefinite article than writing with no feedback?

4. Does ME followed by rewriting lead to more accurate use of the hypothetical conditional and the indefinite article than ME alone?

5. Is there any difference in the effect of WCF and ME with and without opportunity for rewriting on learners' accurate use of the two grammatical features?

Method

Participants

A total of 214 learners in seven general English classes in a university in Japan agreed to participate in this study. They were all first and second year university students who were majoring in a variety of subjects: mathematics, science, music, physical education, arts, special needs education, and English.[3] All the participants were Japanese speakers who had received at least six years of formal English instruction at their secondary and high schools as well as in the university they belonged to. Although they had substantial metalinguistic knowledge of English grammar, they had limited ability to produce English in free writing tasks. In this sense they were considered to be pre-intermediate level. After eliminating students due to absenteeism from the pretest and posttest phases of the study, the number of participants reached 171. They were 49 male and 122 female students, aged 18 to 21. They were randomly assigned into five groups: the ME group ($n = 40$), the DCF group ($n = 34$), the ME with revision (ME+R) group ($n = 32$), the DCF with revision (DCF+R) group ($n = 28$), and the comparison group ($n = 37$).

Target Structures

Two structures were the focus of this study: (a) hypothetical conditional and (b) indefinite article *a/an*.

Hypothetical Conditional. The hypothetical conditional structure expresses the hypothetical outcome of an event that did not occur in the past. A sentence usually consists of an *if*-clause and a main clause. Celce-Murcia and Larsen-Freeman (1999) argued that acquiring English conditionals is difficult for L2 learners because the structure is complex both syntactically and semantically. They pointed out that acquisition requires knowledge of the tense-aspect system, modal auxiliaries, and negation. Izumi, Bigelow, Fujiwara, and Fearnow (1999) argued that the difficulty of this structure lies in the fact it that requires learners to encode two functions: hypotheticality and past time reference. They noted that it is the most difficult of the three conditional sentence types because it requires an additional marker of pastness in the form of the past perfect to distinguish it from unreal conditional sentences, as in (1):

(1) *If I had money, I would buy a new computer* (unreal conditional).

If I <u>had had</u> money, I would <u>have bought</u> a new computer (hypothetical conditional). They reported that their participants attended to hypotheticality but typically missed the aspect of past time reference. Their results lend support to O'Grady's (1997) Cumulative Complexity Principle, which states that "X is more accumulatively complex than Y if X follows everything that Y does plus something else" (p. 349).

Indefinite Article. English has three articles: the definite article *the*, the indefinite article *a/an*, and the so-called zero article *ø*. Articles convey different meanings that are captured by two binary features: whether the article and associated noun phrase (NP) refer to a specific entity [±specific referent], and whether the article and associated NP are deemed to be already known (from the previous discourse or from context) to the listener [±hearer knowledge] (Hawkins, 2001).

English articles, particularly the indefinite article, are difficult for Japanese speakers for several reasons. First, articles do not accord with the One to One Principle (Andersen,

1984), which guides learners in early L2 acquisition. For example, as Ogawa (2008) pointed out, [±hearer knowledge] determines whether or not *the* can be used (i.e., it cannot be used in noun phrases with [-hearer knowledge]). However, neither [±specific referent] or [±hearer knowledge] exclude the use of *a/an* when the noun phrase is [+specific referent/-hearer knowledge] (e.g., I saw *a* funny looking dog today) or when it is [-specific referent/+hearer knowledge] (e.g., *A* lion is a ferocious animal), or when it is [-specific referent/-hearer knowledge] (I don't have *a* car). Second, there is no grammatical Japanese equivalent to articles in English.

Studies have shown that the indefinite article is frequently late acquired (e.g., Huebner, 1983; Parrish, 1987). Based on studies investigating the acquisition of English articles, Hawkins (2001) suggested the following developmental sequence: bare NP (no article) → specificity (marked by *the*) → definiteness (marked by *a/ø*). According to Hawkins, *the* emerges earlier than *a/an* or *ø* because it can take all types of noun phrases (except for proper nouns) regardless of countability ([±count]) and number ([±singular]). On the other hand, the correct use of *a/an* and *ø* requires learners to judge nominal countability and number correctly.

A number of recent WCF studies have investigated articles (e.g., Bitchener & Knoch, 2010; Sheen, 2007). However, as noted above, they examined both *a/an* for first mention and *the* for anapahoric mention (e.g., *David bought <u>a</u> balloon. <u>The</u> balloon burst.*). The problem here is that learners tend to overgeneralize the use of the definite article, making it difficult to determine if they have really acquired *the* to realize a specific grammatical function. Restricting the analysis to *a/an* allows for more reliable scoring of the effect of feedback on acquisition.

