Language Learning & Technology
ISSN 1094-3501

June 2019, Volume 23, Issue 2
pp. 65–83

ARTICLE

The effect of two forms of computer-automated metalinguistic corrective feedback

Jianwu Gao, *Capital Normal University*
Shuang Ma, *Capital Normal University*

Abstract

This study investigated whether the effect of two forms of computer-automated metalinguistic corrective feedback in drills transferred to subsequent writing tasks. The English simple past tense, a learned structure, was selected as the target structure. Participants included 117 intermediate learners of English as a foreign language assigned to two feedback groups, one no-feedback group, and one control group. These groups completed writing tasks before the drills, immediately after the drills, and two weeks after the drills. In the drills, the feedback groups completed an untimed error correction test (ECT 1) in which they received either metalinguistic feedback or metalinguistic feedback with corrections. After that, the feedback groups completed another untimed ECT (ECT 2). The no-feedback group completed the two ECTs without receiving any feedback. The results showed that the feedback groups performed better than the no-feedback group on ECT 2. However, no effect for group was found on the learners' improvement from the first to the second writing task and from the first to the third writing task.

Keywords: *Computer-Assisted Language Learning, Writing, Syntax, Grammar, Second Language Acquisition*

Language(s) Learned in This Study: *English*

APA Citation: Gao, J., & Ma, S. (2019). The effect of two forms of computer-automated metalinguistic corrective feedback. *Language Learning & Technology*, *23*(2), 65–83. https://doi.org/10125/44683

Introduction

The efficacy of metalinguistic corrective feedback (CF) has drawn growing interest in the field of second language acquisition. Researchers have studied written metalinguistic CF (e.g., Shintani, Aubrey, & Donnellan, 2016), computer-automated metalinguistic CF (e.g., Zhao & MacWhinney, 2018), oral metalinguistic CF (e.g., Rassaei, 2015), and metalinguistic CF in computer-mediated communication (CMC; e.g., Monteiro, 2014).[1] However, most studies have not examined the efficacy of metalinguistic CF as independent input. Rather, they have introduced it as an addition to direct written CF (e.g., Diab, 2016) or oral recasts (e.g., Rassaei, 2015). The findings of these studies, therefore, cannot speak to the efficacy of metalinguistic CF alone. The few studies that have examined metalinguistic CF as independent input are largely divergent in the types of target structures and research designs. Their findings are far from conclusive. The current study examined whether the effect of computer-automated metalinguistic CF, by itself or with correction, transferred from drills to subsequent writing tasks. The English simple past tense—a structure that participants had been taught before—was selected as the target structure.

We aimed to further the understanding of metalinguistic CF—particularly computer-automated metalinguistic CF—in four ways. First, although previous studies found that computer-automated metalinguistic CF facilitated the learning of new structures (e.g., Presson, MacWhinney, & Tokowicz, 2014), its impact on the correct use of previously-learned structures remains understudied. The only computer-automated metalinguistic CF study that selected a learned structure (Heift, 2010) provided insufficient evidence, because it did not include any control or comparison group. On the other hand, metalinguistic CF studies outside the field of computer-assisted language learning (CALL) selected learned structures and found metalinguistic CF alone to be either ineffective or unable to produce a durable effect (e.g., Shintani et al., 2016). By studying whether computer-automated metalinguistic CF facilitates the correct use of a learned structure, we compared our findings with those of previous CALL studies that analyzed new target structures and of non-CALL studies that analyzed learned structures.

Second, our study examined whether the effect of CF in drills transferred to subsequent written production. In drills, the primary focus of CF is the linguistic form. In communicative tasks, the focus is meaningful communication, with attention to form occurring only incidentally, if at all (Li, 2010). According to Li (2010), different task types in which CF is given can affect learning outcomes, with drills generating a larger effect size than communicative tasks. Most previous investigations into the effects of metalinguistic CF on the use of learned structures adopted communicative tasks, and their results tended to negate, rather than affirm, the efficacy of this type of CF. The only two studies that adopted drills reported conflicting findings. Our study offered further insight on the role of metalinguistic CF on a learned structure when drills are adopted.

Third, previous metalinguistic CF studies that incorporated a control or comparison group in their designs predominantly used either comparison groups that participated in practice[2] without receiving form-focused feedback (e.g., Shintani, et al., 2016) or control groups that did not participate in any practice (e.g., Shintani, Ellis, & Suzuki, 2014), but not both (for an exception, see Wong, Zhao, & MacWhinney, 2018). The divergence in group design gave rise to concerns of comparability across studies (e.g., Monteiro, 2014). It would therefore be helpful to incorporate both a no-feedback comparison group and a control group and compare their performance with that of the feedback groups.

Last, our study adopted one type of free constructed responses, written production tasks, to measure the effect of computer-automated metalinguistic CF in drills. Free constructed responses have been extensively used in written metalinguistic CF studies, but rarely with computer-automated CF. By selecting this type of elicitation procedure, we can compare our findings with those of previous written metalinguistic CF studies.

Literature Review

Metalinguistic CF is defined and operationalized differently depending on the medium in which it is given. Oral metalinguistic CF is immediate and may take the form of either comments or questions generally indicating the existence of an error (e.g., *Can you find your error?* or *No*). It can also be slightly more specific information indicating the nature of the error (e.g., *It's masculine*; see Lyster & Ranta, 1997, p. 47). Written metalinguistic CF, on the other hand, is delayed and may be operationalized at various degrees of specificity, ranging from at-length explanations of linguistic rules (e.g., Bitchener & Knoch, 2010) to codes indicating the nature of errors (e.g., Lalande, 1982). Metalinguistic CF in CMC contexts is operationalized, like its oral counterpart, as immediate metalinguistic comments on the nature of the error (e.g., *Be sure to use past tense*; see Monteiro, 2014). However, it differs from oral metalinguistic CF, in that it may involve multiple modalities such as text-chat and pictures, thus allowing more processing and planning time (Sauro, 2009). Compared with the other three, computer-automated metalinguistic CF is perhaps more intensive, for it is both immediate and often operationalized as a detailed grammatical explanation with examples (for an example, see Sanz & Morgan-Short, 2004, p. 56).

Despite the divergent operationalizations of metalinguistic CF across different media, one feature remains unvaried: correct forms are withheld to push learners to self-correct. This form of metalinguistic CF is believed to (a) promote learners' noticing of target forms, thereby setting it apart from meaning-focused feedback; (b) provide declarative knowledge of target forms, thereby setting it apart from implicit CF; and (c) encourage learners to "proceduralize" the explicit knowledge themselves (Shintani & Ellis, 2013, p. 288), thereby setting it apart from direct correction.

