Types of Information in Written Corrective Feedback and Its Efficacy on L2 Acquisition

Sung-Soo Jang
(Inje University)


This study investigates whether the effect of written corrective feedback (CF) on learners’ acquisition of English article usage is mediated by the type of information CF provides. To tap into this issue, four types of information on error were identified: its existence, location, correct form and nature. Written CF was categorized into three types according to the sorts of information it contained: indirect CF that supplies information on existence and location, direct CF that provides information on existence, location and correct form, and metalinguistic CF that includes metalinguistic explanation on error nature and all the other information types. Three CF groups were accordingly formed along with a control group in a quasi-experiment, where the groups engaged in three narrative-writing tasks. Error correction and narrative writing tests were administered to measure the acquisition of explicit and implicit knowledge of English articles respectively. The results revealed that written CF was beneficial to acquisition of both L2 explicit and implicit knowledge and that such positive effects emerged only in the metalinguistic CF group. The findings added counterevidence to Truscott’s (1996) argument against written CF utility and suggested that metalinguistic explanation on error may play a pivotal role in written CF-triggered L2 acquisition.

I. INTRODUCTION

Language acquisition theories have not taken a uniform view of learners’ errors and their corrections. Behaviorists, who paralleled habit formation with learning, claimed that errors are detrimental to learning processes and correcting them is necessary. In contrast, those under the influence of nativist views such as Universal Grammar did not believe that

* This work was supported by the 2011 Inje University research grant.
error correction is essential as acquisition is fundamentally driven by positive evidence. Interactionists emphasized that errors can be corrected through more competent interlocutors’ feedback arising in their interaction with learners. They focused on this particular type of corrective feedback (CF) and its value of facilitating acquisition.

CF can be provided in oral or written form. Since Chaudron’s (1977) seminal work, however, a majority of CF studies in the SLA literature have empirically and theoretically focused on learners’ oral errors. Overall, empirical research has shown that oral CF plays a positive role in L2 acquisition although its efficacy is mediated by various factors such as research settings, target features, and individual learner differences including proficiency and anxiety (Ammar & Spada, 2006; Egi, 2007; Leeman, 2003; Long, Inakagi, & Ortega, 1998; Lyster & Mori, 2006; Mackey & Philp, 1998; Sheen, 2006, 2008). On the theoretical front, such beneficial effectiveness of oral CF was supported by several SLA theories such as noticing hypothesis (Schmidt, 1995), comprehensible output hypothesis (Swain, 1995), interaction hypothesis (Long, 1996) and skill learning theory (DeKeyser, 1998, 2007).

Theories, however, differ in terms of what type of CF is optimal. Interaction hypothesis prioritizes CF providing correct form while comprehensible output hypothesis places more emphasis on CF withholding it. Sociocultural theory denies the existence of a single best type, claiming that CF efficacy depends on how finely it is tuned to the learner’s stage of development. Some researchers even took a counter view of CF benefits, arguing that CF is harmful, rather than helpful, to L2 acquisition as it raises the learner’s affective filter (Krashen, 2003).

Research into the effect of written CF on L2 acquisition has recently increased but is still scarce in comparison with oral CF. It is partially because, until recently, writing skill was not a major target area of SLA research and CF did not constitute an important component of L2 writing instruction. This study is one attempt to alleviate current imbalance between oral and written CF research by focusing on written CF and comparing the efficacy of its different types. To contextualize the present study in SLA research paradigm, details are given in the following section of how written CF is compared with oral CF in terms of characteristics and classification and of how they were differentially treated in L2 acquisition research.

II. BACKGROUND

1. Distinction between Oral and Written CF

Oral and written CF contrast strikingly in terms of how they are characterized, classified,
Types of Information in Written Corrective Feedback and Its Efficacy on L2 Acquisition

and researched.

1) Differences in Characteristics

Oral CF is provided online in that it constitutes an immediate response to the learner’s erroneous utterance during the interactional exchanges. Written CF, on the other hand, is given offline. Its provision is delayed as it requires some time to review and correct the learner’s writing. This difference leads to oral CF typically imposing more cognitive load on the learner’s short-term memory than written CF as the former forces the learner to engage in making a spontaneous cognitive comparison between his/her ill-formed output and information delivered in CF (Sheen, 2010).

In typical instructional settings, learners are likely exposed to oral CF on other learners’ errors, let alone on their own errors. Such exposure is generally not available in the written mode unless CF is intentionally induced through a particular activity such as peer correction (Sheen, 2010).

Due to the differences in delivery of and exposure to CF, its efficacy may vary across oral and written modes. However, rarely has SLA research empirically addressed this issue with the exception of Sheen (2010), who failed to find any mediating impact of CF mode. Further research is definitely in order before reaching any compelling argument regarding the issue.

