Effects of Different Types of Written Feedback on Acquisition of Two Target Structures Measured by Revision and New Writing

YingYing BAO  
Graduate School, Miyagi University of Education

Kounosuke SATO  
Graduate School, Miyagi University of Education

Adrian LEIS  
Miyagi University of Education

Wataru SUZUKI  
Miyagi University of Education

Abstract

A growing amount of recent second language (L2) research has shown both empirical evidence for the effectiveness of written feedback (WF) on L2 writing and acquisition and mediating roles played by various factors in WF efficacy (see Kang & Han, 2015 for a review). Out of a great number of variables, this study focuses on three: (a) type of feedback (direct feedback vs. metalinguistic explanation), (b) type of grammatical structures (i.e., indefinite article vs. past hypothetical conditional) and (c) type of measurement (i.e., revision vs. new writing), and examined relationships among these three variables. Forty-one Japanese university students of English were asked to reconstruct a text (first draft) in Week 1 of the study. They were then given either direct feedback (DF) or metalinguistic explanation (ME) and then asked to revise the first draft (revision) in Week 2. Finally, participants reconstructed a new text (new writing) in Week 4. Results showed that (a) DF promoted more accurate revision than ME, (b) DF and ME worked differently according to different structures, and (c) types of WF had no differential effects on promoting accuracy in new writing. These findings are discussed with reference to L2 research. This article concludes by providing limitations, future directions, and pedagogical implications.

1. Introduction

In response to Truscott’s (1996) claim that written feedback (WF) is unnecessary, ineffective, and potentially harmful, many theoretical arguments in favor of WF have been provided (e.g., Ferris, 1999, 2006) and many empirical studies have been conducted to examine the effectiveness of WF
on L2 writing and acquisition. A growing number of recent L2 studies (e.g., Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007; Shintani, Ellis, & Suzuki, 2014; Van Beuningen, de Jong, & Kuiken, 2012) has shown empirical evidence for the effectiveness of WF on L2 writing and acquisition, and mediating roles of numerous variables in the effectiveness of WF (see Kang & Han, 2015 for a review). These variables, among many others, include type of feedback, type of linguistic structures, and type of measurement.

The type of feedback has been an object for many empirical studies. Earlier studies focused on the effectiveness of direct feedback (DF) and indirect feedback (IF) on L2 writing accuracy development (e.g., Chandler, 2003; Robb, Ross, & Shortreed, 1986). Only recently, Shintani and her colleagues examined the effectiveness of another type of WF, metalinguistic explanation (ME), which is thought to have a pedagogical value in large classrooms (see Shintani & Ellis, 2013, 2015; Shintani et al., 2014). Although there is a growing number of L2 studies investigating the relative effectiveness of different types of WF, the research base so far has been limited in certain grammatical features (e.g., articles in English). The extent to which WF is effective for different domains and categories of linguistic knowledge should be fully explored in future L2 research. Another controversial issue is the type of measurement. The effectiveness of WF may differ from short-term revision to a new piece of writing (e.g., Bitchener & Knoch, 2010a, 2010b; Sheen, 2007; Truscott & Hsu, 2008). In this study, we focused on relationships among types of feedback (DF vs. ME), types of target structures (indefinite article vs. past hypothetical conditional), and types of measurements (short-term revision & new writing), each of which will be discussed in detail below.

### 2. Literature review

#### 2.1 The type of WF

According to Bitchener and Ferris (2012), WF can be provided in a variety of forms such as DF, IF, and ME. DF gives correct target forms or structures near the errors by crossing out an unnecessary word and inserting a missing word or morpheme. IF provides no information to students as to the correct way of writing but simply indicates errors by, for example, underlying, highlighting, or coding the error. ME involves the use of metalinguistic clues as to the nature of the errors that learners have made without providing the actual correction.

