Effects of teacher-scaffolded and self-scaffolded corrective feedback compared to direct corrective feedback on grammatical accuracy in English L2 writing

Jill A. Boggs

Abstract
This study aims to increase written grammatical accuracy by facilitating learners' use of corrective feedback (CF). A quantitative quasi-experimental study design is used to compare effects of traditional (teacher-provided) scaffolding and self-scaffolding, and compares these to unscaffolded direct written CF. Participants were Korean EFL university STEM students (n=109) in compulsory academic English writing classes in South Korea (hereafter *Korea*). To elicit metalinguistic reflections, traditional scaffolding was provided in 1-to-1 conferences, and self-scaffolding was provided by worksheets. Unscaffolded direct written CF was provided to a standard-treatment comparison group. The study was designed for ecological validity, with data being derived mainly from classroom writing samples, worksheets, and audio-recordings of conferences. Results provide support for previous CF research: mixed ANOVA results suggest that all three groups experienced similar, significant, and durable increases in grammatical accuracy. The study contributes to existing knowledge by a) using linear regression to demonstrate that quality of metalinguistic reflections does not necessarily predict an increase in grammatical accuracy; b) establishing that there may be difficulties in scaffolding oral metalinguistic reflections with the described population; and c) drawing on data from the background survey and interviews to inform the interpretation of the results.

Key words: corrective feedback; second language writing; scaffolding; languaging; metalinguistic reflections; metalinguistic explanations

Introduction

Corrective feedback (CF) is '[a]n indication to a learner that his or her use of the target language is incorrect' (Lightbown & Spada, 2013, p. 216). That is, the purpose of CF is not merely to provide the correct linguistic form, but also to draw learners' attention to it. Truscott's (1996) review of CF studies highlighted how poorly designed CF studies tended to be. Almost ten years later, it was suggested that the then-recent research with improved designs indicated that CF could indeed be beneficial, and that the issue to focus on was how and in what ways it could be *most* beneficial (Ferris, 2004). Research then turned to comparing effects of various types of CF: direct CF provides the location of errors and their corrections; indirect CF provides, without correction, indication of the existence of errors and sometimes the location as well e.g. with a tick mark at the location of the error or at the end of a line where there is an error; metalinguistic CF provides corrections as well as explanations about the errors (Bitchener & Storch, 2016, p. 1). While CF studies tend to demonstrate positive effects of CF in general, they also produce contradictory results about which type is most effective (Bitchener, 2012). This has led to increased interest in how learners use CF; and to calls for methodologically sound, ecologically valid CF, and for research to include consideration of learner context (Bitchener & Storch, 2016; Goldstein,

2016; Liu & Brown, 2015). The current study responds to this call by using quantitative methods to measure effectiveness, and then using participant data from a background survey and interviews to inform the interpretation of these results.

The contradictory results in CF research might be explained by differing philosophical approaches; by the type of CF provided; or by learner-specific factors. Philosophical approaches to CF research tend to be classified into either second language acquisition (SLA) approaches or L2 writing approaches (Ferris, 2010). SLA approaches may prefer tasks which are tightly controlled, such as picture descriptions or translations, and may limit their focus to one or two grammatical features (Sheen, 2007; Bitchener and Knoch, 2009). The generalisability of these studies to classrooms can be unclear. Contrastingly, L2 writing approaches tend to be classroom-based (Ferris, 2010), where learners might engage in a writing process, producing drafts and revisions. Tasks in L2 writing studies might allow for more personalisation, possibly resulting in more varied content between participants, which could require varied types (or uneven distribution) of correction. While a strength of L2 writing approaches is ecological validity, a corresponding weakness is the loosening of control; and while the control of SLA approaches tends to result in rigorous designs, a corresponding weakness is difficulty applying the results to classrooms (Liu & Brown, 2015).