2) Differences in Classification

There are several ways to classify oral CF strategies. One common way is explicit/implicit distinction with explicit type being more overt in showing its corrective nature. Another criterion is whether correct form of error is contained in CF, which serves to distinguish input-providing from output-prompting CF. The former offers error-related information (e.g., its correct form) while the latter intends to supply opportunities for learners to self-correct their own errors. What is notable is that the two classifying dimensions are not exclusive to each other and a CF strategy can be described in both dimensions. For instance, recasts are a CF strategy which is implicit and input-providing type while elicitation can be categorized as explicit and output-prompting. An additional dimension involves focused/unfocused division according to how many linguistic features CF targets. Focused CF responds to only one or a limited number of pre-determined specific types of errors whereas unfocused CF is free of such restrictions.

Research has overall shown that explicit CF is more effective than implicit CF (Carroll & Swain, 1993; Carroll, Swain, & Roberge, 1992; Ellis, Loewen, & Erlam, 2006). However, findings are mixed over the relative efficacy of input-providing vs. output-
prompting CF. Some researchers argued for the superiority of input-providing CF particularly in the form of recasts (Doughty, 2003; Long, 2007) while others found more benefits in output-prompting type (Ammar & Spada, 2006; Lyster, 2004).

With regard to written CF, focused/unfocused division can be made while explicit/implicit distinction is not feasible as written CF is intrinsically explicit. No matter how written CF is delivered, its corrective intention is most likely to be perceived by the learner due to its visual and sustained nature. Written CF is most commonly categorized as two types, direct and indirect. Direct CF is one that provides correct form, corresponding to oral recasts in this regard. In contrast, indirect CF withholds correct from and, in this respect, is similar to oral output-prompting CF. However, the former is more limited in its capacity to prompt the learner’s output than the latter due to the inherent nature of written CF to delay the delivery of corrections.

Written CF studies produced mixed findings about the relative effectiveness of direct and indirect CF (Chandler, 2003; Ferris, 2006; Robb, Ross, & Shortreed, 1986; Semke, 1984). Ferris (2006) contended that indirect CF is more beneficial as it facilitates forming hypotheses about L2 form and rules. On the contrary, Chandler (2003) argued for the superiority of direct CF, asserting that it helps confirm previously formed hypotheses. Recent studies showed that focused written CF promotes learners’ linguistic accuracy (Bitchener & Knoch, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007). Ellis et al. (2008) further compared focused with unfocused CF but found no differential effectiveness.

3) Differences in Research Purposes and Design

Oral and written CF differ in several ways concerning how they have been investigated. One difference emerges in terms of who conducted research with what purposes. Oral CF research has predominantly been conducted by SLA researchers while written CF has mostly been studied by L2 writing researchers and practitioners.

SLA researchers initially concentrated on developing taxonomies of oral CF strategies and later shifted their focus to investigating the effectiveness of different types of identified CF strategies on L2 grammatical development. On the other hand, grammatical error correction received little attention in L2 writing instruction where the process-oriented perspective was prevalent. In this process-oriented approach, pedagogical priority was given to improving the quality of the learner’s original writing through recursive processes of drafting, receiving feedback, and redrafting over developing linguistic accuracy. Accordingly, written CF was primarily given on a variety of writing aspects such as content, organization, rhetoric, and mechanics rather than on linguistic accuracy. However, some researchers claimed that linguistic accuracy constitutes a component of L2
writing performance and should be addressed in L2 writing instruction (Eskey, 1983; Horowitz, 1986; Johns, 1995; Leki, 1990). Along this line, Ferris (2010) and other L2 writing specialists (Lane & Lange, 1999; Frodesen, 1991; Reid, 1994) made suggestions of how written CF should be provided for increasing linguistic accuracy in the process-oriented L2 writing class. To be specific, it was maintained that written CF should be “contextualized within the recursive writing process, prioritized to focus on serious and frequent patterns of written error, and personalized to the specific needs of the individual student writer” (Ferris, 2010, p. 185). Most L2 writing researchers and teachers have recently come to at least agree that “there is some role for written CF in L2 writing instruction” (Ferris, 2010, p. 184) that considers linguistic accuracy improvement.

In short, SLA researchers were mainly concerned with the effect of oral CF strategies on L2 acquisition while L2 writing experts mostly focused on examining the value of written CF in improving overall quality of the learner’s revised writing text. These different concerns of the two CF research fields are reflected in the ways that they each design CF studies (Ferris, 2010; Sheen, 2007). With goals to examine the effectiveness of CF on acquisition, SLA research typically has control and treatment groups and pretest-posttest designs. CF generally targets one or a few chosen error types and is consistently provided. In contrast, such SLA-specific design features are not necessarily required in L2 writing studies whose primary concern is whether written CF assists in producing the revised text of a higher quality by the same student writer. Thus, until recently, it remained to be resolved whether written CF could lead to L2 acquisition.