Research on the relative effectiveness of providing different types of WF has tended to dominate WF literature over the last 30 years (see Bitchener & Ferris, 2012). Many previous studies have investigated the relative effectiveness of DF and IF (e.g., Chandler, 2003; Robb et al., 1986), different types of IF (e.g., Ferris & Roberts, 2001), and different DF combinations (e.g., Bitchener & Knoch, 2008, 2009, 2010a). Overall, these studies have reported no differences between various feedback types (see Bitchener & Knoch, 2010b for a review). Recently, however, a growing number of studies has reported a clear advantage for DF, especially for learners of limited L2 proficiency (e.g., Bitchener & Knoch, 2010b; Ellis et al., 2008; Ferris, 2012; Sheen, 2007; Shintani et al., 2014;...
Van Beuningen et al., 2012), while in Lalande’s (1982) study (reported in Ferris, 2002) IF was shown to be more effective in improving learners’ writing accuracy than DF. ME has been recently gaining an interest among L2 researchers because of its pedagogical concern to give DF and IF in a large classroom in typical English as a foreign language contexts such as Japan (Bitchener, Young, & Cameron, 2005; Sheen, 2007; Shintani & Ellis, 2013; Shintani et al., 2014; Suzuki, Leis, & Itagaki, 2014).

The present study investigated relative effectiveness of DF and ME on improved accuracy in L2 writing. DF enables learners to instantly internalize the correct form, whereas IF makes learners notice that there is an error in writing, but they do not know if their own hypothesized corrections are accurate or not (Chandler, 2003). ME involves providing linguistic clues, and thus appeals to learners’ explicit knowledge by helping them to understand the nature of the error they have committed (Ellis et al., 2008). The former has been shown to be effective in recent studies, while the latter has been investigated relatively less often. Only a small body of studies addresses the relative effectiveness of these two types of WF (see Shintani & Ellis, 2013, 2015; Shintani et al., 2014).

2.2 The target structures

Investigations into the effectiveness of WF have been limited in a certain domain of linguistic knowledge (e.g., articles in English). Sheen (2011) claims “…we cannot assume that because CF [corrective feedback] has been shown to assist the acquisition of one grammatical feature it will necessarily do so for all features” (p. 165). Therefore, it is important for L2 researchers to target various linguistic features and languages.

Bitchener et al. (2005) investigated effects of WF on the use of articles, past simple tense, and prepositions in English. They found that WF was effective for English articles and past simple tense but not for prepositions. Based on this finding, they argued that WF may be effective for rule-governed structures, such as articles and tenses. Also, Shintani et al. (2014) focused on two different target structures (the indefinite article and past hypothetical conditionals). In their study, accuracy development was found on conditionals, but not on articles. When explaining their findings, Shintani et al. inferred that conditionals are more salient and semantically more important than articles for certain kinds of writing tasks, such as dictogloss and text reconstruction.

In the current study, we targeted two grammatical structures: the indefinite article, which is less salient, semantically not important, and simple; and the past hypothetical conditional, which is more salient, semantically important, and complex.

2.3 The type of measurement

Previous L2 studies vary depending on whether they use revision or new writing to examine the effectiveness of WF. Revision immediately following WF is considered to be a measure of its short-term effectiveness. However, Truscott (1996) argued that improvement in revisions alone does
not constitute evidence that learning has occurred. As Sheen (2007) also argues, “to claim that WF results in learning, one must examine whether the improvement in revisions carries over to a new piece of writing or if the improvement is manifested on posttest or delayed posttest measures” (p. 258).

Previous L2 studies focused only on revision as a measure of short-term effectiveness of WF (e.g., Ferris & Roberts, 2001; Robb et al., 1986). In contrast, an increasing number of recent studies have been investigating effectiveness of WF using a new writing (e.g., Bitchener & Knoch, 2008, 2009, 2010a, 2010b) or some kinds of delayed-posttests (e.g., Sheen, 2007). For example, Sheen found no significant differences between DF and ME in the immediate posttest, however, found an advantage for ME over DF in the delayed posttest conducted two months later. The present study was designed to examine effectiveness of different types of WF using both short-term revision and new writing.

2.4 Research questions

Relative effectiveness of different types of WF is a matter of intensive debate in current L2 research. Although WF effectiveness may differ depending on linguistic structures, few studies have been conducted to explore this issue. Furthermore, effectiveness of WF may vary depending on whether it is measured using revision or a new piece of writing. In order to shed some insights into these complex issues, the current study attempts to answer the following three research questions:

1. Is there any difference in effectiveness of DF and ME on grammatical accuracy improvement in L2 writing?
2. Does effectiveness of WF vary according to different target structures?
3. Is there any difference in the effect of DF and ME in improving accurate use of two grammatical structures in a new piece of writing?