Design

The five groups (DCF, ME, DCF+R, ME+R and comparison group) participated in

five sessions. In Week 1, all the participants completed the first writing task. In Week 2, the DCF and ME groups received and studied their feedback and were then asked to complete a new piece of writing. The DCF+R and the ME+R groups revised their original writing and then completed the new piece of writing. The comparison group was only asked to complete a new piece of writing. In Week 4 all the groups completed a third piece of writing and in the following week, they completed a background questionnaire (see Figure 1). In order to eliminate any possible task effects, the three writing tasks were counterbalanced; each group was divided into three sub-groups and each sub-group completed a different task at Times 1, 2 and 3.

[INSERT Figure 1: see separate file]

Materials

Writing Task Materials. Three dictogloss tasks were designed for this study. The participants were given a task sheet and requested to make notes in the table provided while they listened to an audio-recorded text twice. They were then given a writing sheet and 20 minutes to reconstruct the text. An example of the tasks is provided in Appendix S1 of the online Supporting Information. Dictogloss tasks were chosen for two reasons. First, it was necessary to create a context for use of both target features. While contexts for the use of the indefinite article arise naturally in writing, those for the use of the hypothetical conditional do not. In a free-writing task of the kind used in a number of other WCF studies, learners are likely to avoid the use of such a complex structure. The reconstruction of the dictogloss passages created contexts for the multiple use of both the indefinite article and the hypothetical conditional. The notes that learners took while listening to the passages prompted the use of the hypothetical conditional. Second, as already noted, the participants had limited English proficiency and thus it was felt that a guided writing task such as dictogloss was needed.

Interview.   The purpose of the interview was to obtain information about how learners similar to those in the main study responded to the ME and the DCF. Ten students not involved in the main study were randomly allocated to either the ME or the DCF treatment (five in each). They all completed the first writing task in class. One week later they received feedback on their writing (i.e., either DCF or ME) and then completed the second writing task. The interview, conducted in Japanese immediately after they had completed the second writing task, probed the use that ME learners made of the metalinguistic information sheet, the use that the DCF students made of the corrections on their first piece of writing, and what both groups of learners were thinking about when completing the second piece of writing. The interview protocol is provided in Appendix S2 in the online Supporting Information.

Background Questionnaire. This requested information about the participants' age, gender, the languages they could speak, the English-speaking countries they had lived in and the length of time they had spent in them.

Treatment

All the experimental groups received one feedback treatment as some studies have shown that a one-shot written CF treatment led to improvement in accuracy in new pieces of writing (e.g., Bitchener et al., 2005; Hartshorn et al., 2010; Sheen, 2007).

The DCF group received direct corrective feedback on the writing they produced at Time 1. At time 2, they were given five minutes to look over the corrections on their writing. Then their written texts were collected and they completed a new dictogloss task. The DCF+R group followed the same procedure as the DCF group but were given 20 minutes to rewrite their first text. They were allowed to look at their corrected original text while rewriting. Then their original texts and the rewritten texts were collected by the researchers before they completed the second writing task.

The ME group did not receive any corrective feedback. In Week 2, they were given a

handout providing an explicit explanation of the two grammatical structures (see Appendix S3 in the online Supporting Information) following the first writing task. They were allocated five minutes to check their errors. The written texts and the handout were collected in by the researchers. Then the students immediately completed the second writing task (Time 2). The ME+R group rewrote the first written text after receiving the ME treatment. They were allowed to look at their original text and the grammar explanation while revising. Then, they completed the new (second) piece of writing.

The comparison group followed the same procedures as the feedback groups but did not receive any corrections on their writing. That is, at Time 1, they wrote the original text, and at Time 2 they were given five minutes to look over it before completing the second text. They did not rewrite their original text.

Coding and Scoring

Hypothetical conditional. When a learner attempted a sentence including an *if*-clause corresponding to one of the hypothetical conditional sentences in the dictogloss passage, one point was scored irrespective of whether the sentence was correct. For example, the sentence *If he marry her, he will live in Italy* was coded as an attempt to use the hypothetical conditional because it corresponded to the sentence *If he had married her, he would have lived in Italy* in the dictogloss passage.

We adapted the scoring system used by Izumi et al. (1999). They awarded one point for each of the seven component features of the hypothetical conditional, namely: (a) the past tense, (b) the perfect aspect, and (c) the past participle (PP) form in the *if*-clause, and (d) a modal, (e) the past tense, (f) the perfect aspect, and (g) the past participle form in the main clause. However, we noted that in scoring the sentence below, it is difficult to decide whether *bought* was used as the past form of verb (where it would score 1 point) or as a past participle form (where it would score 2 points).

(2) *If he marry her, he would <u>bought</u> a big house.*

A further problem arose in scoring sentences such as the following:

(3) *If he married her, he went to Paris.*

According to Izumi et al.'s system, this sentence would gain 2 points even though it does not indicate that the learner could use the hypothetical conditional form. In order to solve the first problem, we revised these researchers' scoring system as shown in Table 1.