2. Integration of Written CF in SLA Research

The issue of whether and how written CF is related to L2 acquisition saw its turning point in Truscott’s (1996) influential study. Truscott (1996) claimed that written CF should be abandoned in L2 writing instruction as it is not only ineffective but even harmful to development of L2 implicit knowledge. His study drew earnest attention to the issue among researchers of SLA and L2 writing alike, sparking heated debate and leading to more subsequent studies. Change in research methodology and scope occurred in both fields to empirically tackle Truscott’s (1996) argument. Many L2 writing experts began to adopt SLA research design (Ashwell, 2000; Chandler, 2003; Ferris, 2006) while SLA researchers expanded their acquisition studies beyond oral to written CF (Bitchener & Knoch, 2008; Ellis et al., 2008; Sheen, 2007). A majority of subsequent studies demonstrated that written CF is effective in L2 acquisition, casting doubt over Truscott’s (1996) strong skepticism of written CF.

Some studies further looked into what types of written CF are more beneficial than others (Ellis et al., 2008; Sheen, 2007). This is closely linked to a question of what and
how much information should be required in CF for its optimal efficacy. Written CF can be categorized as direct, indirect, or metalinguistic CF according to the sorts of information it delivers, which include the existence, location, correct form, and nature of the error. As written CF is inherently explicit, its presence alone is sufficient to signal the occurrence of error irrespective of its type. Indirect CF indicates where error occurred while direct CF shows not only the error’s location but its corresponding correct form. Metalinguistic CF even helps the learner understand the nature of error by providing metalinguistic explanation of it.

Ellis et al. (2008) proved that focused direct CF was beneficial in the acquisition of English articles by L2 learners. Sheen (2007) compared direct CF with and without metalinguistic explanation and found that both were effective in the short-term period but only the former had a long-term effect. However, no study has compared all three types of written CF so far in terms of their acquisition-facilitating value. The present study addresses this limitation and, in so doing, aims to sort out the information that is required in written CF for its optimal effectiveness in L2 acquisition. This goal can only be pursued on the premise that written CF can enhance L2 learning and, as noted above, a variety of previous studies revealed that written CF plays an acquisition-promoting role. Nonetheless, the present study additionally seeks to confirm such CF benefits to consolidate the validity of comparing CF types containing different sorts of information. The following research questions are investigated to bring about the study’s purposes:

1. Is written CF beneficial to L2 acquisition?
2. What sort of information in written CF is most conducive to its efficacy on L2 acquisition?

III. METHOD

1. Design

An experiment with adult EFL learners was conducted to address the research questions. As this study aims to examine the effect of written CF on L2 acquisition, it employed methodological features that have typically characterized SLA research. Thus, a control group was formed to be compared with the treatment groups and a specific linguistic feature was chosen as the CF target. Materials for the experiment consisted of three sets of feedback treatment tasks and three sets of tests. The tasks were designed to produce writing productions in frequent use of the target form. The tests included the pretest, the immediate posttest, and the delayed posttest to trace changes in the learners’ L2
knowledge. Each test was composed of two versions to examine different levels of knowledge-declarative and procedural knowledge. Pretests were administered before the treatment tasks, immediate posttests right after their completion, and delayed posttests two weeks later.

2. Participants

The battery of participants was composed of a teacher and fifty four Korean EFL students who completed all the components of the experiment including three treatment tasks and three testing sessions. They were all sophomores of a university in Southeast Korea and were taking one of the courses by the researcher, who was also their teacher, in the semester that the experiment was carried out. The participants were assigned to one of three treatment groups or a control group. The treatment groups, which were labeled direct, metalinguistic, and indirect, were distinguished respectively according to the sorts of information in CF that they received during the treatment sessions. Further details of the distinctions in this regard are provided in a subsequent section. The control group took part in the treatment tasks but did not receive any CF. The group composition for the study is summarized in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Group (DG)</td>
<td>13</td>
</tr>
<tr>
<td>Metalinguistic Group (MG)</td>
<td>13</td>
</tr>
<tr>
<td>Indirect Group (IG)</td>
<td>14</td>
</tr>
<tr>
<td>Control Group (CG)</td>
<td>14</td>
</tr>
</tbody>
</table>

3. Target Structure

English articles were selected as the target feature. The decision was made as a result of considering methodological and pedagogical advantages the feature could have when written CF is examined in an EFL context involving Korean learners. On the methodological front, Korean language does not have a structure equivalent to English articles, which causes learners to frequently make errors. Nonetheless, they are seldom corrected in typical instructional settings in Korea as articles are neither salient graphically or phonologically nor carry high communicative value. In fact, correcting article error did not constitute part of instruction for the English course that the participants of the study were taking, which made it possible to isolate the effect of CF offered in the experiment. From a pedagogical perspective, targeting articles has a high validity as they turned out to
be one of the most frequently used structures in English (Sinclair, 1991).