The second factor contributing to the inconsistency of CF study results are different CF types, each with benefits and drawbacks. Direct CF has repeatedly been seen to increase grammatical accuracy (Bitchener, 2012), a strength being the unambiguous information it provides about correction. A weakness, however, is that learners may not understand why the correction was necessary, and thereby fail to learn from it (Ferris, 1995). Another weakness is that direct CF can fail to engage learners, allowing them merely to copy the correction when revising their work without analysing the underlying cause of the error (Shintani & Ellis, 2015).

Indirect CF is argued to address some of the weaknesses of direct CF, as it can be seen as engaging learners in analysing their work (Ferris, 2010). Because learners are not provided with the correction, they are prompted to analyse their work closely to identify the problem in order to correct it. However, as with direct CF, no explanation is given, so learners may not understand why the correction was necessary. While some research suggests this disadvantage applies particularly to lower-proficient learners, at least for oral CF (Li, 2009),

2

there is evidence that learners with higher second language (L2) proficiency can also experience difficulties (Storch & Wigglesworth, 2010). Furthermore, indirect CF may frustrate learners who struggle to identify the error (Author, 2018; Hedgcock & Lefkowitz, 1996); or it may result in learners missing the target of the CF, creating unexpected problems, e.g. as learners change correct linguistic forms to incorrect ones (Ruegg, 2016).

Metalinguistic CF is a technique which can resolve the problem of insufficient information, a weakness of both direct and indirect CF: the error is explicitly corrected, reducing ambiguity; and the metalinguistic information explains the reason for the correction. This information can help learners 'to understand the nature of the error they have committed' (Ellis, Sheen, Murakami, & Takashima, 2008, p. 356). Despite this additional information, however, studies comparing metalinguistic CF to other forms often fail to find significant differences between treatment groups (Bitchener, 2008; Bitchener & Knoch, 2008; 2009a; 2009b). A possible explanation for this might be found in processing theory (McLaughlin, Rossman, & McLeod, 1983); that is, to effect greater learning, the information may need to be more deeply analysed by the learner than the mere provision of the information. That is, while metalinguistic explanations provide *opportunity* for deeper analysis, they cannot compel learners to take up this opportunity. Learners might skim metalinguistic CF, spending no more time looking at metalinguistic CF than they spend looking at direct CF (Shintani & Ellis, 2013). Any benefit offered in theory may not be experienced by learners if they cannot or do not use CF advantageously.

As suggested by the preceding discussion, the third factor which may contribute to the inconsistency of CF research results is what learners do with provided CF. One way of helping the learner use the CF advantageously may be to provide scaffolding, which is a 'process that enables [a] novice to solve a problem, carry out a task or achieve a goal which would be beyond his [*sic*] unassisted efforts' (Wood, Bruner, & Ross, 1976, p. 90). While there are differing ideas about how scaffolding should be conceptualised, there is general agreement in the field that scaffolding can be used to assist learners in increasing their level of understanding or ability (Lepper, Drake, & O'Donnell-Johnson, 1997; Puntambekar & Hubscher, 2005; Stone, 1998).

However, traditional views of scaffolding can be prescriptive (Littleton, 2013) or deceptively linear (Mascolo, 2005), or dismissive of learners' agency (Granott, 2005). A more suitable

conceptualisation might be 'functional scaffolding' (Bickhard, 2005); that is, viewing scaffolding as a sequence of resources which differ in level of accessibility to the learner. Expert-provided resources might be replaced with more autonomous ones as learning progresses, with different resources becoming more appropriate over time. It is self-evident that '[a] person cannot provide to him or her self-knowledge [*sic*] that is otherwise not available' (Bickhard, 2005, p. 170); however, learners can learn to use existing resources for self-scaffolding (Bickhard, 1992a; 1992b). In practice, the less autonomous end of the scaffolding continuum might be teachers scaffolding understanding through instruction, while the opposite end is learners independently using resources to support their own learning.