Despite theorists' optimistic predictions (e.g., Ellis, 1994; Schmidt, 1990), investigations into the efficacy of metalinguistic CF remain inconclusive, as shown by the summary of studies in Table 1.

Table 1. *Studies of Metalinguistic CF as Independent Input*

Study	Medium of Metalinguistic CF	Research Design					Effect of Metalinguistic CF[c]
		Target Structure	Task Type	Research Setting	Control Comparison Group[a]	Data Elicitation Procedure[b]	
Lalande (1982)	Written	Learned	Communicative	Classroom	0	FC	0
Robb, Ross, and Shortreed (1986)	Written	Learned	Communicative	Classroom	C2	FC	0
Carroll and Swain (1993)	Oral	New	Drill	Lab	C2	CC	2
Nagata (1993)	Computer-automated	New	Drill	Lab	0	CC	2
Kim and Mathes (2001)	Oral	Learned	Drill	Classroom	0	CC	0
Chandler (2003)	Written	Learned	Communicative	Classroom	C1	FC	0
Sanz (2003)	Computer-automated	New	Drill	Lab	C2	MC+CC+FC	1
Sanz and Morgan-Short (2004)	Computer-automated	New	Drill	Lab	C2	MC+CC+FC	1
Ellis, Loewen, and Erlam (2006)	Oral	Learned	Communicative	Classroom	C1	GJT+CC	2
Loewen and Erlam (2006)	CMC	New	Communicative	Classroom	C1	GJT	0
Ammar (2008)	Oral	Learned	Communicative	Classroom	C2	MC+FC	2
Sauro (2009)	CMC	New	Communicative	Classroom	C2	MC	0
Heift (2010)	Computer-automated	Learned	Drill	Lab	0	CC	1
Storch and Wigglesworth (2010)	Written	Learned	Communicative	Classroom	0	FC	2
Lyddon (2011)	Computer-automated	New	Communicative	Lab	C2	CC	1
Goo (2012)	Oral	New	Drill	Classroom	C1	GJT+CC	?
Stafford, Bowden, and Sanz (2012)	Computer-automated	New	Drill	Lab	C2	MC+GJT	2

Shintani and Ellis (2013)	Written	Learned	Communicative	Classroom	C2	GJT+FC	0
Lado, Bowden, Stafford, and Sanz (2014)	Computer-automated	New	Drill	Lab	C2	MC+GJT+RT	2
Monteiro (2014)	CMC	Learned	Communicative and Drill	Classroom	C2	GJT+CC	1
Presson et al. (2014)	Computer-automated	New	Drill	Lab	C2	CC+RT	2
Shintani et al. (2014)	Written	Learned	Communicative	Classroom	C1	FC	0
Diab (2015)	Written	Learned	Communicative	Classroom	0	FC	1
Shintani and Ellis (2015)	Written	Learned	Communicative	Classroom	0	FC	0
Shintani et al. (2016)	Written	Learned	Communicative	Classroom	C2	FC	0
Zhao and MacWhinney (2018)	Computer-automated	New	Drill	Lab	C1	CC+RT	1
Wong et al. (2018)	Computer-automated	New	Drill	Lab	C1+2	CC+RT	2

[a] *0 = no control or comparison group, C1 = control, C2 = comparison group; the distinction followed Norris and Ortega (2000).*

[b] *MC = multiple choice, GJT = grammaticality judgment test, CC = constrained constructed responses, FC = free constructed responses, RT = reaction time; the classification followed Norris and Ortega (2000).*

[c] *0 = no effect or no durable effect, 1 = durable effect, 2 = more effective than other forms of CF, ? = no proof of durability of effect*

The metalinguistic CF studies included in Table 1 differed largely in four aspects of research design: the participants' familiarity with the target structure, task type, the presence of a control or comparison group, and elicitation procedures. The following sections are devoted to a review of their findings in line with these four factors.

Learned Versus New Target Structures

Of the 27 studies that investigated the effect of metalinguistic CF as independent input, 13 selected new target structures, and 14 selected learned structures. The 13 studies that selected new target structures reported divergent findings depending on the medium of the metalinguistic CF under examination. The most positive results came from the computer-automated and oral metalinguistic CF studies. All 10 computer-automated metalinguistic CF studies found this CF form effective, and five of the studies (Lado et al., 2014; Nagata, 1993; Presson et al., 2014; Stafford et al., 2012; Wong et al., 2018) reported metalinguistic CF to be more effective than indirect CF. The two oral CF studies reported similarly positive results. Carroll and Swain (1993) found metalinguistic explanation significantly more effective than indirect CF, recast, and indication of errors. All the feedback groups in their study performed better than the comparison group. Goo (2012) found that the metalinguistic CF group performed better than the control in the immediate post-test. The two CMC studies (Loewen & Erlam, 2006; Sauro, 2009), on the other hand, did not find efficacy for metalinguistic CF. Also notable is that none of the written metalinguistic CF studies investigated the learning of new structures.

Of the 14 studies that selected learned target structures, 12 examined written and oral metalinguistic CF. Of the nine written metalinguistic CF studies, seven found metalinguistic CF ineffective. The remaining two studies provided evidence in favor of the efficacy of metalinguistic CF, but that evidence is less definitive upon closer examination. Diab (2015) did not include a comparison or control group and Storch and Wigglesworth (2010) did not adopt new writing tasks to measure their outcomes. The three oral metalinguistic CF studies reported mixed results. Ammar (2008) and Ellis et al. (2006) found metalinguistic CF more effective than other CF forms. Kim and Mathes (2001), however, found metalinguistic CF ineffective. Meanwhile, there was only one CMC study (Monteiro, 2014) and one computer-automated metalinguistic CF study (Heift, 2010) that adopted learned structures, and both found metalinguistic CF effective. However, it is important to note that Heift (2010) did not incorporate a control or comparison group design, and Monteiro (2014) found no difference in performance between the feedback groups (metalinguistic CF and recast) and the comparison group.

Task Types

The divergent results may be easily attributable to another factor: the adoption of different types of tasks. Of the 15 studies which implemented communicative tasks, nine of them found metalinguistic CF ineffective. On the other hand, the 13 studies[3] that adopted drills largely affirmed the efficacy of metalinguistic CF. An initial explanation can be that different task types are associated with different research settings. Of the 15 studies that adopted communicative tasks, 14 were conducted in the classroom, whereas 10 out of the 13 studies that adopted drills were conducted in the lab. As Li (2010) points out, feedback is more salient and noticeable in the lab than in the classroom because there are fewer distractions.