English article rules are extremely complicated in nature and hard to figure out particularly for learners whose L1 does not have an equivalent structure such as those in the current study. In this light, written CF was given only on errors involving two major functions of the articles: indefinite article *a* used to refer to someone or something for the first time and definite article *the* with a function of indicating that its following noun was previously mentioned (i.e., anaphoric reference). Example 1 is a sentence that includes both functions.

Example 1: Article functions of first and second mention

I recently got a new computer and I will use the computer mostly for surfing the Internet.

4. Sorts of Acquisition for Measurement

Although L2 writing experts and SLA researchers sought to clarify the impact of CF on L2 ‘learning outcome’, they had a different view of what it meant. As noted early, it was considered an equivalent of acquisition for SLA researchers. In contrast, L2 writing specialists evaluated it on the basis of improvements in overall quality of the learner’s revised writing text. Yet some L2 writing practitioners recently adopted the SLA view and written CF is increasingly investigated within SLA research framework.

The notion of ‘acquisition’ also has yet to be well defined in the literature. Frequently, acquisition is connected with two different levels of knowledge. For example, skill learning theory makes distinction between declarative and procedural knowledge (DeKeyser, 1998, 2001). The former refers to knowledge about the facts while the latter represents ability to use declarative knowledge. Declarative knowledge may be retrieved from short-term memory, which demands controlled processing that requires a lot of attentional resources and conscious awareness. For this reason, declarative knowledge is viewed as more of the explicit type. Procedural knowledge, on the other hand, is established through practicing declarative knowledge and is retrieved from long-term memory, which involves automatic processing that does not heavily rely on attention and awareness. Procedural and automatized knowledge is therefore more of the implicit type. In skill learning theory, acquisition is defined as proceduralizing and automatizing declarative knowledge. Some researchers took quite a different view of acquisition, claiming it to be progress along a sequence of developmental stages for chosen target features (Mackey & Philp, 1998; Pienemann, Johnston, & Brindley, 1988).

The majority of CF research has examined acquisition only in terms of procedural knowledge. The present study, however, looked into both declarative and procedural
knowledge. Accordingly, two different versions of tests were developed and administered at each testing session: an error correction test for assessing the learners’ declarative knowledge of the target structure and a narrative writing test for procedural knowledge.

5. Tests for Assessing Acquisition

1) Error Correction Test

An error correction test contained 25 items. Each item had two sentences and one of them was underlined, indicating that it was grammatically wrong. The learners were asked to correct and rewrite the underlined sentences. Twenty items involved misuse of English articles. Example 2 below is one item from the test.

Example 2: An item of the error correction test

There may be restaurant around here. I want to eat something soon because I am really hungry now.

The other five items were distracters and their ungrammaticality was caused by errors in various grammar features such as tense, word order, and verb form. The article-targeting items only were scored by giving 1 point to each correctly changed sentence, the perfect score being 20 points. As the test imposed no time limit, the learners were able to review the items as many times as they wanted and to mobilize as much attentional resources as they needed to retrieve knowledge about article-related rules to complete the test. In this sense, the test was regarded as measuring the learners’ declarative/explicit knowledge of the two major functions of English articles.

2) Narrative Writing Test

This test was adapted from those used in Muranoi (2000) and Sheen (2007). It consisted of two sets of four sequential pictures with a few word prompts next to each picture. For each set, the learners were asked to write a coherent story by making reference to the pictures and accompanying words. A learner’s score was calculated as a percentage by dividing the number of articles correctly supplied by the number of contexts where articles should be obligatorily used and contexts where articles are not obligatory but used by the learner. As the constrained free writing, like ones in this study, does not guarantee that articles are only usable in obligatory contexts, the following coding guidelines were adopted as presented in Sheen (2007).
1. When one article was followed by two noun phrases as in ‘a shirt and skirt’, only the first noun phrase was coded as it was not clear whether the second noun was the result of elision.
2. When an article and a determiner coexisted as in ‘a his dog’, the noun phrase was not coded.
3. Idiomatic expressions containing articles as with ‘at the moment’ were not coded.

There was a five-minute time limit to write a story for each set. This pushed the learners to rely on the kind of knowledge retrievable through subconscious and automatic processing. In addition, creating a coherent story required them to focus preferentially on words of high communicative value. Given a low communicative value involved in articles, the learners were not likely to consciously pay attention to article usage with such a time-limit. For this reason, scores of this test were regarded as reflecting the learners’ level of procedural/implicit knowledge of the target structure.

6. Treatment Tasks

Three treatment sessions continued over a two-week period, at each of which the learners performed a narrative writing task. Three stories each of 78 words with 24 indefinite articles and 30 definite articles in total were created. In each session, the learner’s task was to read a story and write it as accurately as possible. Specifically, the teacher distributed a story with a writing sheet to each learner, who then read it for 3 minutes. After collecting the story, the teacher read it aloud. The learners then wrote the story on the writing sheet as closely to the original one as they could remember for 5 minutes. The sheet had a few key words from the story. The teacher’s act of reading the story aloud and provision of key words were intended to reduce the learner’s processing load on retrieving the story. Lastly, the teacher collected the learners’ written stories and later corrected article errors only.

