English Language & Literature Teaching, Vol. 22, No. 1 Spring 2016

Effects of Direct Feedback on Grammatical Accuracy and Explicit/Implicit Knowledge of Target Forms[*]

Daniel Corks[**]
(Dongshin University)
Eun Sung Park
(Sogang University)

Corks, Daniel & Park, Eun Sung (2015). Effects of direct feedback on grammatical accuracy and explicit/implicit knowledge of target forms. *English Language & Literature Teaching, 22*(1), 1-22.

This paper describes a case study that examined learner response to unfocused written corrective feedback and its effects on grammatical accuracy for high-beginner learners of L2 English. The study was motivated by the ongoing debate (i.e. the "Truscott debate") over the effectiveness and potentially harmful effects of grammar correction in L2 writing. Two participants wrote weekly journals over a 9-week period and received direct corrective feedback on each journal entry; five grammatical features were targeted for feedback without the knowledge of the participants. The effect of the corrective feedback was measured by a number of instruments aimed at gauging the participants' explicit and implicit knowledge of the target features. In addition, the journal entries were analyzed for potential secondary effects of the feedback. The results showed small improvements for accuracy and explicit/implicit knowledge, but a lack of consistency for individual features, suggesting that the efficacy of feedback was minimal overall. Secondary effects were found to be either marginal or attributable to factors other than corrective feedback. The current results support Truscott's position and suggest that incidental learning of redundant grammatical features does not take place.

[grammar correction/explicit knowledge/implicit knowledge/L2 writing/direct feedback/written corrective feedback/second language acquisition]

[*] The current research is based on the first author's MA thesis.
[**] Daniel Corks: first author; Eun Sung Park: corresponding author

I. INTRODUCTION

Many years have passed since Truscott (1996) called for an end to the common practice of correcting written grammatical errors of second language (L2) learners. Arguing that grammar correction is ineffective and even harmful, he cited, among many other points, theoretical problems regarding the interface between explicit and implicit knowledge and practical problems pertaining to determining where each learner is in their developmental sequence. Much research has since been conducted on the topic but has yet to produce conclusive findings.

The controversy surrounding the efficacy of corrective feedback that Truscott's original paper and his numerous follow-ups (1999, 2001, 2007, 2010a, 2010b) stimulated has been criticized for distracting researchers away from other areas of importance (see Bruton, 2009). However, we can credit Truscott with motivating the shift in focus of the correction debate from a learning-to-write perspective, focused on producing proficient and independent writers, to a writing-to-learn perspective, focused on whether the process of writing and receiving feedback can aid the development of a learner's interlanguage. While learning-to-write is primarily the domain of language instructors, writing-to-learn considers the possibility of developing grammatical knowledge through writing and whether this knowledge can be of use to learners in other communicative settings, such as speaking.

The focus of the present study is to examine whether corrective feedback on grammatical errors in L2 writing (henceforth "CF") causes improvement in accuracy in future writing or leads to the development of learners' explicit or implicit knowledge. This study also attempts to answer the broader question of whether CF should be used in L2 pedagogy by combining the two questions above with concerns regarding potential negative effects of receiving CF on a learner's writing behaviour. Each of these three factors will be considered here in turn.

II. LITERATURE REVIEW

1. Theoretical Background

One way to summarize the paradigm change of CF studies from learning-to-write to writing-to-learn that occurred as a result of Truscott (1996) is to say that the focus of attention shifted from writing produced by learners to the learners themselves. For instance, prior to Truscott's critique of CF, the majority of papers on CF considered showing improvement from one revision of a paper to the next to be sufficient for arguing for the

efficacy of CF, in essence valuing individual pieces of writing more than the individuals producing them. The writing-to-learn perspective, by contrast, has considered improvement in future writing to be the more apt measure of the effect of CF, as it demonstrates that the learner has acquired some knowledge or ability that they are able to use going forward.

A second example of the difference between these two approaches is the change from considering all errors in a piece of writing to focusing specifically on the grammatical errors. In papers prior to Truscott (1996), errors of orthography and punctuation, which are relatively easy to self-correct and only of use in writing, were not distinguished from grammatical errors, which are notoriously difficult to self-correct and reflect a learner's core linguistic competence (cf. Robb, Ross & Shortreed, 1986).

A final example of this difference was the introduction of a secondary concern to the correction debate regarding potential negative impacts of CF on a learner's subsequent writing. Although Truscott's later papers did not address this concern, his original critique of CF argued that corrections on a learner's writing can be actively detrimental, especially if those corrections are numerous or are on items that learners often have difficulty with. In these cases, CF can be burdensome and demotivating for learners, to the point where they purposely avoid using particular structures that they are unsure of (and thus fail to push the limits of their L2 knowledge). Alternatively, learners may either shorten or simplify their writing in order to receive fewer corrections. This criticism has received less attention in the literature, but its importance remains since arguing that a learner's grammatical accuracy has improved as a result of receiving CF requires that researchers be able to prove that any gains were not simply the result of the learner avoiding more difficult structures.