However, the choice of task type may also be related to learners' familiarity with the target structure, possibly affecting the research results. Metalinguistic CF may be given as part of instruction for the learning of new structures or as reinforcement of previously learned structures. As reinforcement, it may be given in both drills and communicative tasks. As part of instruction for the learning of new structures, however, metalinguistic CF by itself has an important limitation: Without prior knowledge of the target structure, it is impossible for learners to independently produce the correct forms when prompted (Lyster, 2004). This might be the reason why that all the studies incorporating new structures adopted drill activities that exposed the participants to the correct forms (e.g., the binary choices in Lado et al., 2014). Such exposure, however, might have had a confounding effect on the findings about metalinguistic CF itself as part of instruction.

When we narrow our examination to the studies into metalinguistic CF as reinforcement, we find that for both drills and communicative tasks, there has been an almost equal amount of evidence for and against the efficacy of the CF. Of the 12 studies that adopted communicative tasks, seven reported negative results. Of the three studies that adopted drills, one (Kim & Mathes, 2001) found metalinguistic CF ineffective, whereas two (Heift, 2010; Monteiro, 2014) found metalinguistic CF effective.

Control Versus Comparison Groups

In addition to the selection of target structures and task type, metalinguistic CF studies differ in their incorporation of control and comparison groups. The typology of control groups adopted for our review generally followed those by Norris and Ortega (2000), but we classified the operationalizations of the comparison groups into three further categories, as follows:

- The group participates in practice and receives no feedback (e.g., Monteiro, 2014);
- The group participates in practice and receives meaning-focused feedback only (e.g., Lyddon, 2011); or
- The group participates in practice, and receives *general form-focused CF* (i.e., feedback that draws learners' attention to forms in general without indicating the target forms), such as indicating the number of errors next to a line (e.g., Robb et al., 1986) or indicating the existence of errors (Wong et al., 2018).[4]

Using the above categorization, we found that seven of the 27 metalinguistic CF studies did not incorporate control or comparison-group designs (e.g., Shintani et al., 2014). These studies therefore could only demonstrate the efficacy of metalinguistic CF compared to that of other forms of CF. Of these studies, one special case was Diab (2015). Although Diab claimed that what she called the control condition was not exposed to any form-focused CF from the teacher, the learners were explicitly asked to correct their own erroneous use of two target structures. Therefore, this group should have been classified as an experimental condition with brief metalinguistic clues rather than as a comparison group. Similarly, Shintani and Ellis (2015) mentioned that they incorporated a "control group" (p. 113), but neither the operationalization nor the results from that group were discussed.

Results of the studies that incorporated control- or comparison-group designs were again mixed. Of the seven studies with control groups, three (Chandler, 2003; Loewen & Erlam, 2006; Shintani et al., 2014) reported no progress for any group, while three (Ellis et al., 2006; Wong et al., 2018; Zhao & MacWhinney, 2018) reported that the treatment groups outperformed the control group. The remaining study (Goo, 2012) did not incorporate any delayed post-test. Of the 14 studies with comparison groups, nine reported that the comparison groups improved in performance, four observed no progress in the performance of either the feedback groups or comparison groups, and eight studies reported that the metalinguistic feedback groups outperformed the comparison groups at least in some measures.

The above summary points out the possible discrepancy between the performance of control groups and that of comparison groups. To our knowledge, the only study into metalinguistic CF that incorporated both types of groups in the research design was the one by Wong et al. (2018). However, that study did not compare the performance of the two conditions. Therefore, a study that implements both a control group and a comparison group is needed to isolate the effect of metalinguistic CF as independent input.

Elicitation Procedures

A final possible source of divergence in the results of the metalinguistic CF studies was the adoption of different types of elicitation procedures. Constrained constructed responses and free constructed responses were adopted most frequently. However, the studies that adopted these two procedures reported quite different findings.[5] Of the 13 studies that adopted constrained constructed responses, as many as 12 confirmed the effectiveness of metalinguistic CF. Of the 12 studies that adopted free constructed responses, seven did not find efficacy for metalinguistic CF.

The findings of the 12 studies that adopted free constructed responses were, again, largely divergent depending on the medium of the metalinguistic CF under examination. Seven of the nine written CF studies found metalinguistic CF ineffective, whereas the only oral CF study (Ammar, 2008) affirmed the efficacy of the same CF form. The positive results reported by the two computer-automated metalinguistic CF studies, on the other hand, should be interpreted with caution. Sanz (2003) and Sanz and Morgan-Short (2004) operationalized the practice as binary choices. Consequently, in these two studies, the practice exposed the comparison groups to the correct forms.

The above literature review reveals four areas that deserve further research. First, only one computer-automated metalinguistic CF study (Heift, 2010) adopted a learned target structure, a subject amply explored in non-CALL metalinguistic CF studies. Second, relatively few studies that adopted learned target structures employed drills, and that research reported inconclusive findings. Third, only one metalinguistic CF study incorporated both a control group and a comparison group (Wong et al., 2018), although the two conditions were shown to diverge in performance (Lyster, 2004). Last, few computer-automated CF studies adopted written production to measure the outcome, and those that adopted this form of elicitation either lent inadequate support for the efficacy of the CF form or were susceptible to ambiguity in research design. The current study, therefore, delved into these areas by (a) investigating whether two forms of computer-automated metalinguistic CF in drills (by itself and combined with direct correction) facilitated the correct use of a learned structure, (b) incorporating both a control group and a comparison group, and (c) adopting written production tasks to measure outcomes.

Research Questions

1. Does the effect of computer-automated metalinguistic CF, by itself and with direct correction, transfer from drills to subsequent writing tasks?
2. Does the participation in form-focused drills affect the accuracy of using the English simple past tense in subsequent writing tasks?

Method

Participants

The participants in this study were 117 first-year undergraduate learners of English as a foreign language (EFL) aged 18–19 from a Chinese university, including 85 students from the Department of English Language and Literature and 32 from the Department of English Education. The curricula for the first-year undergraduates of the two departments were identical. All the participants were native speakers of Mandarin Chinese who had received at least six years of formal EFL instruction. The participants were of intermediate proficiency, with extensive explicit knowledge of English grammar but little experience in English writing.