When correcting errors, the teacher gave different types of CF to the treatment groups. Direct CF was provided to the direct group in the way that article error was underlined and its correct form was written above it. For the metalinguistic group, metalinguistic feedback was supplied that indicated error with a coding symbol and then added metalinguistic explanation the symbol represented. Indirect CF was presented to the indirect group in the way that marked error by only underlining it (see Appendix for CF samples).

The learners received their writings with corrections in the following treatment session and reviewed their errors and corrections for 5 minutes. For the metalinguistic group, an additional sheet with metalinguistic explanations was provided. After reviewing, the learners engaged in the next task that followed the same procedure as the previous one.
with a different story. Stories written in the third session were given back with corrections to the learners prior to the first posttest that was administered two days later. The control group took part in the same writing tasks as the treatment groups but did not receive any CF.

7. Procedure

The learners were assigned to one of the treatment groups or a control group before the experiment began. Each group constituted one of four classes the researcher was teaching during the period of the experiment, meaning that the learners in each group were from the same class. This decision was made on the ground that the classes were similarly at low-intermediate English proficiency level according to the university’s placement test results.

The learners completed the pretests in the first week of the experiment to make sure that levels of knowledge about English article usage were comparable across all groups and to obtain base data to be compared with those of the posttests. The learners then engaged in three treatment tasks over the next two weeks. The immediate posttests were administered right after the learners reviewed their third narratives and corrections, which occurred two days after the last writing task. The learners completed the delayed posttests two weeks later. The whole procedure is summarized in Figure 1.

**FIGURE 1**

Whole Procedure of the Experiment

- Pretest

- Treatment session 1
  - Treatment groups: Narrative writing 1
  - Control group: Narrative writing 1

- Treatment session 2
  - Treatment groups: Receiving CF on narrative writing 1 / Narrative writing 2
  - Control group: Narrative writing 2

- Treatment session 3
  - Treatment groups: Receiving CF on narrative writing 2 / Narrative writing 3
  - Control group: Narrative writing 3

- Immediate posttests
  (Treatment groups: Receiving CF on narrative writing 3 prior to tests)

- Delayed posttests
8. Data Analysis

Percentage scores were computed from the learners’ performance on the narrative writing tests. However, raw scores were used for the error correction tests as the possible score range was the same across all groups, making it unnecessary to convert raw scores into percentages. SPSS was used to analyze all scores, performing one-way ANOVAs, posthoc pairwise comparison tests using Tukey, and two-way ANOVAs. The alpha level determining significance was set at $p < .05$ for inferential statistics.

IV. RESULTS

Table 2 presents the mean scores and standard deviations for the four groups over three error correction tests. A two-way ANOVA was performed with mean scores as the dependent variable and four groups and three testing times as independent variables. The results revealed that there were both the main effect for group differences, $F(3, 150) = 8.561, p < .001$, and the main effect for time, $F(2, 150) = 59.999, p < .001$. A one-way repeated measures ANOVA showed that the gains the control group made were significant, $F(2, 26) = 10.129, p = .001$. Given that all the groups produced significant longitudinal gains, it is instrumental to examine whether there was a time-group interaction. Results of a two-way ANOVA showed that the time-group interaction did not reach a statistically significant level, $F(6, 150) = 2.024, p = .066$, ns.

One-way ANOVAs were performed to see how the groups performed differently at each testing time. An analysis of mean scores on the pretest showed no statistical significance, $F(3, 50) = .023, p = .995$, ns, confirming that all groups were at the comparable level of explicit knowledge regarding article usage when the experiment began. Analyses of posttest means revealed that the groups significantly differed on both the immediate and delayed tests, $F(3, 50) = 5.387, p = .003; F(3, 50) = 4.521, p = .007$, respectively. Tukey’s posthoc pairwise comparisons, which were carried out to locate where the significant difference lies, revealed that the metalinguistic group performed better than the indirect and control groups on the immediate test. The metalinguistic group’s superiority over the control group was sustained on the delayed test.
Table 2 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

Table 3 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

Table 3 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

Table 3 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

Table 3 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

Table 3 displays the mean scores and standard deviations for the narrative writing tests over time. A two-way ANOVA with mean scores as the dependent variable and four groups and three testing times as independent variables revealed that there were the main effect for group differences, $F(3, 150) = 4.975, p = .003$, and the main effect for time, $F(2, 150) = 63.917, p < .001$. As in error correction tests, a one-way repeated measures ANOVA showed that even the control group made significant progress, $F(2, 26) = 15.674, p < .001$. As all the groups increased test scores over time, a check over time-group interaction was in need. Results of a two-way ANOVA revealed that the statistically significant time-group interaction did not exist, $F(6, 150) = 1.807, p = .101$, ns.