Table 1.Criteria for Scoring the Hypothetical Conditional

Clause	Criteria	Features	Components	Point
if-clause	a	the perfect aspect	*have* (aux) + verb	1.0
(maximum 2	b	the past tense	*had*	0.5
points)	c	the past participle form	correct PP	0.5
main clause	d	the modal in the past tense	past modal	1.0
	e	the perfect aspect	*have* (aux) + verb	1.0
(maximum 3	f	the modal form	correct form of *have* (aux)	0.5
points)	g	the past participle form	correct form of PP	0.5
			Total possible	5

To solve the second issue, we decided to score the sentences in the following way:

1) The *if*-clause and the main clause were scored separately.

2) For the *if*-clause, students scored one point if the clause satisfied criteria 1 (i.e., it included the auxiliary *have* + verb). If they satisfied this criteria, they scored .5 for each of the criteria 2 and 3, that is, if the auxiliary *have* was in the correct tense (i.e., *had*) and if the past participle form was correct.

3) In a main clause, the students needed to first satisfy both criteria 4 and 5 (i.e., the clause included a past modal, an auxiliary *have*, and a verb) in order to obtain scores by satisfying criteria 6 or 7.

In other words, students did not get a point unless the *if*-clause contained two components (i.e., auxiliary *have* + verb) or the main clause contained three components (i.e., past modal + auxiliary *have* + verb). A total percentage score for each student was calculated as follows:

$$\frac{\text{Number of points scored}}{\text{Number of points possible (i.e., number of conditional sentences} \times 5)} \times 100$$

Table 2 gives some examples of how the scores were awarded.

Table 2

Examples of the Scoring System

Sentence	a	b	c	d	e	f	g	Total
If he marry her, he'll go to Italy.	0	–	–	0	0	–	–	0
If he married her, he would visited Eiffel Tower.	0	–	–	1.0	0	–	–	1.0
If he would have marry her, he had go to Italy.	1.0	0	0	0	1.0	–	–	2.0
If he has became a vet, he would had become very happy.	1.0	0	0	1.0	1.0	0	0.5	3.5
If he had passed the exam, he would have became a doctor.	1.0	0.5	0.5	1.0	1.0	0.5	0	4.5

Note: Column header spans "Criteria" over columns a–g.

Article. An obligatory occasion was defined as any singular noun denoting [+specific referent] and [-hearer knowledge] that appeared in the listening texts in the tasks. There were

seven or eight such occurrences in the dictogloss passages. Then suppliance of the indefinite article in each context was checked. Finally, incidences of overuse were counted. Accuracy was calculated using Pica's (1994) formula for target-like use:

$$\frac{\text{Number of grammatical morphemes supplied accurately}}{\text{Number of obligatory contexts } + \text{ Number of overused forms}} \times 100$$

To examine the reliability of the scoring, 80 texts (over 15% of the 510 texts) were re-scored by a second researcher. The overall agreement rate between the two scorers (near-native and native speakers of English) was 98.2% for the conditional scores and 92.8% for the article scores.

Data Analysis

A number of the participants experienced difficulty with the first dictogloss task. Thirty-three of them failed to attempt even a single hypothetical conditional sentence. As a result, the students in the DCF groups received no feedback on this structure and the ME groups had nothing in their texts to relate the metalinguistic explanation to. While these students continued with the other dictogloss tasks, it was later decided to exclude them from the analyses reported below. Thus, the final sample was 140 participants (ME = 29, DCF = 27, ME+R = 26, DCF+R = 24, and Comparison = 34).

The scores for the writing tasks were subjected to a series of statistical analyses. After confirming the assumption of normality and homogeneity, repeated measures ANOVAs tested the comparative effects of the treatments for the writing tasks' scores. A *post hoc* Bonferroni adjustment test was used to investigate differences between pairs of groups when there was a significant time × group effect. Effect sizes for the ANOVAs were calculated as eta-squared (η^2) with values of .01, .06, and .14 indicating small, moderate, and large effects respectively (Cohen, 1988). Effect sizes for the pairwise-comparisons were estimated using the correlation coefficient *r* with values of .1, .3 or .5 indicating small, moderate, and large effects, respectively.

Results

The research questions asked the effects of each of the four types of written corrective feedback (research questions 1-4) and the comparative effects of the four types of treatment (research question 5) on learners' accurate use of the hypothetical conditional and the indefinite article in new pieces of writing. The results for the five research questions are reported for each target structure separately

Acquisition of the Hypothetical Conditional Form

Table 3 shows the descriptive statistics for the accuracy scores of the hypothetical conditional in the written tasks. A repeated measures ANOVA showed that there were significant main effects for group (F (4, 135) = 9.428, $p < .05$, $\eta^2 = .156$) and time (F (2, 270) = 113.574, $p < .05$, $\eta^2 = .429$), and a significant time × group interaction (F (8, 270) = 11.331, $p < .05$, $\eta^2 = .196$). Bonferroni pairwise comparisons showed that the group differences were only found at Times 2 and 3. This indicates that there were no significant group differences at Time 1 but there were at Times 2 and 3. At Time 2, there were significant differences with large effect sizes between each of the four experimental groups and the comparison group: the ME ($r = .74$), the DCF ($r = .75$), the ME+R ($r = .78$), and the DCF+R ($r = .76$). However, at Time 3 the differences diminished with only the DCF+R differing significantly from the comparison group with a medium effect size ($r = .39$). None of the other experimental groups' scores differed significantly from the comparison group at Time 3 (ME: $r = .20$, DCF: $r = .23$, and ME+R: $r = .18$).