Demonstrating CF's effectiveness through improvement in grammatical accuracy in a new writing may be sufficient from a pedagogical perspective, but in order to build upon existing second language acquisition (SLA) theory, more robust proof is required. It is from this viewpoint that Truscott, taking the position of an L2 researcher, levelled a set of theoretical arguments against the use of CF. The most significant of these is the question of whether CF can stimulate the development of explicit knowledge and further, implicit knowledge of grammatical features. He argued that feedback in the form of direct correction may help learners improve their grammatical accuracy in future writing or perform well on a grammar test, but that it is unlikely to be of use in communicative settings. In other words, CF may aid the development of explicit, declarative knowledge but will not become implicit, procedural knowledge. This argument takes a very specific stand on the question of the nature of the interface between explicit knowledge and implicit knowledge – i.e., that there is no interface. Most SLA researchers adopt one of two alternative views: that there is a weak interface that allows explicit knowledge to supplement and support implicit knowledge when difficulties arise; or that there is a strong

interface that allows explicit knowledge to become automatized with practice and over time eventually become implicit knowledge (see Kim, 2014, for a discussion of each type of interface). Shintani and Ellis (2013) identify this uncertainty regarding the interface between the two knowledge types to be central to the debate of whether CF can aid the development of overall language proficiency.

Assuming the existence of an interface, however, does not answer the empirical question of implicit knowledge development, for it would only say that such development is possible in theory. In practice, a spoken dialogue often involves a need for spontaneous communication, thereby providing motivation for the development of automaticity and implicit knowledge. However, writing (in most cases) affords learners ample time to access their declarative knowledge if necessary and is without the social pressure of having to reformulate their output on the spot for a live interlocutor who is unable to comprehend an utterance. If this interactional pressure is absent, or if the output is not "pushed" in Swain's terms (1985), the lack of demand on the learner's implicit knowledge system means it would not be pushed to handle the more challenging environment. In sum, beyond the theoretical argument that implicit knowledge *can* develop from receiving CF, empirical data is required to prove that implicit knowledge *does* develop from receiving CF. Writing tasks and CF, as typically employed in the L2 classroom, rarely involve a strong demand on a learner's implicit knowledge. Hence, instruments specifically designed to measure implicit knowledge must be included in the experimental design in order to address this question.

2. Previous Studies

Despite a flurry of research carried out in response to Truscott (1996), conclusive results have yet to surface (see reviews by Bitchener & Knoch, 2008; Truscott, 2007; Van Beuningen, 2010). The primary issue with this body of work is that the differences between the research focus, methodology, and operationalizations are so broad that the results are incomparable (Ferris, 2010; Van Beuningen, 2010). These differences include whether a given study looked at revisions or new writing, the type of CF provided, (i.e. direct vs. indirect), the numbers of targeted linguistic features for CF treatment (i.e. focused vs. unfocused), the categorization of errors (e.g. grammatical vs. nongrammatical, global vs. local, etc.), and whether the study involved just a single treatment or multiple treatments with a longitudinal design.

One set of research by Bitchener and colleagues has attempted to rectify this situation by conducting a series of studies using similar methodology and targeting the same grammatical features (e.g. Bitchener & Knoch, 2008, 2009, 2010; Sheen, 2007; Shintani & Ellis, 2013). These studies were all single treatment studies using a pre-test, post-test and delayed post-test design, and provided direct, focused feedback specifically targeting the

same linguistic features – i.e. two functions of the English article system. Out of the nine studies published in this set, only one, Shintani and Ellis (2013), failed to find a positive effect for CF in a new writing, though the authors did not interpret the results as evidence against the effectiveness of CF. These results are still preliminary and limited in scope, but they have succeeded in establishing a methodology that future researchers can use to test other grammatical features or other types of CF. For instance, a study by Park (2015) employed a similar methodology to explore differential responses to feedback of the two functions studied by Bitchener and colleagues. It is worth noting that Shintani and Ellis (2013) employed additional instruments to assess learners' explicit and implicit knowledge, but their lack of conclusive findings for accuracy overall eliminated the possibility of gaining further insights from their study through these instruments.

Another recent study which is relevant to the current research is Van Beuningen, De Jong and Kuiken (2012) which targeted Dutch high school students. It is one of the few longitudinal studies that separately analyzed grammatical and nongrammatical errors. Their results showed that only direct CF (i.e., correcting errors by providing the correct form) produced significant improvement for grammatical errors, while only indirect CF (i.e., indicating the location and type of error without providing the correct form) produced statistically significant improvement for nongrammatical errors, such as lexical, orthographic, or pragmatic errors. Their results suggest some important implications for treating grammatical and nongrammatical errors as separate categories.

Van Beuningen (2010), in her extensive review of existing literature on written corrective feedback, makes two suggestions for future research which may provide useful insights on the practice of corrective feedback. The first of these was the suggestion that future studies conduct individual analyses on the grammatical features as individual grammatical features differ greatly in terms of their complexity, redundancy and salience. There is potentially much that can be learned from analyzing individual features separately, just as conducting separate analyses of grammatical and nongrammatical errors provided new insights in Van Beuningen *et al*. (2012). The second suggestion was her call for longitudinal studies of CF which allows a qualitative look at learner responses to CF and their improvement over time. Most studies thus far have been single treatment studies or purely quantitative in their reporting. Given the potential insights these suggestions can offer, we have incorporated each of these suggestions into the present study.