A one-way ANOVA analysis of the groups’ pretest mean scores produced no statistical significance, $F(3, 50) = .073, p = .974$, ns, indicating that all groups were comparable in their implicit knowledge of English article usage at the beginning of the experiment. Additional one-way ANOVAs for the immediate and delayed posttest scores revealed that the group differences were statistically significant on both tests, $F(3, 50) = 4.536, p = .007$ for the immediate posttest; $F(3, 50) = 3.375, p = .025$ for the delayed posttest. Tukey’s posthoc pairwise analyses found that the significant difference in mean scores emerged between the metalinguistic and control groups. The former group remarkably outperformed the latter group on both the immediate and delayed tests.

V. DISCUSSION

Research question 1 was concerned with whether written CF is facilitative of L2
acquisition. To answer the question, the study measured the learners’ abilities to monitor and use two functions of English articles (i.e., explicit and implicit knowledge of articles), using error correction tests and narrative writing tests respectively. The results revealed that the metalinguistic group performed better than the control group on both types of posttests. This means that written CF was beneficial to explicitly developing knowledge of article usage and to implicitly increasing accuracy of article use. Moreover, such positive effects continued for long as the metalinguistic group outperformed the control group not only on the immediate tests but also on the delayed tests. Thus, this study provides another empirical counterevidence to Truscott’s (1996) claim about the ineffectiveness of written CF on the development of L2 implicit knowledge.

Research question 2 asked what information should be contained in written CF to maximize its efficacy in L2 acquisition. As described previously, CF offered to the treatment groups differed in terms of the four aspects of information-existence, location, correct form, and nature of the learner’s error. CF for the indirect group contained information on existence and location of error while CF for the direct group provided correct form of error in addition to information on its existence and location. The metalinguistic group’s CF had all the four aspects of information including information on error nature (i.e., metalinguistic rule explanations).

The results of the experiment showed that only the metalinguistic group outperformed the control group. Performances among the other treatment groups and the control group were not significantly differentiated. This raises two interrelated issues: why the metalinguistic group made progress and why the direct and indirect groups performed similarly and were not distinguishable from the control group. These questions can be answered within the theoretical framework of Schmidt’s (1993, 1995, 2001) noticing hypothesis, whose main argument is that attention to L2 grammatical features is a necessary path to their acquisition. Attention involves conscious awareness which occurs at two different levels (Schmidt, 2001). One is the level at which the learner ‘notices’ a new form and the gap between his/her interlanguage output and target language form. The other level refers to ‘understanding’ rules and patterns involving an L2 form. It is argued that understanding requires more complex cognitive processing and leads to greater L2 learning in comparison with noticing (Schmidt, 2001; Sheen, 2010).

It should be noted that as for CF, perceiving its corrective intention is critical for its efficacy on L2 acquisition. This was aptly pointed out in Carroll’s (2001) autonomous induction theory, which claimed that CF cannot work for L2 learning if learners do not recognize its corrective function. It is because, without its recognition, CF is unlikely to lead to noticing and understanding. Thus, upon receiving CF, the learner’s attention needs to be paid preferentially to its corrective force to trigger further cognitive processes necessary for L2 acquisition. In short, three facets of cognitive processing need to be
sequentially involved in CF-facilitated L2 acquisition: recognizing corrective force of CF, noticing correct form of error, and understanding rules that govern the form.

With regard to perceiving the corrective force of CF, it is unlikely that there was notable difference among the treatment groups because written CF of any type is invariably explicit in nature as noted previously. This facet of processing therefore does not account for the metalinguistic group’s greater gains in knowledge. In terms of the two levels of conscious awareness (i.e., noticing and understanding), Sheen (2010) contrasted written direct CF with written metalinguistic CF. She argued that both types of CF can raise awareness at the level of noticing due to correct form they provide. Visual juxtaposition of error with its correct form in written mode is most likely to lead to learners’ noticing and noticing-the-gap with little difficulty. However, with an aid of metalinguistic explanation, metalinguistic CF can further enhance learners’ awareness at the level of understanding. The increased understanding was most likely a main source of the metalinguistic group’s improvement. In contrast, direct CF, in the absence of such metalinguistic explanation, can only promote understanding “if learners reflect on their errors and work out why they were incorrect and what the rule is” (Sheen, 2010, p. 226). Then, it is highly probable that although the learners in the direct group noticed correct form, they did not go so far as to try to figure out the rules that govern the article functions, which in turn may have brought about little understanding and, consequently, little progress in their knowledge.