Research has investigated the scaffolding of language learners' use of CF by engaging learners in *languaging* (Knouzi, Swain, Lapkin, & Brooks, 2010), defined as 'the process of making meaning and shaping knowledge and experience through language' (Swain, 2006, p. 98). Importantly, engaging learners in languaging could address the issue of compelling learner attention to CF; the deeper analysis of grammatical features that languaging requires could effect greater learning, as proposed by information processing theory (McLaughlin et al., 1983). As studies on languaging are relatively new, there is some inconsistency in terminology; here, *metalinguistic explanation* describes the provision of linguistic explanations <u>to</u> a learner as a form of CF (Bitchener & Knoch, 2008; 2009a; 2009b), while the term *metalinguistic reflection* refers to the product of languaging <u>by</u> the learner (Simard, French, & Fortier, 2007).

Languaging could be particularly useful for L2 learners who have lower L2 proficiency or are less experienced using feedback (Williams, J. and Kane, 2009; Esteban and Roca de Larios, 2010). While some Western educational cultures emphasise writing skills as key to academic success, other cultures may not share this view, which could impact on how well learners from such academic cultures are able to use feedback when confronted with it. Some East Asian educational systems, such as Japan and Korea, have tended to de-emphasise writing (Williams, C., 2017), which could result in students placing low value on writing. Such learners may 'not perceive EFL writing as important for their future study or careers', and university students who are not in English programmes of study 'often do not perceive a need for writing in English' (Reichelt, 2009, p. 196). If learners have little experience writing, they logically might have correspondingly little experience receiving and using CF.

There can be several explanations for writing instruction being de-emphasised in parts of East Asia; the need for students to pass high-stakes multiple-choice university entrance examinations is given as one reason (Reichelt, 2009; Shintani, Ellis and Suzuki, 2014). In Korea, the College Scholastic Ability Test (CSAT, or *suneung*) is the national university entrance examination. It is a multiple-choice test of academic subjects, including English, science, and maths (KICE, 2015). A full discussion of Korean educational culture and the impact of the CSAT is beyond the scope of this paper (but Song, 2011, offers an excellent discussion). To glimpse the importance of the CSAT, consider the action taken on the day of the examination:

> [On the day of the CSAT,] the government rescheduled working hours, airplanes and military jets were grounded, and live-fire military drills were suspended near test sites during listening tests. Government offices and companies [began work an hour later] and the stock market opened an hour later to help ease morning rush hour traffic so students could arrive at test sites in time. Construction work was also halted during Korean and English listening test hours (Lee, 2011).

Because of the importance of the CSAT, secondary school teachers focus on helping students succeed on this multiple-choice test (Ro, 2017); to do otherwise would be a disservice to the students. Nonetheless, this can leave little time for other areas of study. L2 learners with weak L1 writing skills may be at a disadvantage when learning L2 writing (Cumming and Riazi, 2000). With little writing experience and little experience receiving feedback, it is possible that learners with test-centric educational backgrounds might benefit from being prompted to engage in languaging as a way of scaffolding their use of CF.

There is empirical support for oral and written languaging, but the support tends to be limited. Oral languaging studies tend to be small scale (Storch, 2008: n=22; Swain et al., 2009: n=9; Knouzi et al., 2010: n=2), meaning that effects are not usually statistically measured, so inferences are not usually made. Oral languaging studies also tend to involve tasks that prompt learners to produce similar texts, such as text reconstructions (Storch, 2008); or which focus on a single grammatical feature (Knouzi et al., 2010; Swain, Lapkin, Knouzi, Suzuki, & Brooks, 2009). Results of such studies might not apply to larger classes where participants have written unique texts or are learning multiple grammatical features. Written languaging is defined as 'learners' self-explaining of the target grammar rule in writing' (Ishikawa & Suzuki, 2016, p. 98). Like oral languaging studies, written languaging studies also indicate a potential positive contribution to L2 development (Ishikawa, 2013; Suzuki, 2012). For example, of all instances of grammar correction (n=322) given to twenty-four university-aged

Japanese L1 participants in Suzuki (2012), most (n=293) errors were successfully resolved on a revision task following a languaging activity. However, written languaging studies also tend to be small scale and have limited generalisability. For example, results of studies conducted with participants in English degree programmes, as those in Suzuki (2012) appear to be, might not apply to populations preparing for science degrees (Hyland, 2013; Rao & Liu, 2011; Wong & Nunan, 2011). Another issue is the task; common tasks in languaging studies include revision tasks (Suzuki, 2012); translation tasks (Ishikawa, 2013); and dictogloss tasks (Ishikawa, 2018). Such highly controlled tasks cannot indicate how languaging might influence accuracy in tasks where learners produce their own writing, which might be more common in typical classrooms (Liu & Brown, 2015). Furthermore, studies often measure improvement with revisions, receptive tests, or fill-in-the-blank tests, so effects of languaging on accuracy in freer writing, where it is arguably most needed, are unclear.