In light of the foregoing reviews, the current study sought to explore the impact of direct, unfocused corrective feedback on learners' written accuracy, and the development of their explicit and implicit knowledge of select grammatical features. In addition, the possibility of corrective feedback having a negative impact on learners' writing, such as avoidance or general depreciation in quality and complexity, was also examined. The following three research questions guided the current study:

1. Does the provision of unfocused direct CF improve learners' grammatical accuracy in writing?
2. Does the provision of unfocused direct CF affect learners' development of explicit and implicit knowledge?
3. Does the provision of unfocused direct CF have any secondary effects on a learner's writing behaviour?

III. METHOD

1. Participants

Two first-year university students, henceforth 'Tom' and 'Alex' (both 19 years of age) were selected from a pool of L1 Korean-speaking students with no prior relationship with the researcher. Both were of high-beginner proficiency, able to complete a 300-500 word writing task in approximately one hour. Both had recently completed a low-level English course at a university in Seoul, South Korea. The course is offered to lower ability students to prepare them for the university's mandatory 'core' English classes.

Both Tom and Alex's English learning experiences typify that of South Korean students in their age range: roughly 10 years of formal English education with a heavy focus on complex grammar and academic vocabulary for exam preparation, especially during the high school years, with little attention given to productive language skills or practical communication. Neither participant had overseas experiences or any similar extra-curricular exposure to English.

2. Target Structures

As noted, the participants received unfocused feedback on their writing. For the purpose of the current study, five grammatical structures were chosen for analysis: the indefinite article (henceforth "Article"), the regular plural suffix ("Plural"), the third person singular present simple inflectional suffix ("Third Person"), the regular past tense inflectional suffix ("Past Tense"), and negation ("Negation"). These structures were chosen based on the following reasons: a) they often appear in the type of writing tasks used in this study, b) they are persistent sources of error for both participants, and c) they were deemed to be addressable at the participants' level.

3. Treatment

This study spanned over a period of 16 weeks during which nine weekly journal-

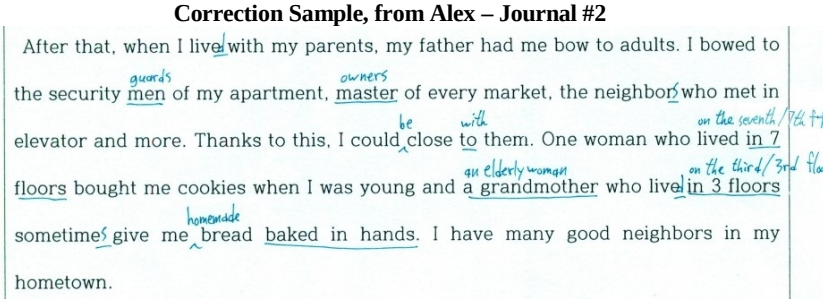
response writing tasks were collected, and feedback on errors on each journal entry was provided to each of the two participants in the form of direct CF. Each week of the 9-week treatment phase consisted of a written journal on an assigned topic, a one-on-one meeting with each participant to discuss the researcher's feedback on the journal, as well as the participant's revisions to the journal based on the feedback.

The participants were given four days to write and digitally submit a journal response on an assigned topic. The topics provided were ones that the participants could respond to without need for detailed planning or careful argumentation, thereby encouraging them to "just write"[1], the topics were not selected specifically to encourage the use of certain target features. Participants were given a rough minimum length of one typed page and some encouragement to write formally, but that aside there was little effort to constrain their output.

In an effort to model real-world teacher-student interaction, written feedback was made on a printed copy of the submitted journal entry which was then handed to each participant during the one-on-one meetings. This feedback covered all aspects of the participant's writing, including grammar, word choice, structure and content – i.e. unfocused and extensive feedback. Feedback was focused primarily on aspects of the journals that impacted comprehensibility; however, all grammatical errors related to the target features were corrected.[2] These corrections all took the form of direct CF, with the exception of content-related comments. A sample of the corrections is provided in Figure 1.[3]

FIGURE 1

Correction Sample, from Alex – Journal #2


After that, when I live with my parents, my father had me bow to adults. I bowed to the security men of my apartment, master of every market, the neighbors who met in elevator and more. Thanks to this, I could close to them. One woman who lived in 7 floors bought me cookies when I was young and a grandmother who live in 3 floors sometimes give me bread baked in hands. I have many good neighbors in my hometown.

In the individual feedback sessions, each participant was given time to read over the provided feedback and ask questions. These questions were mostly directed toward word

[1] Sample topics included work/life balance, Korean drinking culture, and military service, among others.

[2] The feedback was unfocused from the learners' perspective; however, only five target features were analyzed for the purpose of the study.

[3] For fear of overwhelming the participants, errors that did not impact comprehensibility or readability were generally left uncorrected.

choice or how to clarify unclear phrases, with some questions pertaining to grammatical corrections. Most questions were answered fully; however, questions about the target features were purposely answered indirectly so as to not provide any information about the grammatical rules involved beyond the corrections on the paper. Overall, these sessions were very similar to what would occur if a student sought extra help from their teacher outside of class to ask about feedback they received, with the exception that in a real world setting a student would not likely seek help on a weekly basis.

Under the guise of pronunciation practice, each participant was asked to read his writing aloud at the end of each session, incorporating the provided feedback, thereby guaranteeing that all of the grammatical feedback provided was processed at least once during the session. After each session, each participant was given one day to edit the journal entry and digitally submit a second draft. Feedback was given during the one-one-one meetings as a hard copy, to require that each participant read and type out the received corrections, thereby allowing them to process the feedback a second time.