Given that indirect CF did not even provide correct form, the indirect group may not have been able to engage in noticing let alone understanding. And this could be at least partially responsible for the indirect group’s relatively poor posttest performance. It is not clear, however, whether this group’s weak performance was mainly due to the learners’ failure to work out (i.e., understand) the rule, or to notice, or the combination of both.

One point to be made here is that the control group made significant gains over time, indicating that there was a test practice effect. It arose in part from the same tests being administered over three testing periods. Given that all the groups went through the same testing procedure, it is estimated that the test practice effect also accounted for much of the improvement the treatment groups exhibited over time. However, the fact that only the metalinguistic group did better than the control group suggested that metalinguistic CF had a unique efficacy beyond such test practice effect.

It should be noted that metalinguistic CF facilitated not only the learners’ explicit knowledge but also implicit knowledge. This is particularly significant in rejecting Truscott’s (1996) argument against written CF effectiveness because he looked into the effect of CF only in relation with implicit knowledge. Receiving metalinguistic comments on their errors, the learners in the metalinguistic group may have had little difficulty in developing explicit/declarative knowledge when CF targeted only two functions of articles. However, receiving such CF was probably not sufficient for implicit/procedural
knowledge development, considering the limited number of possible CF instances in the treatment tasks. How then did this group get to develop implicit knowledge? One account is plausible within skill learning theory, according to which, as mentioned early, practicing explicit knowledge is what transforms explicit into implicit knowledge. The metalinguistic group members may have established explicit knowledge through CF on their writings in early treatment tasks and utilized the ensuing writing tasks as opportunities to practice and proceduralize the explicit knowledge, resulting in ultimately developing implicit knowledge.

The second issue raised above was about non-differentiation in posttest performances among the direct, indirect and control groups. The non-differentiation may be taken as evidence that information delivered by CF for the direct and indirect groups (e.g., existence, location, correct form of error) is not sufficient for written CF to contribute to L2 acquisition. However, such conclusion is probably too simplistic and premature when learning process is viewed as involving two separate ways—deductive and inductive. The metalinguistic group was able to engage in deductive learning as their CF provided what was needed for such learning (i.e., metalinguistic accounts of rules). In contrast, the direct and indirect groups did not receive information to enable deductive learning. This, however, does not necessarily justify their inferior test performance. With information offered in CF, the direct and indirect groups could have at least tried inductive learning, struggling to work out the rules of articles as noted earlier (Sheen, 2010). The ineffectiveness of CF seen in the two groups thus demonstrates that they did not engage in inductive learning.

What then prevented the direct and indirect groups from making inductive endeavors to make out the target rules? One plausible explanation is that they did not get enough instances of correction to inductively come up with the article rules. There were only three treatment tasks and a relatively short story of 78 words was used in each task. The number of errors and corrections available may have been too limited to make inductive learning possible. Another explanation is the possibility that the learners did not appreciate CF so much. Articles are quite low in communicative value but highly complex in their rule application as their choice is constrained by consideration of both linguistic and pragmatic environments. This means that misuse of articles rarely disturbed conveying meaning of the story while the rules for their correct use were not easy to capture. Thus, it can be argued that even though the learners were explicitly given CF on articles, they may not have felt a need to focus on such form and discover its related rules on their own as it was deemed more crucial to retain and convey an original storyline in completing their writing task at hand.

The findings of this study could help assure language teachers in general and writing instructors in particular that the effect of written CF is not limited to improving linguistic
Types of Information in Written Corrective Feedback and Its Efficacy on L2 Acquisition

accuracy of a revised writing text but it also works to further interlanguage development. The study also demonstrated that the positive effect does not manifest uniformly across CF types. In providing written CF, a teacher might want to make sure that it contains metalinguistic explanation relating to the learner’s error as the study showed that those receiving CF with such aid made the greatest progress in their knowledge.

It should be noted that choosing a particular pedagogical practice for L2 writing class based upon findings of research conducted in SLA framework could raise a thorny question of practical applicability. SLA-oriented research that produced compelling evidence for the positive role of written CF in L2 acquisition has targeted only a single linguistic feature. However, in a real writing class, there are various writing aspects of content, organization, rhetoric, and mechanics that CF can and should target aside from linguistic errors. Moreover, even when focusing CF exclusively on linguistic accuracy, few writing teachers, in reality, have the luxury of being able to correct every linguistic error with its metalinguistic explanation as implemented in this study and others (Ellis et al., 2008; Sheen, 2007, 2010). Thus, L2 writing practitioners may legitimately doubt the feasibility of applying SLA-oriented research findings to a real writing class (Ferris, 2010). One potential solution to the applicability problem is what is called dynamic written CF advanced by Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, and Anderson (2010). They suggested that the alternative approach is to limit the length of the student writing for which error correction is intensively supplied. This way, feedback becomes practically manageable because “with a shorter piece of writing, teachers can identify all linguistic errors produced by their students, without overwhelming themselves or their students” (Hartshorn et al., 2010, p. 89).