This discussion demonstrates that languaging as a (self-) scaffolding tool can be used to help learners benefit from CF. It also demonstrates the need for studies with rigorous designs to measure effectiveness, such as having intervention groups large enough to support the use of statistical analyses; and for studies with increased ecologically valid designs which can clarify pedagogical applicability. Studies which bridge the gap between the L2 writing and SLA perspectives can contribute to meeting these challenges.

The current study compares effects of expert- and self-scaffolded languaging and effects of a group receiving unscaffolded, direct written CF on new writing; that is, writing which is not the revision of a draft (Storch, 2010; Shintani, Ellis and Suzuki, 2014). Drawing from an L2 writing approach, the study applies ecologically valid tasks and treatments, and includes consideration of participant context and personal variables. Drawing from the SLA approach, the design provides rigour by (a) including sufficient participants to allow inferential statistics to be used; and (b) controlling for known variables, such as L1 writing experience; L2 proficiency; knowledge of an L3; prior knowledge of targeted grammatical features; and time on task. The ecological validity of the study is further strengthened by drawing on data from post-study interviews with participants to inform the discussion of the quantitative results (Liu & Brown, 2015).

The research questions for the current study are:

1. What is the differential effect of using languaging (a) as scaffolding and (b) as self-scaffolding compared to (c) unscaffolded direct CF on accuracy of targeted grammatical features in new writing, as measured by error ratios?

2. Does the quality of metalinguistic reflections predict an increase in accuracy of targeted grammatical features in new writing?

Material and methods

The study schedule

The current study was conducted in a compulsory English L2 academic writing class during the first term of participants' first year at a highly-ranked university in Korea. The aim of the course is to help prepare the students for success in their undergraduate academic careers, where approximately 30% of the courses will be delivered in English. This initially appears to imply a *learning-to-write* context, where instruction is focused on writing skill (Manchón, 2011). This can be contrasted to using writing to work on L2 language skills, or what Manchón calls *writing to learn (language).* However, due in part due to the low productive skill and lack of L1 and L2 writing experience of the participants, as discussed next, the course combines elements from both strands. The current study focuses on the *writing to learn (language)* aspect. Students attended two sessions of the class per week (e.g. Mondays and Wednesdays) for 75 minutes each. The schedule for tasks and interventions is provided in Table 1.

	Session 1	**Session 2**
Week 1	Consent and information	Background survey and proficiency test
Weeks 2 - 6	Content instruction on writing and targeted grammatical features	
Week 7	Instruction (continued)	**Pre-test** (Task 1)
Week 8	CF intervention 1	Task 2 (interim task for feedback purposes; not analysed in the study)
Week 9	CF intervention 2	**Immediate post-test** (Task 3)
Weeks 10-13	Content instruction (on writing only; no further focus on grammar)	
Week 14	Content instruction (continued)	**Delayed post-test** (Task 4)
Week 15	Interviews	

Table 1 Schedule for the study with tests and interventions shaded for emphasis

The participants

Two teachers, English L1 males with eight and twelve years of teaching experience volunteered to facilitate the study. Both teachers had some knowledge of the participants' L1 and held masters' degrees and Teaching English as a Foreign Language (TEFL) qualifications.