4. Instruments and Procedure

1) Grammaticality Judgement Tasks (GJTs)

The GJT contained a series of short sentences containing one of the target features in which the participant was asked to indicate if each sentence was grammatical or ungrammatical. Two GJTs – untimed and timed – were used. Untimed GJT was used to access learners' declarative or explicit knowledge whereas timed GJT was used to access learners' implicit knowledge. Time pressure on a GJT makes accessing explicit knowledge extremely difficult, while an untimed GJT encourages the use of declarative knowledge (Ellis, 2004). The instruments and details of their administration were taken from Ellis (2009), along with some specific test items. Participants were given 4 seconds per sentence on the timed GJTs; on the untimed GJTs they were instructed to underline the problematic areas of sentences they judged to be ungrammatical. There were 50 items in each test.[4]

2) Elicited Oral Imitation Test (EOIT)

This instrument is typically used to tap into learners' implicit knowledge. In this test, each participant heard (only once) a short sentence containing a subjective statement and

[4] Timed GJTs scores have been found to correlate with other tests of implicit knowledge and not with tests of explicit knowledge. Similarly, untimed GJTs scores have been found to correlate with other tests of explicit knowledge and not with tests of implicit knowledge (Ellis, 2009; Han & Ellis, 1998).

was asked to mark on a piece of paper if he agreed or disagreed with the statement then attempt to repeat it using correct grammar. This type of cognitive task between stimulus and reaction is deemed to induce the test-taker to reconstruct the sentence using his/her own grammatical knowledge, relying on the use of one's implicit knowledge (Erlam, 2009). One such stimulus used in the current study was "Korea does not *has* many big buildings" (emphasis added); the participant had to either agree or disagree with the statement before repeating it with correct grammar. The test contained 20 such items: 2 grammatical and 2 ungrammatical for each feature. A response was considered correct if the participant both made an obligatory context for the targeted feature and used said feature in that context correctly.

3) Procedure

The pretests were conducted in Week 1, followed by nine weeks of treatment (i.e. feedback) sessions. The post-tests were conducted in Week 11, and a delayed post-test followed five weeks after the post-test. Table 1 summarizes the procedure of the study.

TABLE 1

Procedure

Time	Session(s)	Instruments used
Week 1	Pretest session	Timed GJT; untimed GJT; EOIT
Week 2 – Week 10	Treatment (9 sessions)	Journal entries and feedback
Week 11	Post-test session	Timed GJT; untimed GJT; EOIT
Week 16	Delayed post-test session	Timed GJT; untimed GJT; EOIT

5. Data Analysis

In order to answer the first research question, accuracy in use of the target features was measured using a formula devised by Pica (1991, as cited in Sheen, 2007), which considers correct usage and incorrect usage in obligatory contexts (including errors of use and errors of absence), and overuse:

$$\frac{\text{number of grammatical morphemes supplied correctly}}{\text{number of obligatory contexts + number of overused forms}}$$

For the second research question, results from the pretest, post-test and delayed post-test were compared. Explicit knowledge was examined through scores from the timed GJTs and EOITs, and implicit knowledge was examined through scores from the untimed GJTs.

To answer the third research question regarding the secondary effects of CF, the journals

were analyzed to look for evidence of avoidance of specific features as a result of receiving CF on those features. The most obvious form of specific avoidance would be a marked decrease in the use of a particular feature after receiving corrections on that feature. This possibility was investigated by comparing usage rates for the features on each journal to error rates on the preceding journal. The second form of specific avoidance considered was a potential unwillingness to make revisions based on the feedback received, and this was investigated by calculating the revision rates for each feature.

Changes in overall quality, quantity and complexity of the writings during the course of the treatment phase were also examined. Holistic quality was measured by three experienced reviewers who were independently given each journal (in a randomized order) and asked to assign a numerical grade. Overall length, average sentence length and structural complexity – operationalized as dependent clauses per T-unit (Wolfe-Quintero, Inagaki, & Kim, 1998) – were calculated using a tool for measuring L2 syntactic complexity developed by Lu (2010). Lexical complexity – operationalized as a lexical TTR (i.e. lexical types/lexical tokens; Wolfe-Quintero *et al.,* 1998) – was calculated using a tool developed by Lu (2012).

IV. RESULTS AND DISCUSSION

1. Improvement in Accuracy of Target Features

The results presented here correspond to the first research question. The accuracy rates and total use of the targeted features across each of the journals are provided in Table 2 (Tom) and Table 3 (Alex).