VI. CONCLUSION

The study looked into whether written CF is effective in L2 acquisition and, if so, what information on the learner’s error should be contained in CF to boost its effectiveness. It was found that written CF is beneficial to improving both explicit and implicit knowledge of English article rules, refuting Truscott’s (1996) argument against the utility of written CF in developing implicit L2 knowledge. However, such positive efficacy surfaced only when CF provided metalinguistic explanation, which shows that only deductive learning, rather than inductive learning, was triggered through CF.

It, however, would be premature to generalize the dependence of written CF on the presence of metalinguistic explanation for its effectiveness based solely upon what the current study found. The study has at least two limitations in this regard. As noted above, the direct and indirect groups perhaps failed to produce gains in knowledge due to
inadequate supplies of error correction over three task sessions. A longer treatment might have enabled inductive learning and produced beneficial effects of non-metalinguistic CF (i.e., direct and indirect CF). Another limitation is that the study did not isolate the effect of pure metalinguistic explanation as metalinguistic CF included other information such as existence, location, and correct form of error. It is therefore still unclear whether or not the effectiveness of metalinguistic CF came exclusively from the workings of metalinguistic explanation.

Future research then should be designed with CF that is given in more treatments over a longer period and that isolates metalinguistic explanation from other error-related information. This way, the question of what constitutes optimal information required in written CF for L2 acquisition will be more adequately addressed. And this in turn can more informatively help L2 writing teachers to implement written CF in real writing classrooms.

There are two additional issues that merit further SLA-oriented writing investigation. Written CF studies adopting SLA methodology have mostly chosen English articles as a target linguistic feature (Bitchener & Knoch, 2008; Bitchener, Young, & Cameron, 2005; Ellis et al., 2008; Sheen, 2007) and the present study was no exception. Inspecting the efficacy of written CF targeting a variety of different grammatical forms and structures is definitely in order. The other issue is concerned with the value of revision. Traditionally, the learner’s capability of revising his/her original writing text to produce the same text of higher quality was taken as a measure for the writing class’s success. Revision was the purpose and object of L2 writing class and treated as a dependent variable in L2 writing studies. Recently, SLA researchers took a different perspective of revision (Ferris, 2004, 2006, 2010; Sachs & Polio, 2007). Ferris (2010) claimed that “revision and editing of a text after receiving written CF may be a helpful and perhaps necessary intermediate step between expert feedback about a target feature and long-term acquisition of that feature” (p. 189). The feasibility and value of treating revision as a pedagogical means rather than goal has yet to be empirically proven. To answer this issue, future research should take revision as an independent variable and examine its efficacy on L2 acquisition.

Despite having a long way to go before getting any clear picture of how written CF is connected with L2 acquisition, research in this line has its unique value. Methodologically, written CF research is relatively easy to conduct in that written errors are quite easy to identify and corrections are not so hard to manipulate in comparison with its oral counterparts or other pedagogical interventions. More importantly, as Sheen (2010) pointed out, CF research is directly and closely relevant to language pedagogy as CF is a pedagogical technique that teachers frequently use in their classrooms. This close association signifies that CF research can inform language teachers of effective ways to correct learners’ errors. In a similar vein, a classroom rich in errors and corrections may
serve as an appropriate site for conducting CF research. In this sense, “CF is an ideal object of inquiry for researcher-teacher collaboration and constitutes an area of inquiry that can connect theory, research, and practice” (Sheen, 2010, p. 177). As written CF is invariably explicit, it is less likely to be missed out by learners than oral CF. Thus, the relevance and connection of CF with pedagogy are most likely stronger when it is provided in written than oral mode. This leaves a certain legitimacy to call for continued and enriched written CF studies, counterbalancing the predominance of oral CF studies to date in SLA literature.

REFERENCES


### APPENDIX

Written Corrective Feedback Samples

1. Direct CF

   \[
   a
   \]

   A moment later, he saw the dog biting the bag.

2. Indirect CF

   He looked into _ bag but a sandwich was not in it.

3. Metalinguistic CF

   A man was traveling in _(*) small town near a(+) Chicago. He went into a fast food restaurant in a(#) town.

\[*\): \(a\) is needed before a singular noun that is first mentioned.

\[\(+\)\): \(a\) is not used before a city name.

\[\(#\)\): \(the\) is needed before a noun that is previously mentioned.
note. 1) All samples are from this study’s data.
2) Metalinguistic explanation was provided in Korean to the learners.

Applicable levels: tertiary education
Key words: written corrective feedback, conscious awareness, noticing, understanding, explicit knowledge, implicit knowledge

Sung-Soo Jang
College of General Education
Inje University
607 Obang-dong, Gimhae
Gyeongsangnam-do, 621-749, Korea

Fax: (055) 339-3734
Email: elansing@hanmail.net

Received in June, 2012
Reviewed in July, 2012
Revised version received in August, 2012