The two teachers together had 195 students, 122 of whom met inclusion criteria (explained next) and gave consent (provided in both L1 and L2). Of these, 109 completed Tasks 1 – 4 (Table 1), and their data were included in the quantitative analyses. English L2 proficiency was measured, as in previous research (Li & Post, 2014), using the Oxford Online Placement Test (OOPT; OUP, 2014), and was found to be low intermediate. Participants were 75% male, as is common in science, technology, engineering, and mathematics (STEM) fields in Korean universities (WISET, 2015). Inclusion criteria were: at least 18 years of age (the oldest was 22; mean age was 19.1); had neither lived abroad nor attended international or foreign language secondary schools; if there was an L3, it was limited to elementary-level Japanese or Chinese; and, if male, had not yet fulfilled their compulsory military service duties. As seen in other studies in East Asia, participants reported little L1 writing experience (Chandler, 2003; Shintani & Ellis, 2015); that is, most (77%) reported writing no or very few 'extended length' L1 essays, defined as three or more typed pages. Most (93.4%) reported coming from

a science-focused secondary curriculum, which can result in less instructional time being apportioned to language arts (González-Becerra, 2017).

A pilot study, conducted at the same institution with different participants in the year prior to the current study, informed the design of this study: targeted grammatical features, time needed for writing and CF provision, and CF treatments. Details are discussed in the relevant sections.

Instruments

The treatment of the targeted grammatical features (hereafter *targeted features*) was informed by the study's aim for ecological validity in two primary ways, mentioned briefly here and expanded on below: instruction on and quantity of targeted features. Some CF researchers might consider the inclusion of instruction problematic. Research from an SLA perspective might prefer to provide CF with no instruction of the targeted features being given in advance of the tasks, termed *CF-for-new-knowledge* (Lyster, Saito, & Sato, 2013). This approach can result in metalinguistic CF resembling typical classroom instruction, delivered after the task *en masse*, either orally to the class (Bitchener, 2008; Bitchener & Knoch, 2008; 2009a; 2009b), or on handouts (Shintani, Ellis, & Suzuki, 2014). However, it seems that CF might be of 'little use when there is no initial learning or surface information' (Hattie & Timperley, 2007, p. 104). There is empirical support for this argument in the literature; e.g., participants in Ammar and Spada (2006) who exhibited some prior knowledge of the targeted forms responded better to CF treatments. This suggests that there may be differences in learners' processing of CF on new grammatical features compared to CF on features previously studied. These different notions of CF might usefully be conceptualised separately as *CF-as-instruction* and *CF-as-reinforcement*, each with its own purposes and outcomes. Particularly in some instructed learning contexts, such as that of the current study, CF-as-reinforcement might be preferable. Thus, following Ammar and Spada (2006) and others in viewing CF as a tool for supporting instruction, the current study provides instruction on targeted features.

The second issue regarding targeted features is the quantity of these targeted features. While many CF studies focus on one or two grammatical features, both students and teachers might understandably resist such limited focus. It is relevant to investigate CF as it might be used in real classrooms using 'mid-focused' feedback (Liu & Brown, 2015); that is, neither focused on a single grammatical feature nor on all features. Eight grammatical features were targeted

9

for correction (**Table 2**); five from the required textbook, which were given explicit instruction, and three from students' secondary school curriculum, which, having been taught previously, were not given further instruction. These three previously-taught features were included because a needs analysis conducted during the piloting phase revealed that this population still struggled to use these with accuracy; other studies in the region have also reported students lacking productive command over elementary-level features (e.g. Ishikawa, 2013). An error correction pre-test confirmed that all eight targeted features presented difficulty.

Grammatical features drawn from textbook

1. **at/ on/ in:** prepositions of time
2. **adverbs of frequency**: always, never, etc.
3. **most (of)**; some (of); all (of); etc.
4. **comparatives and superlatives**
5. **definite and indefinite articles:** limited range of applications, such as first and second mention and using *the* with things in nature – they sky, the ocean, etc.