TABLE 2

Accuracy Rate for Tom, with Total Use in Parentheses

Journal	Article	Plural	Third Person	Past Tense	Negation
1	42% (12)	52% (29)	0% (1)	90% (10)	67% (6)
2	26% (19)	43% (23)	0% (2)	83% (6)	100% (5)
3	60% (5)	76% (17)	33% (6)	0% (4)	90% (10)
4	0% (2)	84% (25)	25% (8)	N/A (0)	100% (7)
5	53% (17)	68% (22)	36% (11)	N/A (0)	71% (14)
6	57% (7)	47% (17)	13% (8)	N/A (0)	80% (5)
7	36% (11)	61% (18)	13% (8)	0% (1)	82% (17)
8	50% (6)	50% (20)	0% (7)	0% (1)	82% (11)
9	43% (14)	57% (7)	20% (5)	25% (16)	80% (10)

TABLE 3

Accuracy Rate for Alex, with Total Use in Parentheses

Journal	Article	Plural	Third Person	Past Tense	Negation
1	43% (14)	54% (13)	N/A (0)	75% (4)	75% (4)
2	50% (2)	90% (10)	N/A (0)	63% (8)	100% (1)
3	N/A (0)	90% (20)	100% (1)	100% (2)	80% (5)
4	100% (2)	100% (6)	N/A (0)	100% (1)	100% (2)
5	33% (3)	50% (10)	67% (3)	50% (2)	75% (4)
6	0% (1)	71% (14)	33% (3)	N/A (0)	100% (3)
7	67% (6)	93% (14)	67% (3)	N/A (0)	80% (5)
8	50% (6)	83% (12)	0% (1)	100% (1)	100% (5)
9	33% (6)	87% (15)	67% (3)	83% (6)	100% (4)

As shown in the tables, accuracy rates fluctuated over the course of the study and did not show any definite trends. Negation was the most stable feature for both participants. A clear change from the first journal to the last was not seen with any feature for either participant.

For each participant, however, brief periods of stable change, lasting over 2-3 journal entries, were found. For Tom, this was an increase in accuracy with Plural, Third Person and Negation observed on journals 3, 4 and 5. During this period, his accuracy for these features rose to a level markedly higher than on the rest of the journals and was stable at that level for 3 weeks before resuming to the original rate. The clearest example of this is from the data for Plural. Tom's average accuracy for Plural was 60% over the 9 journals. On journal #3, however, his accuracy rate increased notably to 76% and remained high before dropping down below the average on journal #6. Nearly all of his errors for Plural were instances of obligatory but absent plural morphology (96%), and of those the majority (76%) were obligatory from sentential context. The types of errors made did not change during the temporary period of improvement.

Alex showed a similar pattern of temporary change in accuracy on journals #5 and #6 for Plural, but in his case it was a notable decrease in accuracy. His accuracy for Plural on average was 80% but it dropped to 50% on journal #5 and at 71% on journal #6. This decrease in accuracy was primarily the result of a sharp increase in overuse on uncountable nouns (3 instances), but by journal #7 this overuse had completely disappeared.

During Tom’s brief period of stable improvement, it appears that he attended to the CF he received and used that to inform his subsequent writing in the short term, but on the subsequent journal entries, his level of engagement fell and with it his accuracy. As Shintani and Ellis (2013) point out, “… explicit L2 knowledge, like other types of declarative knowledge, can be easily forgotten” (p. 291). The results for Plural show this more clearly as it was a particular aspect of Plural use – obligatory contexts based on sentential cues – that produced the increase and following decrease in accuracy.

Alex's drop in accuracy for Plural on journal #5 and quick recovery demonstrates a critique of CF that CF in general, but especially focused CF, will lead to a higher rate of overuse on corrected grammatical structures. While that appears to have in fact happened here, CF received by Alex on his overuse (which was reflected in his revised draft) successfully remedied this overuse in a short time frame.

In addition to considering improvement in accuracy percentages, it is possible to consider the developmental stages of each participant and the types of errors that learners made when using the features to look for evidence of improvement. To that end, the developmental stages outlined in Cazden (1968) were used. In her research of children acquiring nominal and verbal inflections in L1-English, Cazden noted that development of most inflections can be broken down into four categories, which was later confirmed in L1 and L2 development studies (Mervis & Johnson, 1991; Jia, 2003). Jia assigned the following names to these stages: 1) pre-form stage – learner does not use rule; 2) transitional pre-rule – rule is used occasionally in obligatory contexts but with few if any errors or overregularizations when used; 3) transitional post-rule – rule is used more often in obligatory contexts, but with overregularizations; and 4) rule mastery – rule is used with at least 90% accuracy.

In this framework, Tom was in the transitional pre-rule stage for Article, Plural and Past Tense (fewer than 5% of errors involving overregularization), and transitional post-rule for Third Person (18% of errors being overregularizations). Alex's infrequent use of some of the targeted features makes judgements regarding developmental stages difficult. Despite this, the data show that he was in the post-rule transitional stage for Plural (35% of errors being overregularizations). Neither participant progressed from one stage to another during the course of the study.

As an aside, we can note that each feature analyzed responded differently to CF, showing that there are indeed merits to analyzing grammatical features individually. For example, L1-Korean learners typically face additional difficulties with English Article and Plural acquisition given the typological differences between the two languages. Korean is without an article system, and plural morphology is used much less frequently and is rarely obligatory. Hence, the frequency of use and complexity of these systems in English make them especially challenging for L1-Korean learners of English. The ease with which a given grammatical feature is acquired can depend greatly on characteristics of a learner's L1, and individually analyzing the grammatical features enables researchers to account for factors such as this.