3. Method

3.1 Participants

A total of 41 Japanese learners of English agreed to participate in this study. They were all first- and second-year students at a national university located in northeast Japan. Their major was education but their specialties were all different (i.e., math education, science education, arts education, special needs education, and English language education) and they were considered to be low intermediate level of proficiency. At the time of this experiment, the participants were taking English class as a required subject. There were 11 male and 27 female students, aged 18 to 21. All participants were randomly divided into two treatment groups with a total of 19 members for each group. Three students were excluded from the analyses because they failed to produce even a single instance of the two target structures in writing and thus we could not give feedback.
3.2 Linguistic target

The current study investigated two different grammar structures: the past hypothetical conditional and the indefinite article in English. A past hypothetical conditional sentence is used to express past hypothetical situations and their consequences, which did not really occur (Celce-Murcia & Larsen-Freeman, 1999). A typical past hypothetical sentence usually consists of an "if"-clause (if + had past participle) and a main clause (would + have + had past participle). An example is shown below:

Ex) If I had had some money, I would have bought a new computer.

We also focused on the indefinite article in English. The indefinite article (a / an) is used to indicate an unspecific or first mentioned referent (Celce-Murcia & Larsen-Freeman, 1999). An example is given below:

Ex) My friend had a little daughter. She wants to be a pianist.

There were three reasons why we chose these two grammatical features. First, both target structures are considered to be difficult for Japanese learners of English. This is because the grammatical usages of conditionals in English were much more complex than that in Japanese. There is also no grammatical Japanese equivalent to articles in English (Celce-Murcia & Larsen-Freeman, 1999). Because of their novelty and abstractness, articles are difficult for learners whose L1s do not have them (DeKeyser, 2005). Second, the past hypothetical conditional and indefinite article are considered as complex and simple grammatical structures, respectively. The past hypothetical conditional is a rule-governed but syntactically and semantically complex grammar feature. In contrast, the indefinite article is relatively simple in terms of its description. Third, both structures have been well-researched in previous SLA studies (e.g., Bitchener & Knoch, 2010a, 2010b; Shintani et al., 2014; Uggen, 2012).

3.3 Design and procedures

Table 1 shows the experimental procedure of this study. In the first week, all participants completed a dictogloss task (i.e., first draft). In the task, first, the participants were provided with a sheet of paper and asked to take notes while listening to an audio-recorded text. The text was played twice at normal speed. Then, another sheet of paper was distributed to all the participants, and the participants were requested to reconstruct the whole text as accurately as they could, based on their notes. The reconstruction phase was completed within 20 minutes. Both sheets were collected at the end.

In the second week, the two groups received their respective feedback and revised the original writing. The DF group had their papers returned with the ill-formed target structures directly
corrected. All the errors of the indefinite article and the past hypothetical conditionals were underlined or indicated with symbols inserted and correct forms were provided nearby. Errors on other grammatical structures were left uncorrected. Two examples are given below.

Ex) Original: Norika was good cook.
   Original with feedback: Norika was good cook.

Ex) Original: If I have married her, I would have enjoyed fantastic meals every night.
   Original with feedback: If I had married her, I would have enjoyed fantastic meals every night.

The ME group was given back their first drafts without any corrections. Instead, they were provided with a brief explanation handout, written in Japanese, about the indefinite article and the past hypothetical conditionals. Both groups were asked to study their feedback for five minutes and to revise the first draft for 20 minutes. They were allowed to keep the original writing with DF and the ME sheet but not allowed to use other resources (e.g., dictionary).

In the fourth week, both groups completed another dictogloss task (delayed posttest). The procedure was exactly the same as the first dictogloss task. In order to counterbalance task effects, each group was further divided into two sub-groups and each subgroup received a different task across the first draft and the new piece of writing. Each task took approximately 30 minutes.

Table 1
The Procedure of the Study

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Writing task 1 (First draft)</td>
<td>Writing task 1 (First draft)</td>
</tr>
<tr>
<td>Week 2</td>
<td>DF (5 min) + Revision task (20 min)</td>
<td>ME (5 min) + Revision task (20 min)</td>
</tr>
<tr>
<td>Week 4</td>
<td>Writing task 2 (New writing - posttest)</td>
<td>Writing task 2 (New writing - posttest)</td>
</tr>
</tbody>
</table>

Notes. DF: Direct Feedback; ME: Metalinguistic Explanation.