Table 3

Descriptive Statistics for the Accuracy Scores of the Hypothetical Conditional

		Time 1			Time 2			Time 3		
	N	*Atm.*	*Mean%*	*SD*	*Atm.*	*Mean%*	*SD*	*Atm.*	*Mean%*	*SD*

ME	29	4.87	28.11	27.65	5.37	77.69	26.28	5.22	54.69	38.02
DCF	27	5.22	18.72	19.92	5.55	81.84	28.01	5.07	53.91	36.48
ME+R	26	5.01	21.31	28.74	5.09	80.60	22.60	4.78	54.56	40.64
DCF+R	24	5.11	19.92	27.06	4.85	84.15	28.37	5.91	60.57	36.49
Comp.	34	4.98	26.94	30.04	5.00	19.62	26.65	5.35	31.87	32.70

Note. ME = metalinguistic explanation with no revision, DCF = direct corrective feedback with no revision, ME+R = metalinguistic explanation with revision, DCF+R = direct corrective feedback with revision, Comp. = comparison group, *Atm.* = mean number of attempts to use the hypothetical conditional.

Table 4 shows the within-group effect sizes for all the experimental groups. These increased from Time 1 to Time 2 with large effect sizes but in the long-term (Time 1 to 3) only the DCF group ($r = .51$) and the DCF+R group ($r = .54$) showed large effect sizes (i.e., $r > .50$). The pairwise within-group comparisons showed that all the experimental groups significantly improved from Time 1 to Time 2, but then their scores significantly all decreased from Time 2 to Time 3. However, the scores at Time 3 were still significantly greater than those of Time 1 for all the experimental groups. The comparison group did not show any significant differences between Times 1, 2 and 3.

Table 4

Within Group Effect Sizes for the Accuracy Scores (Hypothetical Conditional)

	N	Time 1–Time 2	Time 2–Time 3	Time 1–Time 3
ME	29	.68*	-.33*	.37*
DCF	27	.79*	-.40*	.51*
ME+R	26	.76*	-.37*	.43*
DCF+R	24	.76*	-.34*	.54*

| Comp. | 34 | -.13 | .20 | .08 |

* $p < .05$

In order to examine the relative effect of ME versus DCF, the two ME groups (i.e., ME and ME + R) and the two DCF groups (i.e., DCF and DCF +R) were combined and compared. A repeated measures ANOVA for the writing tasks produced a significant effect for group ($F (2, 137) = 18.912, p < .05, \eta^2 = .216$) and time ($F (2, 274) = 73.038, p < .05, \eta^2 = .348$) and a significant time × group interaction ($F (4, 274) = 22.537, p < .05, \eta^2 = .248$). Bonferroni pairwise comparisons showed that there were no significant differences between the three groups at Time 1. At Time 2, both the combined ME groups ($r = .76$) and the combined DCF groups ($r = .76$) outperformed the comparison group but there was no significant difference was between the two experimental groups. At Time 3, the combined DCF group outperformed the comparison group ($r = .34$) while there were no significant differences between the combined ME groups and the comparison group ($r = .18$) or between the ME and the DCF groups ($r = .11$). To sum up, overall both the ME and the DCF treatments proved to be effective. However, the effectiveness of the DCF proved longer lasting than that of the ME.

In order to examine whether requiring the learners to rewrite their first piece of writing had any effect on accuracy in subsequent new pieces of writing, the two experimental groups that undertook revision (i.e., ME+R and DCF+R) and the two groups that did not revise (i.e., ME and DCF) were combined and compared. A repeated measures ANOVA resulted in a significant effect for group ($F (2, 137) = 18.940, p < .05, \eta^2 = .217$) and time ($F (2, 274) = 72.605, p < .05, \eta^2 = .346$), and a significant time x group interaction ($F (4, 274) = 22.136, p < .05, \eta^2 = .244$). Bonferroni pairwise comparisons showed that there were no significant differences between the three groups at Time 1 ($r = .11$). At Time 2, both the No-Revision group ($r = .75$) and the Revision group ($r = .77$) outperformed the comparison

group, but the difference between the two reconstituted experimental groups was non-significant ($r = .06$). At Time 3, there was again no significant difference between the two reconstituted experimental groups ($r = -.04$) but only the Revision group outperformed the comparison group ($r = .33$).

Acquisition of Article

Table 5 shows the descriptive statistics for the accuracy scores in the written tasks. A repeated measures ANOVA showed that there was no significant effect for group ($F(4, 135) = .388$, $p > .05$, $\eta^2 = .011$) or time ($F(2, 270) = .457$, $p > .05$, $\eta^2 = .003$), and no significant time x group interaction ($F(8, 270) = 1.000$, $p > .05$, $\eta^2 = .029$).