In addition to L1-related factors, it was observed that accuracy for Negation, the only targeted feature that carried some communicative value, was notably higher than accuracy for the other targeted features for both participants. The fact that the target features in this study (Negation excepted) are not meaning-laden, i.e. 'redundant' in most contexts makes

them difficult to acquire for any learner of L2 English. According to VanPatten (2004), redundant grammatical forms with no communicative value such as the plural marker (which is often marked by –*s*, and by determiners such as "two" or "many") are processed after the less meaningful forms, or not processed at all. Schmidt's (1990) Noticing Hypothesis extends this idea by suggesting that the acquisition of redundant features is aided by or perhaps requires learner engagement and attention. Then, for either participant to clearly progress from one stage to another for any of the redundant features analyzed, a significant amount of engagement and focus would have been required on their part, especially considering the frequency of treatments and the total length of this study. As such, it is not entirely surprising that no such progression was observed – a point that will be visited again in the next section.

2. Development of Explicit and Implicit Knowledge

This section pertains to the second research question. The results from the untimed GJT, EOIT and timed GJT are provided in Table 4 (Tom) and Table 5 (Alex).

TABLE 4

Explicit / Implicit Knowledge Test Scores for Tom

	Explicit Knowledge			Implicit Knowledge					
	UGJT (maximum=10)			EOIT (maximum=4)			TGJT (maximum=10)		
	Pre	Post	Del.	Pre	Post	Del.	Pre	Post	Del.
Article	4	7	5	3	2	2	4	5	6
Plural	4	6	8	1	2	3	6	7	8
Third Person	9	10	8	1	1	2	9	8	10
Past Tense	9	8	9	1	0	3	9	8	8
Negation	9	8	10	2	4	3	10	7	10

Note. UGJT: Untimed grammaticality judgement task; EOIT: Elicited oral imitation test; TGJT: Timed grammaticality judgement task; Del.: Delayed post-test

TABLE 5

Explicit / Implicit Knowledge Test Scores for Alex

	Explicit Knowledge			Implicit Knowledge					
	UGJT (maximum=10)			EOIT (maximum=4)			TGJT (maximum=10)		
	Pre	Post	Del.	Pre	Post	Del.	Pre	Post	Del.
Article	6	6	9	2	2	3	5	5	7
Plural	7	8	9	2	4	4	6	5	7
Third Person	9	10	10	3	2	1	10	9	10
Past Tense	10	10	10	1	2	1	10	10	10
Negation	9	9	9	4	3	4	10	9	10

Tom's performance for Article on the GJTs fluctuated but overall showed a small gain on the untimed GJTs and a moderate and consistent gain on the timed GJTs. For Plural, Tom had a large overall gain of 4 points on the untimed GJTs from the pretest to the delayed post-test, and a moderate gain of 2 points on the timed GJTs over the same period. His EOIT scores matched this by showing a moderate gain as well.

Alex's test score patterns differed greatly from Tom's. For Article, Alex showed no change from the pre-test session to the immediate post-test session on both GJTs, but an increase of 2-3 points between the immediate post-test and the delayed post-test. For Plural, Alex showed moderate overall gains of 2 points for all three test types but a different pattern for each. His untimed GJT scores improved gradually, his timed GJT scores actually fell on the post-test, and his EOIT scores improved sharply on the immediate post-test and remained high.

Before discussing the results, some observations should be made for the sake of giving context to the results above. Comparatively low performance on the EOIT may be attributed to the participants' unfamiliarity with this type of instrument. Also, both participants excelled at the secondary tasks for the untimed GJTs (locating problematic areas of ungrammatical sentences [100% accuracy] and identifying the grammatical rules involved on those same items [Tom: 86%, Alex: 97%]); hence data provided for the untimed GJTs is limited to the primary task of determining grammaticality. Finally, a ceiling effect for the features of Third Person, Past Tense and Negation for both GJTs precluded an analysis of performance on these features using the results of the GJTs. The EOIT was deemed to not have enough items per feature to draw conclusions based on its results alone.

For both Tom and Alex, performance on Article and Plural showed increases in both explicit and implicit knowledge. Tom's performance on Plural was the clearest of these, increasing gradually and steadily over the three tests. Improvements for the other three given were less clear, involving much delayed improvement or a lack of stability, but were improvements nonetheless. However, recall that for the first research question, none of the features analyzed showed a durable improvement in accuracy over the course of the treatments.

In order to argue against Truscott's (1996) position and say that learners can acquire the target form from written feedback, it is essential to observe their improvements in both explicit and implicit knowledge. Additionally, an improvement in written accuracy would be expected once implicit knowledge has improved, if not earlier with explicit knowledge. The fact that there was no consistency between the results for accuracy on the one hand and explicit and implicit knowledge on the other suggests that the treatment did not achieve the dual goals of improving accuracy and aiding the acquisition of grammar rules.

An important issue that arises concerns why studies on CF such as Bitchener and Knoch (2010) and Van Beuningen *et al.* (2012) found a positive effect for accuracy, which was not the case for the current study. A number of possible explanations can be ruled out. First, it is difficult to argue that the participants did not notice the CF, as the read-aloud task and revision task ensured that each instance of CF was processed by the participants at least twice. Second, it is difficult to argue that the CF given was insufficient for Article and Plural. Errors with these features produced a minimum of 20 corrections for each form over the course of the CF treatment. Third, accuracy for these two features was higher than 50% at the outset of the study, so each participant was at least in the pre-rule stage, and thus 'developmentally ready' to process the forms. Last, both participants were actively engaged in the writing exercises themselves and showed strong interest in improving their ability, as was made clear from their questions during the feedback sessions in particular.