Because the participants were from two departments that had different class hours, we assigned the 85 participants from the Department of English Language and Literature randomly to Groups 1 through 3 and assigned all the 32 participants from the Department of English Education to Group 4 (the control). Group 1 (n = 28) completed an untimed error correction test (ECT 1) with computer-automated metalinguistic CF and then completed a second ECT (ECT 2). Group 2 (n = 28) differed from Group 1 only in receiving computer-automated metalinguistic CF with direct correction in ECT 1. Group 3 (n = 29) participated in the two ECTs without receiving any feedback. Group 4 (n = 32) did not participate in either ECT.

Target Structure

In this study, we selected the English simple past tense as the target structure. Our choice was based on three considerations. First, although recent studies found that English simple past tense is among the first linguistic forms acquired by EFL learners (Pienemann & Lenzing, 2015), this form remains challenging for the population of this study: intermediate Chinese EFL learners. From our discussions with the faculty members who were teaching the participants, we found that the participants had been explicitly taught the simple past tense in high school, but they kept making errors in this structure in their university course assignments. The choice of simple past tense was therefore especially pertinent to the examination of the effect of metalinguistic CF as independent input, requiring the participants to provide correct forms by themselves upon receiving metalinguistic CF alone. Second, the simple past tense was relatively easy to elicit from participants in writing tasks focusing on meaning. The adoption of this structure could forestall the problem of error avoidance, for which many written CF studies have been rightly criticized (Truscott, 1996). Third, the simple past tense had received substantial attention in oral and written metalinguistic CF studies and in studies in CMC contexts, but to our knowledge, it had not been investigated by any studies focusing on computer-automated metalinguistic CF.

Design

The study was conducted during an instructional term at a Chinese university. A stand-alone program was installed on the server of all the lab PCs to automatically administer the drills and CF. We adopted a one-shot treatment design that was adopted by most (six out of nine) of the computer-automated CF studies reviewed above.[6] Four groups (metalinguistic CF, metalinguistic CF plus correction, no feedback comparison, and control) completed three writing tasks, and three of the groups (metalinguistic CF, metalinguistic CF plus correction, and no feedback comparison) completed two ECTs (see Table 2). In Session 1, all the groups completed the first writing task (WT 1), and the feedback groups and the comparison group participated in ECT 1 (untimed). After the completion of ECT 1, the feedback groups immediately received their respective feedback on ECT 1, and were required to read the feedback quietly for 12 minutes. In the meantime, the no-feedback (i.e., comparison) group was directed to an external website to play an online game. When the 12 minutes were over, the feedback groups and the no-feedback group were asked to complete ECT 2 (untimed) and then finish the second writing task (WT 2).The control group completed WT 2 without participating in the two ECTs. In Session 2 (two weeks after Session 1), all the groups completed the third writing task (WT 3).

Table 2. *The Design of the Study*

<table>
<tr><th></th><th>Time (min)</th><th>Metalinguistic CF</th><th>Metalinguistic CF Plus Correction</th><th>No-Feedback Comparison</th><th>Control</th></tr>
<tr><td rowspan="5">Session 1</td><td>20</td><td colspan="4">WT 1</td></tr>
<tr><td>Untimed</td><td colspan="3">ECT 1</td><td>N/A</td></tr>
<tr><td>12</td><td>Metalinguistic CF on ECT 1</td><td>Metalinguistic CF plus Correction on ECT 1</td><td>Online Game</td><td>N/A</td></tr>
<tr><td>Untimed</td><td colspan="3">ECT 2</td><td>N/A</td></tr>
<tr><td>20</td><td colspan="4">WT 2</td></tr>
<tr><td>Session 2</td><td>20</td><td colspan="4">WT 3</td></tr>
</table>

Treatment

The Drills

At the beginning of the drills, participants were automatically administered an untimed ECT (ECT 1). Participants' successful completion of ECT 1 depended on their mastering of two rules governing the use of simple past tense (see Figure 2 and Figure 3). By consulting the textbooks and the faculty members, we made sure that the words contained in all the items of ECT 1 were likely known to the participants. The test was reviewed and modified by three native speakers from the UK and 10 experienced EFL teachers from China. ECT 1 included two types of sentences of similar length, namely, 12 target sentences containing incorrect use of simple past tense and 12 distracters involving an incorrect use of pronoun reference, a new rule to the participants. No feedback would be provided for errors made in distracters. Special care was taken to ensure that the sentences within each type were presented in a random sequence for each participant, thereby reducing the potential for order effects. In addition, we made sure that each target sentence was followed by a distracter. The above arrangements enabled us to encourage a general focus on form without alerting the comparison group of the target structure. Figure 1 illustrates how a test item was presented on the computer screen.

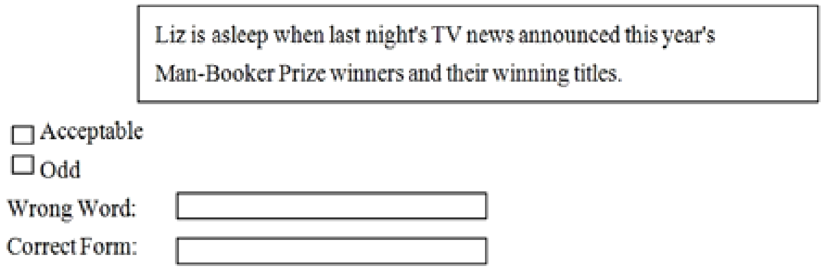


Figure 1. Sample item from ECT 1

As shown in Figure 1, participants were instructed to (a) judge whether the sentence was grammatically acceptable or odd, (b) double-click the erroneous word to make it appear in the "Wrong Word" slot, and (c) type the correct form in the "Correct Form" slot. Each sentence could be rectified by altering one erroneous word. Compared with tests that only required a grammaticality judgment, this form of error correction test precluded the possibility of participants making the right judgment for the wrong reasons.[7]

Computer-Automated Metalinguistic CF

Computer-automated metalinguistic CF was presented to participants automatically upon their completion of all the items of ECT 1. The participants were not presented with the distracters again during the 12-minute CF treatment. The administration of the CF was based on the automatic scoring for ECT 1, which in turn was based on the comparison between the subjects' input and a list of forms that were pre-written into the program. The CF was given to participants, if they (a) wrongly judged any target sentence as acceptable, (b) failed to identify any targeted wrong word, or (c) identified the targeted wrong word but did not make an apparent attempt to change the erroneous tense into simple past. By *apparent attempt*, we mean any ostensible attempt to add the past tense suffix -*ed* to the base form of the verb in question. Any other attempts, including the ambiguous cases, were calculated as erroneous and scored as zero. For example, participants would receive CF if they corrected *lags* as *lagging* but would not if they corrected it as *laged*. The target sentence with the participant's wrong response was displayed on the top of the screen, followed by a detailed explanation of the rule applicable to the sentence in question, one example of correct use of the rule, and one typical learner error. The CF was individualized in that it was provided only if a learner's response to the target sentence was wrong and in that only the metalinguistic explanation pertaining to the target sentence in question was provided. Figure 2 and Figure 3 illustrate the metalinguistic CF on erroneous responses to two target sentences. (Note the match between the target sentences and the rules explained with the examples.)