3.4 Materials
3.4.1 Writing task

The current study adopted dictogloss as a writing task. Two passages used in this study shared almost the same level of vocabulary and structures, and the frequency of occurrences of the target structures (see Shintani et al., 2014). There were 188 words in the passage titled “The bachelor” and 207 words in the passage titled “A holiday never was” (see Appendix for a sample).

Dictogloss was chosen in the current study for two reasons. First, dictogloss can prompt learners to produce target structures instead of avoiding them. If learners had avoided using the past conditional and indefinite article in the writing task, we would have faced difficulty in providing
feedback on the targeted forms. Therefore, dictogloss can help us provide feedback on the targets compared with other types of tasks such as free writing (e.g., Van Beuningen, de Jong, & Kuiken, 2008) and picture description (e.g., Hanaoka, 2007; Swain & Lapkin, 2002). Second, most learners participating in this experiment belonged to non-English majors and were considered to have limited writing skills. Dictogloss, a kind of guided composition, can help learners produce language in L2. In short, our current experiment materials were designed with the specific purpose of ensuring participants to produce two target structures as many as possible, thereby making possible this investigation into effectiveness of WF in L2 writing.

3.4.2 Metalinguistic explanation (ME)

In this study, the participants in the ME group received a sheet of paper on which the grammar explanation of the two target structures were provided in Japanese (see Shintani et al., 2014). The grammar explanation sheet includes a short introduction about the purpose and the linguistic knowledge about two target structures. Both the indefinite article and the past hypothetical conditionals were provided with definitions followed by grammatical rules and examples explaining when and how to use them.

3.5 Scoring and analyses

Three writing samples (i.e., first draft, revision, new writing) were analyzed in terms of accuracy of the two grammatical structures. To obtain accuracy scores of the first draft and new writing, we first counted the total number of correct uses, incorrect uses, and overuses of the two target structures. Then we divided the number of correct uses by the total number and then multiplied by 100 to the final accuracy rate.

\[
\text{(Number of correct uses ÷ Total number of use of target structure) × 100}
\]

The accuracy of revision was calculated differently from that of the above analysis. That is, we divided the number of successful uses in the revision (i.e., correct uses of the target structures and successfully corrected errors) by obligatory context use in the first draft (incorrect uses plus overuses of the target structures), and multiplied that score by 100.

\[
\frac{\text{Number of successful use of target structure in revision task}}{\text{Total number of use of target structure in the first draft}} \times 100
\]
Table 2

Examples of the Scoring System of Articles

<table>
<thead>
<tr>
<th>First Draft</th>
<th>The number of errors</th>
<th>Revised draft</th>
<th>The number of successful corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norika has rich family.</td>
<td>1</td>
<td>Norika has a rich family.</td>
<td>1</td>
</tr>
<tr>
<td>...he had lived big house</td>
<td>1</td>
<td>...he would have lived big house with her.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 gives an example in which we scored the indefinite article for the accuracy of revision. First, all the incorrect uses and overuses of the indefinite article were counted from this participant’s first draft. In this example, each of the two sentences included an error on indefinite article, thus the obligatory occasion was two. Then the number of successfully corrected errors in revision was counted. One error was corrected successfully in the example, and another was not corrected. Therefore, the accuracy rate was 50% (i.e., 1÷2×100).

Table 3

Examples of the Scoring System of the Past Hypothetical Conditionals

<table>
<thead>
<tr>
<th>First draft</th>
<th>The number of errors</th>
<th>Revised draft</th>
<th>The number of successfully corrected errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>If I would have go to Egypt,</td>
<td>2</td>
<td>If I have gone to Egypt,</td>
<td>0</td>
</tr>
<tr>
<td>I would visited pyramids.</td>
<td></td>
<td>I would visited pyramids.</td>
<td></td>
</tr>
<tr>
<td>If he married her,</td>
<td>2</td>
<td>If he had married her,</td>
<td>1</td>
</tr>
<tr>
<td>he would has a big house.</td>
<td></td>
<td>he would have has a big house.</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