Grammatical features drawn from needs analysis

6. **subject/ verb agreement**
7. **plural -s**
8. **select tenses**: simple present; present continuous; simple past; future

Table 2 Selection of targeted grammatical features

The writing tasks were scheduled as part of the regular curriculum (Table 1), and were the main source of data. Reflective worksheets for the worksheet intervention were adapted from Santos, López-Serrano and Manchón (2010). Student–teacher conferences for the conferencing intervention were audio-recorded. Table 1 provides the schedule of the tasks and interventions. Strengthening the study and enhancing ecological validity, the two teacher-participants collaborated with each other and with the researcher to develop one set of lesson plans and accompanying PowerPoint presentations for the term, which they then both used; the researcher confirmed that grammar lessons were delivered similarly by viewing the class videos recorded by the institution. All participants completed a background survey, used to confirm eligibility for the study and to gain deeper understanding of learning histories. Post-study interview data from approximately 10% (n=11) of the participants was obtained, and is used here only to add depth to the understanding of the results. A full report of this aspect of

the study is forthcoming (Author, 2019). The semi-structured interviews were formed around the following questions:

1. How did you study English before university (at academy/ at school)?
2. Can you tell me about your English classes (academy/ school)? Did you write for your classes? What did your teachers do about grammar errors?
3. Your teacher taught you some grammar this semester. What did you think about these lessons?
4. How did you feel when you received feedback?

Procedure

Intervention groups. Given the empirical evidence for the value of CF to learning, the teacher-participants were understandably uncomfortable about withholding it, so the current study applies a *standard-treatment control group* (henceforth *Comparison group*) rather than a *no-treatment control group* (Vogt, Gardner and Haeffele, 2012). The three groups were stratified by known variables (English proficiency, prior education, teacher, gender, and educational history (Cohen, Manion, & Morrison, 2011, p. 154)), allowing similarity between groups to be confirmed. Stratification was achieved by randomly assigning participants to groups, and then moving participants to balance the variables. Expert-scaffolded participants (n=37) received 4-5 minute 1-to-1 conferences, with teachers instructed to scaffold participants' identification and understanding of errors by engaging participants in scaffolded languaging. Self-scaffolded (n=34) participants completed Reflective worksheets (Figure 1), self-scaffolding through written languaging. Comparison group (n=38) participants received unscaffolded direct CF. Interventions are explained in detail below.

Writing sessions. For the pre-test, participants chose from four local current-issues topics (same-sex marriage; regulation of antibiotics; local pollution problems; or *Iodo*, a disputed local island territory). Participants had ten minutes with their phones to make notes, e.g. dates and translations, on a provided notecard. Following the note-taking, mobiles were put away, and lined paper was provided. As explained, the current study aimed to investigate CF ecologically, taking a CF-as-reinforcement view of feedback; the targeted features had been explicitly taught, either in secondary school or in the current classroom, and participants were now being asked to use them. As might be done in a typical classroom, the teachers listed the features (without explanations or examples) on the whiteboard, reminding participants to use them.

Participants wrote one paragraph (8-10 sentences) in the remaining approximately 50 minutes, seen in the piloting phase as appropriate. Participants left the room after submitting their work (Storch, 2008). Subsequent writing tasks followed the same procedure, with each focusing on a different aspect of the same broad topic, assigned at the start of the writing session. After each session, teachers collected the notecards (to confirm appropriate use) and paragraphs.

Preparation of feedback. The researcher provided written CF for all tasks, focusing only on targeted features, and then passed the work to the teachers. The teachers reviewed the researcher's feedback and provided additional comments on content only, using the same colour of pen. Pilot study participants had demonstrated difficulty differentiating between grammar feedback and other types of feedback (e.g. content, formatting), which resulted in them writing metalinguistic reflections (MLRs) about features other than grammatical errors. To resolve this, grammar errors were numbered in the current study on the work of participants in all three groups.

Direct written CF was provided on the original work of the self-scaffolded group and the comparison group. For the expert-scaffolded group, direct written CF was marked on photocopies, leaving the students' original work clean for use during the conferences.