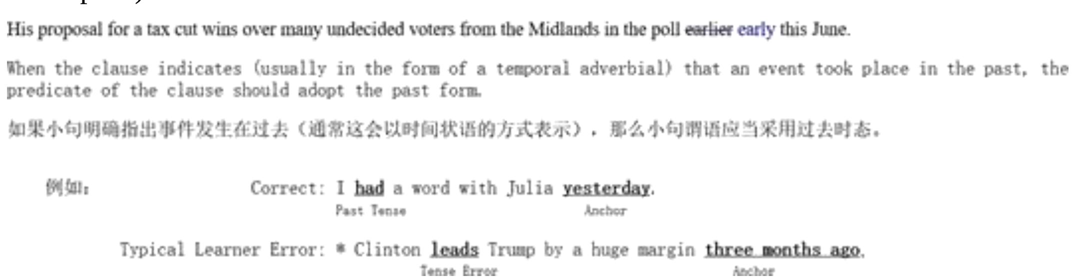


Figure 2. Computer-automated metalinguistic CF on the use of simple past tense when only one clause is involved

The ~~elderly~~ older lady in the picture is our neighbour from the Clarendon city centre before we moved away three years ago.

For a sentence with two clauses (Clause A and Clause B), the predicate of Clause B should adopt the past form if:

1) The predicate of Clause A indicates that an event took place in the past, and

2) Clause B introduces an event that took place no later than the event in Clause A.

对包含两个小句（设为小句A和小句B）的复杂句而言，如果小句A明确指出事件发生在过去，而且小句B的事件发生时间不晚于小句A所指的事件发生时间，则小句B的谓语应当采用过去时态。

例如： Correct: I didn't (Event A) know back then (Anchor) that you two were (Event B) in Paris too.

Typical Learner Error: * Yesterday (Event A) he got (Anchor) to the station as soon as his father's train arrives (Event B).

Figure 3. Computer-automated metalinguistic CF on the use of simple past tense when two clauses are involved

Computer-Automated Metalinguistic Feedback with Direct Correction

The operationalization of computer-automated metalinguistic feedback with direct correction was the same as that of the computer-automated metalinguistic CF, except that the correct form was also provided. Figure 4 illustrates this form of CF on the erroneous judgment of one target sentence. (Note the correct form was provided beneath the participant's incorrect response.)

The legendary Charles Foster Kane goes ~~bankrupt~~ bankrupted during the Great Depression because of his purchases all over the States.

The legendary Charles Foster Kane ~~goes~~ went bankrupt during the Great Depression because of his purchases all over the States.

When the clause indicates (usually in the form of a temporal adverbial) that an event took place in the past, the predicate of the clause should adopt the past form.

如果小句明确指出事件发生在过去（通常这会以时间状语的方式表示），那么小句谓语应当采用过去时态。

例如： Correct: I had (Past Tense) a word with Julia yesterday (Anchor).

Typical Learner Error: * Clinton leads (Tense Error) Trump by a huge margin three months ago (Anchor).

Figure 4. Computer-automated metalinguistic feedback with direct correction

No Feedback

In the operationalization of this condition, participants were required to complete the two ECTs without receiving any CF. At the end of ECT 1, the participants were directed by the program to an external website to play an online game. Figure 5 shows the online game with its instructions.

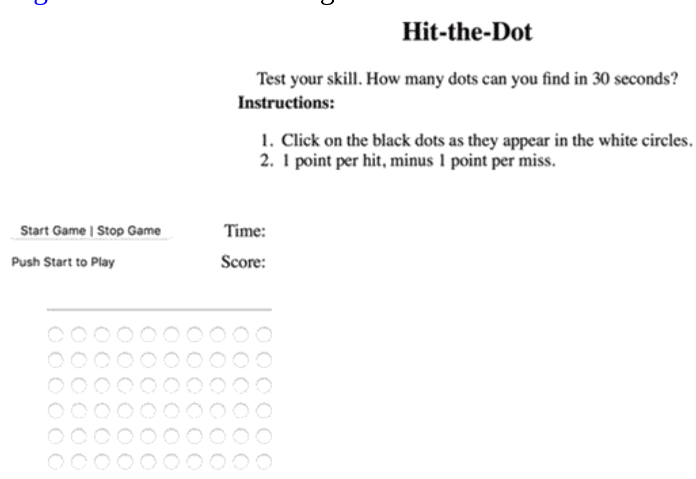


Figure 5. Game for no-feedback group (from Hit-the-Dot, Copyright © Birk, 2000; reproduced with permission; all rights reserved)

Control

The control group did not participate in the ECTs.

Instruments and Measures

To test whether the effect of the two CF forms transferred to subsequent written production, we selected the accuracy rates of the two ECTs[8] and of the three writing tasks as our outcome measures. ECT 2 consisted of 24 new sentences which were designed and reviewed in the same manner as those in ECT 1. The same scoring mechanism was applied to both ECTs (see Table 3).

Table 3. *Scoring for ECT 1 and ECT 2*

Example of Participant Input	**Accuracy Rate**
The UK ~~lags~~ lagged behind the US in childcare support for women until the British government passed a law in the 60s.	1/1
The UK ~~lags~~ laged behind the US in childcare support for women until the British government passed a law in the 60s.	1/1
The UK lags behind the US in childcare support for women until the British government ~~passed~~ past a law in the 60s.	0/1

As is shown in Table 3, full scores were awarded as long as the participants judged a target sentence as *odd* and made an apparent attempt to change the erroneous tense into simple past (e.g., full scores were awarded even if participants misspelled the simple past form of *lagged* as *laged*). Two researchers reviewed the scores automatically generated in the ECTs, and both agreed that the scores were 100% accurate. Because the internal consistency (Cronbach's alpha) for ECT 1 was below .70, three test items were deleted, and the resulting reliability was .71. The internal consistency (Cronbach's alpha) for ECT 2 was .75.