Table 5

Descriptive Statistics for the Accuracy Scores in the Written Tasks (Article)

		Time 1			Time 2			Time 3		
	N	*Obl.*	*Mean%*	*SD*	*Obl.*	*Mean%*	*SD*	*Obl.*	*Mean%*	*SD*
ME	29	5.54	47.14	32.40	5.65	50.93	38.90	5.79	51.31	31.32
DCF	27	5.22	50.48	34.41	5.43	52.04	39.79	5.39	47.63	28.63
ME+R	26	5.87	46.15	36.71	5.55	61.50	37.41	5.68	58.19	27.04
DCF+R	24	5.35	57.58	37.87	5.38	60.38	34.75	5.52	45.04	35.07
Comp.	34	5.51	54.00	39.27	5.78	48.94	29.95	5.51	59.74	31.11

Note. Obl. = the mean number of obligatory occasions.

Summary of the Results

Tables 6, 7, and 8 summarize the results for both between-group and within-group comparisons.

Table 6

Group Differences (All Groups)

	Hypothetical conditional	Indefinite article
Time 1	ME = DCF = ME+R = DCF+R = Comp.	ME = DCF = ME+R = DCF+R = Comp.
Time 2	ME = DCF = ME+R = DCF+R > Comp.	ME = DCF = ME+R = DCF+R = Comp.
Time 3	DCF+R > ME = DCF = ME+R = Comp.	ME = DCF = ME+R = DCF+R = Comp.

Note. "A=B" indicates no significant differences between A and B (at $p < .05$). "A > B" indicates A obtained significantly greater scores than B.

Table 7

Group Differences (DCF vs. ME; +Revision vs. –Revision) for Hypothetical Conditional

	ME vs. DCF	+/-Revision
Time 1	ME = DCF = Comp.	+R = –R = Comp.
Time 2	ME = DCF > Comp.	+R = –R > Comp.
Time 3	DCF > ME = Comp.	+R > –R = Comp.

Note. "A=B" indicates no significant differences between A and B (at $p < .05$). "A> B" indicates A obtained significantly greater scores than B.

Table 8

Improvement in Accuracy over Time

Group	ME	DCF	ME+R	DCF+R	Comparison
Hypothetical conditional					
New writing (Immediate)	Yes	Yes	Yes	Yes	No
New writing (Delayed)	Yes	Yes	Yes	Yes	No
Article					
New writing (Immediate)	No	No	No	No	No
New writing (Delayed)	No	No	No	No	No

Note. Yes = the group improved significantly in the test compared to the pre-test scores. No = the group did not significantly improve.

Discussion

Polio (2012) noted that learners can use both their implicit knowledge (i.e., knowledge that is intuitive and available for automatic processing) and their explicit knowledge (i.e., conscious knowledge accessed via controlled processing) while writing but argued that written corrective feedback is more likely to increase their explicit knowledge. We consider it very unlikely that the dictogloss tasks used in this study elicited implicit L2 knowledge. As Van Beuningen et al. (2012) pointed out such tasks encourage the display of correct language. Thus, any increase in accuracy probably reflects the learners accessing their explicit knowledge when formulating sentences and monitoring for accuracy. Also, as a result of many years of traditional grammar instruction, the learners in this study were oriented to treating English as a set of explicit rules.

The feedback required attention to two structures. As explicit knowledge requires controlled processing (i.e., processing that is effortful because it requires conscious attention to form), attending to two grammatical structures simultaneously is likely to have been burdensome. We speculate the learners chose to attend consciously to the hypothetical conditional rather than to the indefinite article both when processing the input provided by the dictogloss tasks and when reconstructing the texts. The hypothetical conditional was more salient in the dictogloss passages and more crucial for reproducing the content. Also, it is likely that the corrections of the verb forms of the hypothetical conditional were more noticeable than the corrections of the indefinite article as they were more numerous. Thus, we propose that the learners may have utilized their explicit knowledge of the hypothetical conditional but, given the processing demands of the writing tasks (i.e., the need to recall the content of the passages from memory without any linguistic or visual support), they failed to

attend closely to the indefinite article.

This interpretation receives some support from the interviews conducted in Japanese with the five DCF learners. All five learners reported attending to the hypothetical conditional in the post-test writing tasks but only one of them reported attending to the indefinite article. Rita, for example, commented:

> I didn't think about the correction at first, and around the middle of the story, I noticed the conditional form and went back to the first sentence and corrected the conditional form. I wasn't sure if it should be the past or present, but because my text had been corrected to "had + PP", so I thought I would follow the corrected form. I didn't attend "*a*" or "*the*" at all.

The interviews with the five ME learners indicated that they did attend to both structures but also that they experienced greater difficulty with the indefinite article. Misa, for example, said:

> All the conditional sentences start with "if he had married her" so I decided to keep the same *if*-clause all the way through and only change the "would have something" part. I thought "came" was impossible, so I changed it into "have come". I also tried to attend the indefinite article for occupations but there were not many.