The explanation that all the evidence points to is a lack of learner engagement and attention in the act of acquiring the grammar rules. During the feedback sessions, the participants' questions were focused on issues impacting meaning, such as word choice or unclear diction, but not on grammar. Aside from the CF itself, there were no external factors drawing the participants' attention to grammatical errors, such as pedagogical interventions or focus on form tasks that might take place in a classroom setting. Additionally, both participants stated in the exit interview that they were completely unaware of the purpose of the study, and thus it was not possible for them to have paid more attention to the targeted features in an attempt to please the researcher. There were also no sources of instrumental motivation, such as grades, nor was there any evidence of a desire to avoid receiving CF. Both participants were quite content to make revisions based on the CF they received and then immediately disregard the CF.[5]

In summary, the participants' grammatical accuracy did not improve, most likely because neither student had any specific interest in it, nor was there anything to encourage or induce them to pay more attention to grammatical accuracy than what was necessary for comprehensibility. Without externally-induced attention that orients learners to focus on form and learner engagement to focus on grammar, only incidental learning of the targeted features was possible, which did not take place.

[5] As reported in the exit interview, the learners' intrinsic motivation was directed towards the content and organization of their writing, and they had no particular desire to focus on grammatical accuracy.

3. Secondary Effects of Corrective Feedback

This pertains to the third research question which examined any secondary effects imposed by the provision of CF. The results revealed no evidence of avoidance behavior regarding the target features. The results also showed that the number of corrections received for a particular feature on a given journal had no impact on the number of times that feature was used in the subsequent journal entry, even if the number of corrections received was relatively large. The overall revision rate was also high (Tom: 89%; Alex: 96%) and not impacted by the given feedback. It is important to note that the features analyzed in this study, particularly Article and Plural, are generally difficult to avoid.

No evidence of an overall loss of quality or complexity was found to have taken place either. On the other hand, some of the measures showed an increase in quality over the course of the study. The holistic quality scores, measured to give an indication of how a teacher of L2 English would rate the journals, showed an increase in quality for Tom, but not for Alex.

For Tom, average sentence length (words per sentence) increased from 10.3 on the first journal to 17.9 on the last, and structural complexity (dependent clauses per T-unit) significantly increased from 0.41 to 1.83. Tom's lexical complexity and all of Alex's complexity measures varied slightly but were overall consistent. Data for these measures of secondary effects are provided in Table 6 (Tom) and Table 7 (Alex).

TABLE 6

Quality and Complexity Scores for Tom

Journal	Holistic quality	Total words	Mean sentence length	Structural complexity	Lexical complexity
1	1.8	711	10.3	0.41	0.60
2	2.0	683	15.9	1.02	0.58
3	3.3	588	15.1	1.30	0.60
4	1.8	502	23.9	1.86	0.54
5	2.8	683	18.5	2.39	0.46
6	2.8	543	16.5	0.98	0.57
7	3.1	556	15.0	1.41	0.50
8	4.0	529	18.2	1.06	0.50
9	3.1	482	17.9	1.83	0.50

TABLE 7

Quality and Complexity Scores for Alex

Journal	Holistic quality	Total words	Mean sentence length	Structural complexity	Lexical complexity
1	3.2	413	10.9	0.26	0.67
2	1.6	235	11.2	0.63	0.79
3	3.0	256	9.1	0.20	0.69
4	3.2	219	11.5	0.89	0.62
5	2.6	282	11.8	0.72	0.62
6	3.4	275	12.0	0.91	0.61
7	4.0	283	11.3	0.21	0.66
8	2.1	293	9.8	0.52	0.69
9	3.7	275	11.5	0.60	0.71

As seen in the tables, the total word count of each journal gradually decreased for Tom, from 711 words on the first journal to 482 words on the last compared to Alex whose word count remained consistent throughout the study.[6] Tom's word-count results, which steadily decreased over the course of the study, deserve some attention. Considering that Tom's holistic quality ratings showed an overall increase in quality, it appears that the decrease in his journal length was simply a refining of his writing ability. While on the first journals his writing was unnecessarily verbose, his ability to communicate his thoughts more clearly and using fewer words increased and it was no longer necessary for him to write long journals to adequately respond to the topic questions. Thus, the decrease in length of his writing seems unrelated to the CF he received.

These findings on secondary effects provide valuable context for the discussion of the earlier research questions concerning accuracy and explicit and implicit knowledge, as they make it possible to argue that any improvements in grammatical accuracy observed in this study can be attributed to CF rather than being the result of avoidance strategies employed by the participants.

V. CONCLUSION

The purpose of the current study was to examine whether the provision of direct correction on grammatical errors in L2 writing produces improvement in grammatical accuracy or stimulates the development of explicit or implicit knowledge of target grammatical features. An additional aim was to explore the possibility of corrective feedback having a negative impact on learners' writing, by causing avoidance behavior or a

[6] The high word count for Alex's first journal was due to a misunderstanding of the amount of time he was expected to put into writing each journal, which was addressed at the first feedback session.

general degradation in quality or complexity. The results showed notable but non-durable changes to accuracy and durable changes to explicit and implicit knowledge. However, the lack of consistency between accuracy measures and explicit/implicit knowledge measures for individual grammatical features suggests that there was no treatment effect overall. The participants were observed to be primarily interested only in those aspects of their writing which directly related to their communicative goals. Combined with the redundant nature of the grammatical features targeted in this study, this means that only incidental learning of the targeted features was possible, and this was found to have not taken place. The negative secondary effects of CF claimed by Truscott were also not found, suggesting that the journal entries reflected the participants' true ability without influence from CF-caused simplification or avoidance behaviour.