All the writing tasks were modeled on Task 1 of the IELTS academic writing test. Each task required the participants to summarize the information in a chart by writing 150 to 200 words within 20 minutes. To provide an obligatory context for the use of simple past tense, we indicated that the data described situations in the past by specifying past years in the instructions and the titles of the charts (see Appendix). Since participants' overuse of the target structure in the writing tasks was minimal, we followed Bitchener and Knoch (2010) in the calculation of accuracy scores. We used the percentage of correct usage for all occasions where the grammatical structure of the sentence written by the participant required it. The scoring of the writing tasks was consistent with those of the ECTs (Table 4).

Table 4. *Scoring for the Writing Tasks*

Example of Participant Input	**Accuracy Rate**
In late 1980s and after 1995, it sharply *rised*.	1/1
The American population *remains* stable from 1950 to 1970.	0/1

As shown in Table 4, usage was rated as *correct* as long as an apparent attempt was made to use the simple past tense in an obligatory context (e.g., the misspelled form *rised* in the first example was not counted as an error; the missing article in *late 1980s* was ignored as well). The accuracy of all the 351 texts was calculated by two researchers, and the inter-rater reliability was 100%.

The accuracy rates on the two ECTs were subjected to a Kruskal–Wallis test, because the assumption of normality and homogeneity was not met. For the analysis of the results of the writing tasks, we first performed a series of non-parametric tests on the obligatory uses (OUs) of the simple past tense to preclude the problem of error avoidance. We then adopted a multivariate analysis of covariance (MANCOVA) with the accuracy scores on WT 1 as the covariate and introduced three dependent variables as measures of the within-subject improvement (i.e., the score improvement from WT 1 to WT 2, from WT 2 to WT 3, and from WT 1 to WT 3).

Results

ECTs

Table 5 shows the descriptive statistics for the accuracy rates on the two ECTs. Because the control group did not participate in the ECTs, only the results of the drill groups are reported.

Table 5. *ECT Results*

		ECT 1		**ECT 2**	
Group	***N***	***M***	***SD***	***M***	***SD***
Metalinguistic CF	28	0.60	0.25	0.77	0.19
Metalinguistic CF plus correction	28	0.63	0.26	0.76	0.19
No feedback	29	0.59	0.30	0.58	0.26

Because the assumption of normality and homogeneity was not met, the data on the two ECTs were subjected to a Kruskal–Wallis test. The test showed that the three groups did not differ in their scores on ECT 1 ($H_{(2)} = .20$, $p > .05$), but they did differ in their scores on ECT 2 ($H_{(2)} = 9.74$, $p = .008$). Post hoc pairwise comparisons using the Mann–Whitney method showed that the difference lay between the no-feedback group and the two feedback groups, with the two feedback groups scoring significantly higher than the no-feedback group ($p = .006$ for the metalinguistic CF group; $p = .009$ for the metalinguistic CF plus correction group). There was no significant difference between the two feedback groups ($p > .05$). To remove the effect of the ECT 1 scores, we also followed one reviewer's suggestion by conducting a Quade's RANCOVA test on the ECT 2 scores with the ECT 1 scores as the covariate. The result showed that there were significant between-group differences ($F_{(2)} = 4.83$, $p = .010$). Post hoc comparisons again showed that the two feedback groups outscored the no-feedback group ($p = .007$ for the metalinguistic CF group; $p = .011$ for the metalinguistic CF plus correction group). The two feedback groups did not differ in their scores ($p > .05$).

Writing Tasks

Table 6 shows the descriptive statistics on the group-specific OU count of the use of simple past tense in the three writing tasks.

Table 6. *Group-Specific OU Count of Simple Past Tense Use in the Writing Tasks*

	WT 1		**WT 2**		**WT 3**	
Group	***M***	***SD***	***M***	***SD***	***M***	***SD***
Metalinguistic CF	13.18	4.35	11.61	3.47	12.07	4.67
Metalinguistic CF plus correction	15.36	3.92	13.39	4.50	12.61	3.66
No feedback	14.00	5.18	12.93	3.89	12.59	4.05
Control	13.76	4.38	12.97	4.15	12.01	3.92

A Kruskal–Wallis test showed that the four groups did not differ from each other in the OU count across the three writing tasks ($p > .05$). One-sample Kolmogorov–Smirnov tests also showed that within each group, there was no difference in the OU count across the three writing tasks ($p > .05$). The problem of error avoidance was therefore excluded.

Table 7 shows the group-specific descriptive statistics on the accuracy rates on the writing tasks.

Table 7. *Group-Specific Accuracy Rates on the Writing Tasks*

	WT 1		**WT 2**		**WT 3**	
Group	***M***	***SD***	***M***	***SD***	***M***	***SD***
Metalinguistic CF	0.56	0.31	0.58	0.34	0.68	0.35
Metalinguistic CF plus correction	0.42	0.39	0.48	0.29	0.60	0.27
No feedback	0.42	0.36	0.59	0.36	0.62	0.35
Control	0.61	0.38	0.64	0.36	0.58	0.33

An ANOVA showed that, on the whole, the four groups did not differ significantly in WT 1, ($F_{(3, 113)} = 2.19$, $p > .05$). However, post hoc comparisons showed that the control group scored significantly higher than the no feedback group ($p = .044$) and the metalinguistic CF plus correction group ($p = .044$). Therefore, we adopted a MANCOVA with the WT 1 accuracy scores as the covariate and introduced three dependent variables as measures of the within-subject improvement: the scores from WT 1 to WT 2, from WT 2 to WT 3, and from WT 1 to WT 3.[9] The descriptive statistics on the improvement in accuracy scores across the three writing tasks for the four groups are shown in Table 8.