Thus, the interviews suggest that both the DCF and the ME learners approached the reconstruction of the texts strategically. They engaged in formulating and monitoring the use of the hypothetical conditional using their explicit knowledge but did not do so for the indefinite article or else did so less effectively.

The Effects of the Feedback on Accuracy in the Writing Tasks

All the experimental groups improved in accuracy in using the hypothetical conditional from Time 1 to Time 2 while there was no change in the comparison group's accuracy scores (see Tables 3 and 4). However, much of the gain in accuracy by the

experimental groups was lost by the time of the final writing task while the comparison group's scores increased at Time 3. However, the difference in scores between Time 1 and Time 3 was still statistically significant for the experimental groups and the effect sizes for two of the groups (DCF and DCF+R) remained large. In contrast, none of the time differences for the comparison group were statistically significant and the effect sizes were negligible. At Time 2, all the experimental groups outperformed the comparison group but at Time 3 only the DCF+R did so. With the exception of this group, then, there was no long-term advantage for the feedback.

As pointed out above, there is evidence from the interviews that the learners approached the writing tasks strategically, formulating and monitoring for accuracy by applying their explicit knowledge. This interpretation is supported by the fact that accuracy declined over time. Explicit knowledge is subject to decay and thus the learners may have not been able to formulate and monitor the hypothetical conditional so effectively in the final writing task. It should be noted, however, that gains in accuracy were still evident at Time 3. In other words, while the effects of the feedback directed at this complex structure declined over time they were sufficient to ensure that some gains in accuracy were maintained.

The results for hypothetical conditional in our study differ from those for other grammatical structures in some previous studies of WCF (e.g., Bitchener & Knoch, 2010; Van Beuningen et al., 2012), which, like this study, involved a one-shot feedback treatment but which reported long-term effects in comparison with control groups. The difference can be explained by that fact that the target structure of our study – the hypothetical conditional – was more complex than the target structures in these other studies and therefore less treatable (Ferris, 1999).[5] It should be recalled that Bitchener et al. (2005) also found that WCF was not effective for their untreatable target feature (i.e., prepositions). If this explanation is correct it

suggests that what constitutes "treatability" is not just a question of whether or not a feature is rule-based but also the complexity of the rule-based structure.

We turn now to consider why the feedback has no effect on the indefinite article. The explanation we offer reinforces our claim that, for these learners, any effect of feedback was due to the use they made of their explicit knowledge of the two target features. The interviews suggested that the learners paid little attention to the indefinite article in their text reconstructions. This is understandable for as Schmidt (2001) pointed out, attention is limited, selective and partially subject to voluntary control. Given the demands imposed on the learners' processing capacity by the need to recall and encode the propositional content of a text, they focused their attention on the structure that was more salient and more semantically important in both the input provided by the task passages and their output (i.e., the hypothetical conditional).

The results for the indefinite article differed from those in Shintani and Ellis (2013). Shintani and Ellis used a similar design to this one (i.e., the effects of the DCF and ME on accuracy in new pieces of writing were compared). They reported that the ME group improved from Time 1 to Time 2 but declined at Time 3. The DCF group did not improve over time. The ME group outperformed the comparison group at Time 2 but not at Time 3. The difference in the results can be explained by the fact that in Shintani and Ellis only one structure – indefinite article – was investigated. Thus, the learners were able to focus on indefinite articles. The ME proved more effective because it made the rule for the use of the indefinite article for [+specific referent] and [-hearer knowledge] contexts explicit and learners were able to use this information to formulate and monitor their output. In contrast, the learners in the DCF group had to work out the rule for themselves and post-task interview data indicated that they were unable to do so. In the current study, the learners were required to attend to both the hypothetical conditional and the indefinite article and, as suggested

above, they elected to focus on the former as it may have been more important for reproducing the propositional content of the passages.

To sum up, the results for the current study point to two main conclusions. The feedback resulted in gains in explicit knowledge of the form of the hypothetical conditional which the learners used to formulate and monitor their production but which atrophied over time. The feedback did not assist the development of explicit knowledge of the indefinite article and the learners were not able to make effective use of whatever knowledge of this structure they initially possessed in the demanding writing tasks because the nature of the tasks may have led them to pay closer attention to the hypothetical conditional.

The Differential Effects of DCF and ME

We turn now to consider the relative benefits of WCF and ME. As there were no group differences for the indefinite article, the discussion only addresses the results for the hypothetical conditional.[6] The analyses that combined ME (i.e., the ME and the ME+R groups) and DCF (i.e., the DCF and the DCF+R groups) showed that overall both the ME and the DCF treatments proved to be effective at Time 2 (i.e., shortly after the provision of feedback). However, the effectiveness of the DCF proved longer lasting than that of the ME as the difference between the DCF groups and the comparison group was still significant at Time 3 whereas that for the ME group was not.