A past hypothetical conditional sentence was scored as two points: one for the If- clause and one for the main clause. Table 3 shows how we scored the past hypothetical conditional. In his/her first draft, this participant had tried to use two target sentences, which were not correct. The total number of targeted errors was 4 (2 for each sentence). In the revision task, one out of four errors was corrected successfully. Therefore, the accuracy rate was calculated as 25% (i.e., 1÷4 ×100).
4. Results

Table 4 shows the accuracy rate of the DF group and the ME group in the learners’ short-term revision task. The learners in the DF group scored 96% in the case of the past hypothetical conditionals and 92% in the indefinite article. In contrast, the ME group scored 70% for the conditional and 52% for the article. The descriptive statistics in revision show (a) that DF was more effective than ME and (b) that the past hypothetical conditional benefited more from WF than the indefinite article did.

Table 4
Accuracy of Short-term Revision Task

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conditionals</td>
<td>Articles</td>
</tr>
<tr>
<td></td>
<td>First draft</td>
<td>Revision</td>
</tr>
<tr>
<td>The number of Errors</td>
<td>170</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>174</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>10</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 5
Accuracy of New Writing Posttest

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conditionals</td>
<td>Articles</td>
</tr>
<tr>
<td></td>
<td>First draft</td>
<td>Delayed test</td>
</tr>
<tr>
<td>Obligatory</td>
<td>170</td>
<td>194</td>
</tr>
<tr>
<td>Success</td>
<td>17</td>
<td>122</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>10</td>
<td>63</td>
</tr>
</tbody>
</table>

Table 5 shows the participants’ writing accuracy of the two target structures in a new piece of writing. The DF group achieved 63% accuracy for the conditionals and 37% accuracy for the indefinite articles. Similarly, the ME group achieved 65% accuracy for the conditionals and 38% for the articles. The descriptive statistics in new writing show that conditionals benefited more from both DF and ME than articles did.
5. Discussion

5.1. The effect of different types of feedback

The first research question was “Is there any difference in effectiveness of DF and ME on grammatical accuracy improvement in L2 writing?” The answer to this question is Yes for the revision task, and No for a new piece of writing. In revision, the DF group outperformed the ME group in terms of accuracy of both target structures. Our finding is consistent with that of previous SLA studies, which showed the advantage of DF on learners’ L2 revision (e.g., Chandler, 2003; Ellis et al., 2008; Ferris & Roberts, 2001). Providing the correct target forms explicitly enables learners to revise their errors. For example, Chandler investigated different effects between DF and IF on revision, and reported positive effects for both types of feedback. She further suggested that “direct feedback is best for producing accurate revisions” (p. 267). In her study, learners were asked to rewrite their first draft based on the feedback they were given. Naturally, it was easy for students who had received DF to understand and make corrections (Chandler, 2003). Also, Ellis et al. (2008) investigated only DF, using a series of three narrative tasks, focusing on English articles, and proved the effectiveness of DF both in an immediate posttest and delayed posttest in new pieces of writing. Taken together, our findings lend some support for effectiveness of DF on accuracy development in L2 writing.

Note that the participants in the ME group achieved a relatively lower accuracy in short-term revision task than those in the DF group. The participants in the ME group received their first drafts without any corrections but a handout with grammar explanations about the two target structures. The handout might have included too much information for the participants to digest in five minutes. Shintani and Ellis (2013) also investigated the difference between DF and ME and found that the ME group gained more accurate scores in the revision task after treatment compared to the DF group. The difference between the current study and Shintani and Ellis (2013) is that the latter only examined accurate use of the indefinite article, although the former investigated both the indefinite article and conditionals. The information provided in this study might have been cognitively demanding and the participants might have thus been overloaded. Effectiveness of WF on new writing will be discussed with RQ3.