Table 8. *Improvement in Accuracy Rates Across the Three Writing Tasks*

	WT 1 to WT 2		**WT 2 to WT 3**		**WT 1 to WT 3**	
Group	***M***	***SD***	***M***	***SD***	***M***	***SD***
Metalinguistic CF	0.01	0.39	0.11	0.36	0.12	0.41
Metalinguistic CF plus correction	0.06	0.37	0.12	0.28	0.18	0.38
No feedback	0.17	0.39	0.03	0.32	0.20	0.33
Control	0.03	0.33	-0.06	0.29	-0.03	0.37
Total	0.07	0.37	0.05	0.32	0.12	0.38

The MANCOVA showed that there was a significant overall effect for the WT 1 scores ($F_{(1, 115)} = 40.32$, $p < .001$) with a large effect size (partial $\eta^2 = .43$). However, there was no significant effect for group, ($F_{(3, 113)} = .65$, $p > .05$) or for group × WT 1 score interaction ($F_{(3, 113)} = .63$, $p > .05$). Tests of between-subjects effects showed that there was a significant effect for WT 1 scores on the score improvement from WT 1 to WT 2 ($F_{(1, 115)} = 53.75$, $p < .001$) with a large effect size (partial $\eta^2 = .32$) and from WT 1 to WT 3, ($F_{(1, 115)} = 66.34$, $p < .001$) with a large effect size (partial $\eta^2 = .37$). There was no significant effect for WT 1 on the score improvement from WT 2 to WT 3 ($F_{(1, 115)} = .40$, $p > .05$).

Discussion

Our research questions asked whether the effect of the two forms of computer-automated metalinguistic CF in drills transferred to subsequent writing tasks. In our study, we adopted the outcome measures of the accuracy rates on the ECTs and on the writing tasks. Since the control group did not participate in the ECTs, our discussion of the first outcome measure is restricted to the feedback groups and the no-feedback group. In our study, the feedback groups improved significantly from ECT 1 to ECT 2, and there was no significant difference between the feedback groups in performance. This finding can be corroborated by the overwhelmingly affirmative results in previous metalinguistic CF studies that adopted similar outcome measures (e.g., Lado et al., 2014).

The no-feedback group, on the other hand, did not improve from ECT 1 to ECT 2. This indicates that participation in drills alone does not suffice to improve learners' accuracy in correcting errors. It is worth noting that our results contrast with the findings of Sanz and Morgan-Short (2004), which affirmed the efficacy of drills. As discussed in the literature review, Sanz and Morgan-Short (2004) exposed participants to the correct forms, which might have created a confounding effect. More importantly, the drills in our study encouraged a general focus on form without drawing attention to the target structure. In contrast, Sanz and Morgan-Short incorporated drills focusing on the target structure, increasing the salience of specific features in the input. According to cognitive interactionist theories, input salience and noticing facilitate language learning and acquisition (Schmidt, 1990). Hence, the lack of CF for the comparison groups in the study by Sanz and Morgan-Short (2004) might have been compensated for by the focused drills. Based on the above considerations, we suggest that when researchers discuss the efficacy of drills, they differentiate those with a general focus on form from those focusing on a particular form.

In our study, the effect of the two CF forms in drills did not transfer to subsequent written production. There was no significant effect for group on the improvement in writing accuracy scores. These results contrasted with the positive findings of other studies on computer-automated metalinguistic CF (e.g., Zhao & MacWhinney, 2018). Differences between the results may be explained by two factors. First, our study adopted a learned structure. According to skill acquisition theory, learners experience a sharp increase in accuracy and decrease in response time when they are first introduced to a rule. When they have been taught the rule, they undergo a much slower, fine-tuning process of automatization, in which the knowledge they draw on becomes highly specific and less transferrable (DeKeyser, 2015, pp. 96–98). The effect of CF provided in this later stage therefore might have been much less observable. Second, we selected different elicitation procedures. While previous studies into computer-automated metalinguistic CF selected highly-controlled practice for elicitation, our study selected both highly-controlled drills and free writing tasks. The difference between the results of previous studies and ours points out the importance of diversifying elicitation procedures to assess participants' learning results more comprehensively.

In a more general context, our findings also contrast with the positive findings of automated writing evaluation studies that adopted free writing tasks for elicitation (e.g., Chodorow, Gamon, & Tetreault, 2010; Lavolette, Polio, & Kahng, 2015; Wang, 2013). However, it was difficult to directly compare these results with ours because different outcome measures were adopted. Automated writing evaluation studies either compared error counts for the first and revised drafts of the same essays (e.g., Chodorow et al., 2010; Wang, 2013) or the holistic scores for the quality of different essays (e.g., Ebyary & Windeatt, 2010).

Finally, compared with the control group, the comparison group did not make significantly greater progress across the three writing tasks. This finding contrasted with those of Lyster (2004), who found the comparison group performed better than the control group. However, the instruction that the comparison group in Lyster's study received was focused on the target structure. Also notable is that Lyster studied the learning of a new rule. The discrepancy in the findings point out the necessity to consider the mediating role of drill focus and learner familiarity with the target structure when isolating the effect of CF from the effect of practice and input enhancement.

Conclusion

Our study investigated whether the effect of computer-automated metalinguistic CF by itself and with direct correction in drills transferred to accuracy of using the target structure on subsequent writing tasks. The results suggest that although metalinguistic CF, by itself or combined with correction, is effective on learners' accuracy in error correction, it does not lead to better accuracy on the subsequent writing tasks. The results also show that drills that encourage a general focus on form do not suffice to improve learners' performance on error correction or on subsequent writing tasks.

Our study had some limitations. First, although great care was taken to keep the conditions identical across the three writing tasks, we did not counterbalance the tasks, so we were unable to eliminate task effect. Second, this experiment was conducted toward the end of the term, so it was impossible for us to allot the participants a longer period between the treatment and the last writing task.

Future researchers can expand on our study by comparing the role of computer-automated metalinguistic CF in drills versus communicative tasks. We also recommend a more refined control- or comparison-group design that considers the mediation of task focus and learner familiarity with the target structure to isolate the effect of CF from the effect of practice and input enhancement.

Our findings show that the effect of computer-automated metalinguistic CF in drills does not transfer to subsequent written production when the target structure is a learned one. Instructors considering the incorporation of computer-automated CF as reinforcement are advised to take caution. Meanwhile, our results suggest that drills with a general focus on form are insufficient for the noticing of a learned structure. Teachers wishing to incorporate these types of drills are advised to combine them with focused CF to promote better noticing.

Acknowledgements

We would like to thank Prof. Charlene Polio from Michigan State University for her insightful feedback on the earlier drafts of this article. We would also like to thank Prof. Hansheng Wang from Peking University for his help with the statistics and Mr. Kehao Zhang for technical support. We thank our two research assistants, Xinwei Zhang and Lan Luo, and all the colleagues and students who participated in or supported the study. Finally, we are grateful to all the anonymous reviewers and the *LLT* editors for their insightful comments and suggestions. All remaining errors are our own.