This lends support to the general finding in the corrective feedback literature, namely that DCF is effective in enabling learners to improve accuracy in new writing tasks. The DCF in this study provided learners with positive evidence showing the correct verb formation for the hypothetical conditional and enabled them to compare this input with their own erroneous attempts to use this structure. In contrast, the ME required the students to apply an abstract account of the hypothetical conditional to identify and correct their errors. This also proved effective in the short-term but its effects were less durable. The advantage of ME is that it is

less demanding of teachers' time as it does not involve correcting each student's script. However, for a complex structure such as the hypothetical conditional, the input that DCF provided appears to have led to a clearer understanding of the verbs forms needed for this structure.

The Effect of Rewriting Following Feedback

Finally, we consider whether requiring the learners to revise their first piece of writing had any effect on accuracy in subsequent new pieces of writing. Both the combined Revision and the combined No-revision groups outperformed the comparison group at Time 2 but at Time 3 only the Revision group did so. This indicates a long-term advantage of requiring learners to rewrite following feedback. Having learners rewrite may have helped them to process the feedback more deeply helping them to consolidate their declarative knowledge of the hypothetical conditional. As a result, its effect diminished less over time. The results, therefore, suggest that the extra time required for revision is of pedagogical value as shown in Chandler's (2003) study. However, it remains a possibility that simply allocating additional time for learners to process the feedback they received (i.e., so they can inspect and reflect on the corrections or ME provided) without any requirement to rewrite would have had a similar effect. The study did not examine this possibility.

Conclusion

The Japanese participants in this study were foreign language learners with extensive explicit knowledge of English grammar but limited ability to use this knowledge in meaning-focused tasks such as dictogloss. Any generalization, then, must be restricted to such learners. Many of the previous studies of focused feedback have focused on a single grammatical structure. A unique feature of this study was that it examined the effect of form-focused feedback on two very different grammatical structures. Both of the structures were difficult for the limited proficiency learners we investigated. The feedback only led to gains in the

hypothetical conditional. This suggests that when a writing task is demanding, they have difficulty in making use of feedback for two structures and will focus their attention on the feedback that addresses the feature that is more salient and semantically functional. Further research is needed to examine the effects of focused feedback directed at two structures. Such research should, like this study, examine structures that differ in complexity and but also investigate structures that are of similar difficulty. In this way, it will be possible to establish whether it is the difference in structural difficulty or just the fact that learners are required to process information about two structures that determines the effectiveness of the feedback.

The study sought to compare the relative effect of two types of form-focused feedback (DCF and ME). ME did lead to short-term improvement in accuracy as measured by a new piece of writing but overall the DCF proved somewhat more effective in the long term possibly because it provided learners with both negative and positive evidence regarding the specific errors they had committed. In future research it would be interesting to investigate the combined effect of combining DCF and ME as they afford very different types of information to the learner, both of which can contribute to learning. Another aim was to investigate whether asking learners to revise had any add-on value to feedback. The results indicated that it did. The most effective of the four types of feedback proved to be direct corrective feedback plus revision. In the delayed writing task only the DCF+R group was more accurate than the comparison group.

This study also had a practical purpose. It sought to contribute to our understanding of the role of written form-focused feedback in language pedagogy. As is evident from an inspection of popular teacher handbooks (e.g., Hedge, 2000; Ur, 1996) there is general agreement that such feedback is needed but also concern about the danger of over-correcting. Focused feedback of the kind investigated in this study addresses this concern. The question arises, however, as to how "focused" the feedback should be. This study suggests that if the

feedback addresses two very different structures it might not be effective for both at least in the kind of writing task used in this study – dictogloss.[7] Other studies have shown that feedback focused on a single grammatical structure is effective. Thus, teachers might be advised to limit the focus to a single structure, switching to different structures over time. However, if form-focused feedback is to have a durable effect, learners may need more than a single-shot of feedback. This study showed that the positive effects of the one-shot feedback on the hypothetical conditional wore off over time.

Another practical concern the study sought to address is the onerous nature of corrective feedback. Correcting individual learners' errors is time-consuming. In contrast, providing metalinguistic explanations in the form of handouts given out to students when specific errors have been found in their writing is much less onerous. Such handouts, once developed, can be used repeatedly when the need arises. In an earlier study (Shintani & Ellis, 2013), ME was found to be more effective than DCF. However, in this study, it was shown to be less effective than the DCF. The main difference was that in Shintani and Ellis the ME focused on a single structure whereas in this study it addressed two structures. This might have resulted in an overload of information. At this point, then, there is insufficient evidence to support the use of ME in place of WCF. Until further evidence is forthcoming, teachers might like to experiment with the two ways of providing feedback.

Final revised version accepted 11 August 2013

Notes

1. A reviewer queried the use of the term form-focused feedback here. We use this term to refer to any attempt to provide learners with feedback on their errors. Traditionally, this has taken the form of corrective feedback (i.e., feedback provided on individual

errors made by individual learners and involving either indication or correction). However, form-focused feedback can also take the form of metalinguistic explanation provided to all the learners following completion of a piece of writing. Thus, we distinguish two types of form-focused feedback (i.e., corrective feedback and metalinguistic explanation).