5.2. The effect of different types of target structures

The second question was “Does effectiveness of WF differ according to different target structures?” The answer appears to be Yes. Overall, it showed that both DF and ME had effects on improving accuracy of the conditionals, but this was not true for articles. Only the DF group performed well in the short-term revision task on the article but the effects leveled out in the new writing. As mentioned earlier, two structures in this study were chosen because both of them are rule-governed and previous studies show WF is effective. However, the conditionals in the tasks were more meaning bearing and salient than the indefinite article. It should be mentioned that there
are many studies focusing on WF on English articles (e.g., Bitchener, 2008; Bitchener et al., 2005; Bitchener & Knoch, 2008, 2009, 2010a, 2010b; Sheen, 2007; Shintani & Ellis, 2013; Shintani et al., 2014), and they found effectiveness of WF on learners’ accuracy improvement in the use of articles. For example, Bitchener et al. (2005) investigated effects of WF on the use of the English articles, past simple tense and prepositions. They found WF to be effective for English articles and past simple tense and they also claimed that WF may be effective for those rule-governed structures. However, these studies above did not focus on more complex, idiosyncratic features such as the past hypothetical conditionals in this study we used. Shintani et al. (2014) investigated DF and ME by focusing on the past hypothetical conditionals and the English indefinite article. The results showed that both types of feedback had effects on conditionals, but had no effects on articles at all. Shintani et al. claimed that the conceptual saliency of structures may have been a crucial factor in drawing learners’ attention. It is also possible for the participants to have avoided and ignored the article, and paid attention to the past hypothetical conditionals, which is more perceptually salient and semantically important in executing tasks in the current study. Besides, the article itself is understandably problematic from a cross-linguistic perspective (Celce-Murcia & Larsen-Freeman, 1999). There is no equivalent of English articles in participants’ first language.

5.3 The effect of feedback in the new piece of writing

The third research question was “Is there any difference in the effect of DF and ME in improving accurate use of two grammatical structures in a new piece of writing?” The answer appears to be No. The participants in both DF and ME groups significantly improved their grammatical accuracy of the past hypothetical conditional in new writing, although there was no difference between them. This finding is consistent with that of previous L2 studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010a, 2010b; Ferris & Roberts, 2001). For example, Bitchener and Knoch (2010b) investigated relative effectiveness of DF, ME and DF+ME on accuracy in the article over 10-month period and found no difference among them. Based on consistent findings across L2 studies, one might want to argue that WF, regardless of its types, promotes L2 writing accuracy development.

If we take a closer look into accuracy in the short-term revision and new writing between the two groups, an interesting trend emerges. That is, the ME group’s accuracy in the conditionals marginally changed from revision (70%) to new writing (65%). However, this is not the case for the DF group who achieved high accuracy in the short-term revision task (i.e., 96%) but a lowly 63% in the new piece of writing. This finding may be an indicative of delayed effects of ME on L2 writing accuracy development. This speculation is supported by Sheen’s (2007) study, which investigated relative effectiveness of DF and ME. No difference was found between DF and ME in the short-term revision task, whereas a clear advantage was found for ME over DF in the new writing posttest administered two months later. The results indicate that the passing of time may be a crucial factor in determining effectiveness of WF on L2 writing accuracy development. However,
Shintani et al. (2014) found that effectiveness of DF proved longer lasting than that of ME, suggesting that DF is preferable for a complex structure. Further research is needed in order to investigate long-term effects of different types of WF on the acquisition of different target structures.

6. Conclusions

Many studies have examined the effectiveness of different types of WF on improved accuracy in L2 writing, but only limited studies investigated effectiveness of DF and ME. Also, previous L2 research has been limited in that it only examines effectiveness of WF on a certain grammar feature (e.g., the article in English). Furthermore, only a few researchers have examined whether effectiveness of WF differs according to a type of measurement (i.e., short-term revision vs. new piece of writing). To fill these research gaps, in the present study, we aimed at examining the effectiveness of DF and ME on improved accuracy in the indefinite and the past hypothetical conditional measured by revision and new writing. Consistent with recent L2 studies, the findings of the present study indicate that WF, under some circumstances, contributes to grammatical accuracy improvement in L2 writing.