Delivery of feedback to intervention groups. After receiving training from the researcher on the conference procedure, the teachers delivered the CF to the participants; the novelty of having someone other than their teacher deliver the CF might have caused participants to react differently (Gravetter & Forzano, 2016, p. 172). Each group received two feedback treatments; previous research with this population indicated that a minimum of two feedback treatments may be needed for increased accuracy to be observable (Author, 2018). The scaffolding structure for the conferences was based on Aljaafreh and Lantolf's (1994) Regulatory Scale, moving through steps to assist the learner only as much as necessary. The initial steps are preparatory: learners self-correct their errors and are invited to the conference. Next, the teacher 'indicates that something may be wrong… [e.g.] "Is there anything wrong with this sentence?"' (ibid.:471). The learner is then given two prompts (Steps 5 and 6) to help them identify the targeted error, with each prompt becoming more specific, before the tutor identifies the error for the learner, if necessary (Step 7). The learner is given two

opportunities to provide the correction (Steps 8 and 9). Steps 10-12 ask the tutor to provide the correct form, an explanation for it, and some examples.

In the current study, the initial preparatory steps in Aljaafreh and Lantolf's (1994) scale were omitted, as the inclusion of this autonomous phase could be a confounding variable in the treatment, and instead began by indicating the general location of the error and prompting student-participants to identify it. An example follows:

1 Teacher Look at the first line, please. Can you find any mistakes?
2 Student [No reply]
3 Teacher It's in the first sentence.
4 Student [No reply]
5 Teacher How many problems is this? (seeming to point to the noun *problem*)
6 Student … four?
7 Teacher Good. So what should we say?
8 Student *A* problem.
9 Teacher Yeah. Why? What does *a* mean?
10 Student One
11 Teacher Yes, are there other problems? Do other problems exist?
12 Student Yes.
13 Teacher Yes, one among many. Ok, line 2. Can you find a mistake in line 2?

The conferences were designed to be focused and use direct questions when eliciting information, as participants in the pilot study had struggled to participate in open-style conferences. Despite this, the level of apparent reticence in the current study was greater than expected, and plans for the analysis of MLRs in the conferences were abandoned due to low participant input. Contributing factors are explored in a separate study (Author, 2019).

Self-scaffolded group participants received direct written CF, with grammar errors numbered as described above. A worksheet (Figure 1) was stapled to their marked work; oral and written instructions were provided. Participants were invited to access resources (i.e. their class notes and textbooks) to complete the task. The first two rows of the worksheet provided information about how to complete it; rows 3 – 21 (two-sided paper) provided blank spaces for students to fill in. The final row in Figure 1 is an example from a student-participant; spelling errors were not addressed.

Figure 1 Reflective worksheet example

The CF interventions took place twice (***Table 1***), once after Task 1 and again after Task 2. Immediate and delayed post-tests did not receive CF until after the study, though teachers provided feedback on content as usual. Participants wrote revisions in class immediately following the interventions. CF studies often involve removing the original, corrected work when assigning revisions in order to measure change between original work and these revisions. However, the current study focuses only on new writing; revisions were not analysed, but, contributing to the ecological validity of the study, were required as best practice (Chandler, 2003; Liu & Brown, 2015; Shintani et al., 2014). Thus, participants could refer to their original work (with corrections) while writing revisions. In order to ensure participants' focus was on the CF rather than on their mark (Taras, 2001; Vardi, 2009) teachers distributed scores on a separate paper as participants submitted their revisions.

Analysis of writing samples

Errors were counted as follows:

- Repetitions of errors were counted separately (e.g. using *sky* without the definite article multiple times counted as multiple errors)

- Errors creating knock-on effects were treated as single errors (e.g. *Apple is my favourite fruit* counted as one error, with *apples* and *are* being corrected simultaneously)

- Article errors were marked for a limited range of guidelines taught in class (e.g. use *the* for things in nature; e.g. the sky, the ocean, etc.). The sum correct obligatory uses was divided by the sum obligatory uses.

- Some writing samples contained no contexts where articles were required, which could be a reflection of the tendency to use generalisations (plurals) in academic writing. As Korean is a language without articles, there was no way of knowing whether omissions of articles were intentional or the result of L1 transference, so instances of obligatory null articles were omitted from analysis.