2. A reviewer rightly pointed out that ME can be individualized and provided in response to specific errors in the same way as DCF. Our purpose in this study, however, was to investigate generic ME that can be provided to all the learners.

3. An anonymous reviewer pointed out that English majors might be very different from other majors especially in their ability to incorporate metalinguistic explanation and that this might be a confounding variable in this study. In this study we conducted a multiple-choice test to measure the participant's decontextualized knowledge of the hypothetical conditional and the indefinite article. The English major groups' scores for the hypothetical conditional and the indefinite article were 10.52 and 6.31 while those for the other majors' groups were 10.79 and 6.40. The results of t-test comparing the major and non-major groups were non-significant ($p > .05$). The test suggests that there was no difference in the major and non-major students' metalinguistic understanding at the beginning of the study.

4. Pseudonyms were used for all the participants referred to in the study.

5. Van Beuningen et al. (2012) only reported the effect of WCF on general accuracy of a range of structures. It is possible that this effect was the result of improved accuracy in treatable structures (e.g., simple morphological features).

6. In Shintani and Ellis (2013) the ME was found to be more effective for indefinite article than the DCF. They proposed that this was because it gave learners a much clearer idea of the rule involved in the use of the indefinite article for first mention.

7. A reviewer pointed out that the difficulty the learners experienced in attended to two structures in our study may have been the result of the particular type of writing involved (i.e., reconstructing texts from memory) and might not apply if learners are generating their own texts. We acknowledge this possibility, which awaits empirical investigation.

References

Andersen, R. W. (1984). The one to one principle of interlanguage construction. *Language Learning, 34*, 77-95.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 69-124.

Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. London: Routledge.

Bitchener, J. & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 193-214.

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*, 191-205.

Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were… *System, 37*, 600-613.

Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System, 38*, 491-498.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The grammar book*. New York: Heinle & Heinle 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*, 267-296.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*, 431-463.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*, 353-371.

Farrokhi, F. & Sattarpour, S. (2012). The effects of direct written corrective feedback on improvement of grammatical accuracy of high-proficient L2 learners. *World Journal of Education, 2*, 49-57.

Fathman, A. & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B.Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge, UK: Cambridge University Press.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1-11.

Ferris, D. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime …?). *Journal of Second Language Writing, 13*, 49-62.

Ferris, D. (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). Cambridge, UK: Cambridge University Press.

Ferris, D. & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161-184.

Frear, D. (2012). *The effect of written corrective feedback and revision on intermediate Chinese learners' acquisition of Englis*h. Unpublished doctoral dissertation. The University of Auckland, New Zealand.

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*, 84-109.

Hashemnezhad, H., & Mohammadnejad, S. (2012). A case for direct and indirect feedback: The other side of coin. *English Language Teaching Journal, 5*, 230-239.

Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford, UK: Blackwell.

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford, UK: Oxford University Press.

Huebner, T. (1983). *A longitudinal analysis of acquisition of English*. Ann Arbor, MI: Korama press.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*, 421-452

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.

Lalande, J. F., II. (1982). Reducing composition errors: An experiment. *The Modern Language Journal, 66*, 140-149.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399–432.

O'Grady, W. (1997). *Syntactic development*. Chicago: University of Chicago.

Ogawa, M. (2008). The acquisition of English articles by advanced EFL Japanese learners: Analysis based on noun types. *Journal of Language and Culture Language and Information, 3*, 132-151.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning, 37*, 361-383.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning, 44*, 493-527.

Pienemann, M. (1998). *Language processing and language development: Processability theory*. Amsterdam: John Benjamins.

Polio, S. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375-389.

Robb, T., Ross, S., & Shortreed, I. A. N. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83-96.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163.

Semke, H. D. (1984). The effects of the red pen. *Foreign Language Annals, 17*, 195-202.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*, 255-283.

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*, 556-569.

Shintani, N. & Ellis, R. (2013). The comparative effect of metalinguistic explanation and

direct written corrective feedback on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing, 23,* 286-306.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Rowley, MA: Newbury House.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327-369.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*, 103-135.

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing, 13*, 337-343.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255-272.

Truscott, J. (2010). Some thoughts on Anthony Bruton's critique of the correction debate. *System, 38*, 329-335.

Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge, UK: Cambridge University Press.

Van Beuningen, C. G., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*, 1-41.

Vyatkina, N. (2010). The effectiveness of written corrective feedback in teaching beginning German. *Foreign Language Annals, 43*, 671-689.

Yang, Y. & Lyster, R. (2010). Effects of form-focused practice and feedback on chinese EFL learners? Acquisition or regular and irregular past tense forms. *Studies in Second Language Acquisition, 32*, 235-263.

Young, R. (1996). Form-function relations in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135-175). Amsterdam: John Benjamins.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: An Example of the Tasks

Appendix S2: Interview Protocol

Appendix S3: Metalinguistic Explanation