The results of this study lead to two observations regarding the practice of administering CF. A critique of focused CF is that it can lead to overuse of a targeted feature in subsequent writing. However, the current results show that this CF-caused overuse can also be quickly rectified by providing CF on this overuse, suggesting that this concern is negligible. Second, each of the targeted features responded differently to the CF provided, and this highlights the potential merits of individually analyzing grammatical features in future studies. This observation can explain why only the series of studies of Bitchener and colleagues, having all targeted the same grammatical feature (cf. Bitchener & Knoch, 2010), has thus far produced consistent results.

As a response to Truscott (1996), the findings of this study support his claim that CF on grammatical errors is not effective. However, this needs to be interpreted with caution since the results can be largely attributed to the participants' focus on communicative intent, and their lack of interest in grammatical precision that did not serve their communicative goals. Thus, these results do not rule out the possibility of CF being effective for situations where extrinsic motivation serves to draw learners' attention to these features, or for errors that involve meaning-bearing features or otherwise impede a learner's communicative intent. These results also do not support the underlying theoretical argument made by Truscott regarding the lack of an interface between explicit and implicit knowledge systems in the mind, as parallel improvements were observed for both types of knowledge.

Some implications for L2 instructors can be drawn from the current findings. As the results have shown, CF on redundant grammatical features seems to provide only minor improvements in promoting learners' accuracy and the development of their explicit/ implicit knowledge of the target forms. In order to reap the most benefits from CF, it would more effective to intentionally direct the learners' attention to the targeted grammatical features in light of the finding that the potential for incidental learning via CF is limited when learners are focused on meaning. Additionally, the current findings also

suggested that providing CF has little to no negative effect on writing behaviour and does not result in avoidance behaviour, implying that the fears of such results were exaggerated. The current results actually yielded positive secondary effects of CF for Tom whose overall writing quality improved in the course of nine weeks. For accuracy improvement, however, it would be best to explicitly direct the learners to pay attention to form either by actually telling them to focus on form or accuracy or via focus on form activities which can orient their attention to grammatical accuracy. Explicitly directing the learners to pay more attention to grammar may not seem like much, but it does orient them to pay more attention to redundant and non-meaning bearing forms which could eventually help their accuracy and develop their implicit knowledge of the forms in the long run.

REFERENCES

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102-118.

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

Bitchner, J. & Knock, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System, 37,* 322-329.

Bitchener, J. & Knoch, U. (2010). The contribution of written corrective feedback to language development: a ten-month investigation. *Applied Linguistics, 31*(2), 193-214.

Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were…. *System, 37*(4), 600-613.

Cazden, C. (1968). The acquisition of noun and verb inflections. *Child Development, 39* (2), 433-448.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275.

Ellis, R. (2009). The measurement of implicit and explicit knowledge. In R. Ellis (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 27-64). Tonawanda, NY: Multilingual Matters.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*(3), 353-371.

Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In R. Ellis (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 65-93). Tonawanda, NY: Multilingual Matters.

Ferris , D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition, 32,* 181-201.

Han, Y. & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2*(1), 1-23.

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research, 46*(6), 1297-1311.

Kim, J. (2014). Explicit grammar instruction and development of implicit knowledge. *English Language & Literature Teaching, 20*(1), 93-112.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190-208.

Mervis, C. & Johnson, K. (1991). Acquisition of the plural morpheme: A case study. *Developmental Psychology*, *27*(2), 222.

Pica, T. (1991). Foreign language classrooms: Making them research-ready and researchable. In B. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 393-412). Lexington, MA: D.C. Heath.

Park, S. (2015). The effects of written corrective feedback on Korean EFL leaners' article accuracy. *English Language & Literature Teaching, 21*(2), 65-82.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*(1), 83-95.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255-283.

Shintani, N. & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing, 22*(3), 286-306.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*(2), 327-369.

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A Response to Ferris. *Journal of Second Language Writing, 8*(2), 111-122.

Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in*

English Literature and Linguistics, 27(2), 93-108.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*(4), 255-272.

Truscott, J. (2010a). Some thoughts on Anthony Bruton's critique of the correction debate. *System, 38*(2), 329-335.

Truscott, J. (2010b). Further thoughts on Anthony Bruton's critique of the correction debate. *System, 38*(4), 626-633.

Van Beuningen C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights and future directions. *International Journal of English Studies, 10*(2), 1-27

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

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.) *Processing instruction: Theory, research, and commentary* (pp. 5-31). Mahwah, NJ: Lawrence Erlbaum Associates.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and syntactic complexity*. Honolulu, HI: National Foreign Language Resource Center.

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

Daniel Corks
Department of Childhood English Education
Dongshin University
185 Geonjae-ro, Naju, Jeonnam
South Korea, 58245
Email: dcorks@gmail.com

Eun Sung Park
Department of English Literature and Linguistics
Sogang University
35 Baekbeom-ro, Mapo-gu, Seoul

South Korea, 04107
Tel: 02-705-8303
Email: eunsungp@sogang.ac.kr

Received in January 15th, 2016
Reviewed in February 26th, 2016
Revised version received in March 5th, 2016