Notes

1. Computer-automated CF differs from CF via CMC in that the former is provided by computers and the latter humans.
2. The definition of *practice* covers not only “task-essential practice” (Loschky & Bley-Vroman, 1993, p. 132), but also practice where the grammatical point is not critical or is not made the focus of attention (e.g., the written production tasks in Bitchener & Knoch, 2010).
3. Monteiro (2014) adopted both types of tasks and was counted in both categories.
4. According to Ellis (2009), this operationalization belongs to the “indication only” (p. 98) subtype of indirect CF.
5. Studies were counted by the individual data elicitation procedures.
6. We admit that more sessions would perhaps be needed to tease out the different effects of the two CF forms, as one reviewer pointed out.
7. For similar practices, see the studies by Muranoi (2000) and Sheen (2010).
8. We did not include a third ECT in the delayed post-test, because one of our research purposes was to find out whether drills could independently produce a durable effect on subsequent writing tasks. The introduction of a third ECT in the delayed post-test would have defeated this purpose.
9. As some of the data failed to meet the assumption of normality (the control group’s improvement from WT 1 to WT 2 and the no-feedback comparison’s improvement from WT 1 to WT 3), they were also analyzed with non-parametric tests using ranked data. The results were equivalent to those obtained from the MANCOVA.

References

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, *12*(2), 183–210.

Birk, J. (2000). Hit-the-Dot [Computer software]. Foster City, CA: QuinStreet, Inc. Retrieved from http://www.javascriptsource.com/games/hit-the-dot.html

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning on linguistic generalizations. *Studies in Second Language Acquisition*, *15*, 357–386.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*, 267–296.

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, *27*(3), 419–436.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). New York, NY: Routledge.

Diab, N. M. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, *24*, 16–34.

Diab, N. M. (2016). A comparison of peer, teacher, and self-feedback on the reduction of language errors in student essays. *System*, *57*, 55–65.

Ebyary, K. E., & Windeatt, S. (2010). The impact of computer based feedback on students' written work. *International Journal of English Studies*, *10*(2), 121–142.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79–114). San Diego, CA: Academic Press.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97–107.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

Heift, T. (2010). Prompting in CALL: A longitudinal study of learner uptake. *The Modern Language Journal*, *94*(2), 198–216.

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal*, *4*, 1–15.

Lado, B., Bowden, H. W., Stafford, C. A., & Sanz, C. (2014). A fine-grained analysis of the effects of negative evidence with and without metalinguistic information in language development. *Language Teaching Research*, *18*(3), 320–344.

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

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, *19*(2), 50–68.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309–365.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*(1), 1–14.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, UK: Multilingual Matters.

Lyddon, P. A. (2011). The efficacy of corrective feedback and textual enhancement in promoting the acquisition of grammatical redundancies. *The Modern Language Journal*, *95*, 104–129.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–426.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *20*, 37–66.

Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, *18*(3), 56–79.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*(4), 617–673.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, *77*(3), 330–339.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528.

Pienemann, M., & Lenzing, A. (2015). Processability theory. In B. VanPatten & J. Williams. (Eds.), *Theories in second language acquisition: An introduction* (pp. 159–179). New York, NY: Routledge.

Presson, N., MacWhinney, B., & Tokowicz, N. (2014). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, *35*(4), 709–737.

Rassaei, E. (2015). Oral corrective feedback, foreign language anxiety and L2 development. *System*, *49*, 98–109.

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.

Sanz, C. (2003). Computer delivered implicit vs. explicit feedback in processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 241–256). Mahwah, NJ: Erlbaum.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, *54*(1), 35–78.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, *13*(1), 96–120.

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

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, *32*, 203–234.

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*, 286–306.

Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, *49*, 110–119.

Shintani, N., Aubrey, S., & Donnellan, M. (2016). The effects of pre-task and post-task metalinguistic explanations on accuracy in second language writing. *TESOL Quarterly*, *50*(4), 945–955.

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103–131.

Stafford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness, practice, and cue learning. *Language Learning*, *62*(3), 741–768.

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, *32*, 303–334.

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

Wang, P. (2013). Can automated writing evaluation programs help students improve their English writing? *International Journal of Applied Linguistics and English Literature*, *2*(1), 6–12.

Wong, M., Zhao, H., & MacWhinney, B. (2018). A cognitive linguistics application for second language pedagogy: The English preposition tutor. *Language Learning*, *68*(2), 438–468.

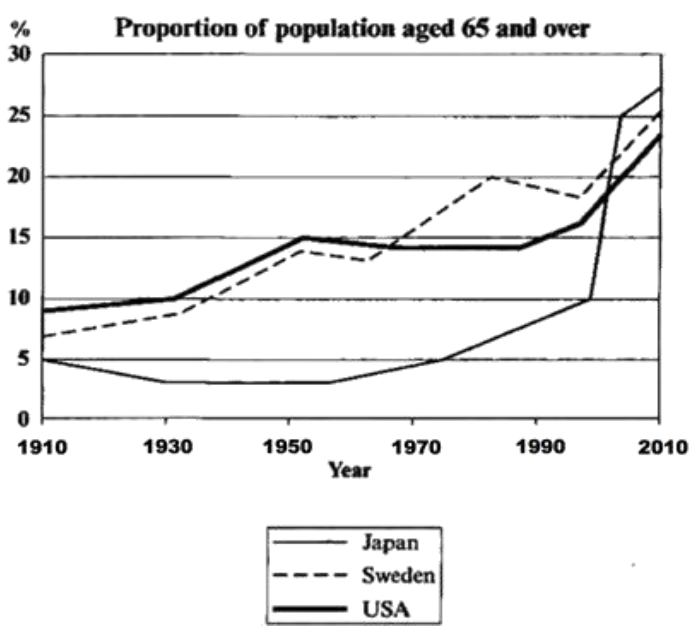
Zhao, H., & MacWhinney, B. (2018). The instructed learning of form-function mappings in the English article system. *The Modern Language Journal*, *102*(1), 1–21.

Appendix. Sample Writing Task

You should spend about 20 minutes on this task.

The graph below shows the proportion of the population aged 65 and over between 1910 and 2010 in three different countries.

Summarise in 150 - 200 words the information by selecting and reporting the main features, and make comparisons where relevant.



About the Authors

Jianwu Gao and Shuang Ma are lecturers at Capital Normal University, Beijing, China. They are currently teaching academic English writing. Their research interests include corrective feedback and L2 writing instruction.

E-mail: gaojianwu@cnu.edu.cn

E-mail: s.ma5384@cnu.edu.cn