There are, admittedly, several limitations of this study which need to be overcome for future research. First, there was no control group, which is not given feedback of any kind in this study. Although a considerable improvement of learners’ accuracy of the conditional from first to new drafts provides indirect evidence for effectiveness of WF, future studies should include a control group to show that accuracy improvement is solely the result of WF. Second, a two-week interval is too short to observe effectiveness of WF. The effectiveness of DF and ME has proven to show no difference at all in a new piece of writing administered one week later. Like Sheen (2007), longer-term intervals (e.g., two months) between WF treatments and posttest writing may be desirable to show differential effects of different types of WF. Finally, like many other studies, this study provided WF to the participants only once. This single shot nature of WF treatment may not be enough to demonstrate long-term effectiveness of WF on L2 writing and acquisition (c.f., Bitchener, 2008; Bitchener & Konch, 2010a). Future research is needed to examine whether proving multiple opportunities for the learners to receive WF on multiple drafts is beneficial for L2 development.

This article concludes by briefly providing three pedagogical implications of this study. First, to improve grammatical accuracy in students’ writing, teachers may wish to provide DF or ME, or both. Second, it is necessary for L2 instructors to observe the effectiveness of WF not only in the short-term (i.e., revision immediately after feedback) but also in new pieces of writing (e.g., following weeks). Finally, L2 practitioners should not only provide WF but also devise a way to help Japanese learners of English acquire some difficult grammatical features such as articles. This may include (a) combining DF and ME, (b) giving multiple opportunities to provide feedback on multiple drafts (e.g., Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson, 2010), and (c) encouraging learners to deeply process WF through self- or peer-explanation (e.g., Suzuki, 2012).
Notes

1Although it is not reported in this article, there are many other variables that are thought to mediate effectiveness of WF. These include proficiency levels (e.g., Li, 2013), aptitude (e.g., Sheen, 2007), and age (e.g., Coyle & Roca de Larios, 2014).

2This is a part of a large scale study which examined complex relationships among (a) type of WF (direct vs. indirect), (b) type of structures (indefinite article vs. past hypothetical conditional), (c) type of measurement (revision, new writing, grammar tests), and (d) learner's language analytic abilities (Shintani & Ellis, 2015; Shintani et al., 2014; Suzuki et al., 2014). The present article is different from these studies, none of which examines effects of WF on grammatical accuracy measured by revision and new writing.

3It is important to note that short-term effectiveness of WF has been measured in L2 research using revision (the current study) and a new piece of writing (e.g., Sheen, 2007; Shintani et al., 2014).

Acknowledgements

The current research was funded by the Miyagi University of Education President’s special research grant awarded to the fourth author of this article. Our sincere thanks go to the students and teachers for their participation in this study. Part of this paper was presented at the 2015 annual conference of the Japan Society of English Language Education in Kumamoto. We would like to thank Konomi Ariji, Rod Ellis, Moe Hashiba, Nobuya Itagaki, Arif Laskar, and Natsuko Shintani for their assistance throughout this project.

References


1Although it is not reported in this article, there are many other variables that are thought to mediate effectiveness of WF. These include proficiency levels (e.g., Li, 2013), aptitude (e.g., Sheen, 2007), and age (e.g., Coyle & Roca de Larios, 2014).

2This is a part of a large scale study which examined complex relationships among (a) type of WF (direct vs. indirect), (b) type of structures (indefinite article vs. past hypothetical conditional), (c) type of measurement (revision, new writing, grammar tests), and (d) learner's language analytic abilities (Shintani & Ellis, 2015; Shintani et al., 2014; Suzuki et al., 2014). The present article is different from these studies, none of which examines effects of WF on grammatical accuracy measured by revision and new writing.

3It is important to note that short-term effectiveness of WF has been measured in L2 research using revision (the current study) and a new piece of writing (e.g., Sheen, 2007; Shintani et al., 2014).


**Appendix: An example of the tasks**

**Listening text**

**A Bachelor’s Life**

Koji Nonaka is 40 years old and a bachelor. When he was younger he had a lot of girlfriends and they all wanted to marry him but he rejected them. There was Norika. She came from a rich family. If he had married her, he would have lived in a big house. Then there was Sayaka. She was a great cook. If he had married her he would have enjoyed fantastic meals every night. Maria was a very beautiful Italian girl. If he had married her he would have gone to live in Italy. Natsue was a great conversationalist. If he had married her he would have had lots of fun talking to her. But the most special woman was Atsuko. She became a very successful actress. If he had married her he would have met lots of famous people. Now he often thinks about all these women and the life he might have had. But he always concludes that it is really better to stay single. He knew that if he had married any of them, he would have had a lot of problems!