Similar studies tend either to analyse student writing by t-units, defined as 'an independent clause and all of its dependent clauses (Sachs & Polio, 2007, p. 79) or by error ratios (Sheen, 2007). Analysis by t-units can be problematic with lower-proficient writers, introducing bias as the researcher interprets and parses the sentences (Gaies, 1980; Ney, 1966). Thus, following previous research (Sheen, 2007) error ratios were used (i.e. numbers of errors per 100 words). A teacher with an MA TESOL, who was not part of the study, served as the inter-rater. She marked 10% (n=12) of the pre-tests, resulting in an intraclass correlation coefficient of .935.

Analysis of worksheets

Following previous research (Sachs & Polio, 2007; Storch, 2008), worksheets were coded for quality, as it is possible that higher-quality MLRs could indicate deeper analysis, which could lead to greater learning. Storch's (2008) two-level coding scheme (Elaborate/ Limited) was modified: a *Null* category was added for missing or inaccurate MLRs (0 points); the Limited category (1 point) was applied when MLRs were close but not quite correct or complete; the Elaborate category (2 points) was used for MLRs which clearly demonstrated understanding (rules or examples). Different participants produced different numbers of MLRs, so a percentage was calculated to normalise the scores. An example of the calculation process follows:

Elaborate MLRs	@ 2 point each	3	= 6
Limited MLRs	@ 1 point each	3	= 3
Null MLRs	@ 0 points each	2	= 0

Total points awarded 9
Maximum possible points 16 (8 explanations)
Final score 9/16 = 56%

A Korean L1 colleague nearing completion of her doctoral studies in applied linguistics served as the inter-rater. She marked 17% (n=6 out of 34) of the worksheets, resulting in an intraclass correlation coefficient of .926. As explained, plans to analyse MLRs in conferences were abandoned due to low student-participant input; this will be examined in Author (2019).

Results

An overview of the results is provided first, followed by the results of each research question in turn. Table 3 provides numbers of words written and errors made at each time point. This information is depicted graphically in Figures 2 – 3.

	Group	n=	Words written by students		Errors (targeted features)	
			Mean	SD	Mean	SD
Pre-test	Expert-scaffolded	37	134.3	31.6	7.5	3.9
	Self-scaffolded	34	156.0	68.8	7.6	4.2
	Comparison	38	133.7	42.2	7.9	4.4

	Group	n=	Words written by students		Errors (targeted features)	
Post-test	Expert-scaffolded	37	130.5	27.4	3.7	3.2
	Self-scaffolded	34	141.4	34.1	4.7	4.2
	Comparison	38	128.2	25.4	5.7	3.9

	Group	n=	Words written by students		Errors (targeted features)	
Delayed post-test	Expert-scaffolded	37	142.5	44.6	4.3	3.2
	Self-scaffolded	34	160.3	48.3	4.1	3.1
	Comparison	38	142.7	45.6	5.6	3.1

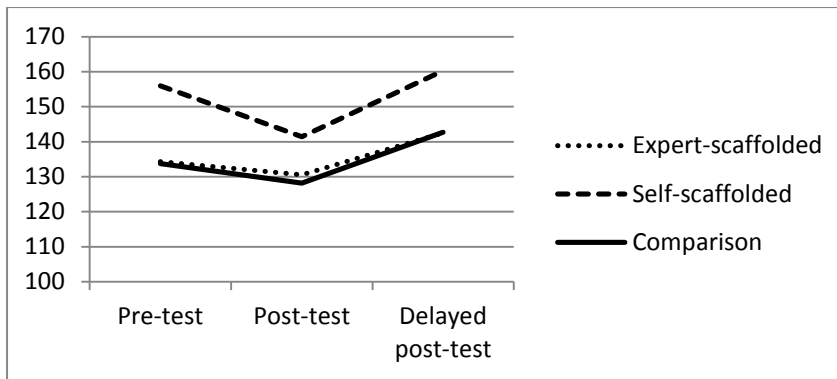*Table 3 Mean words written and errors made at each time point*

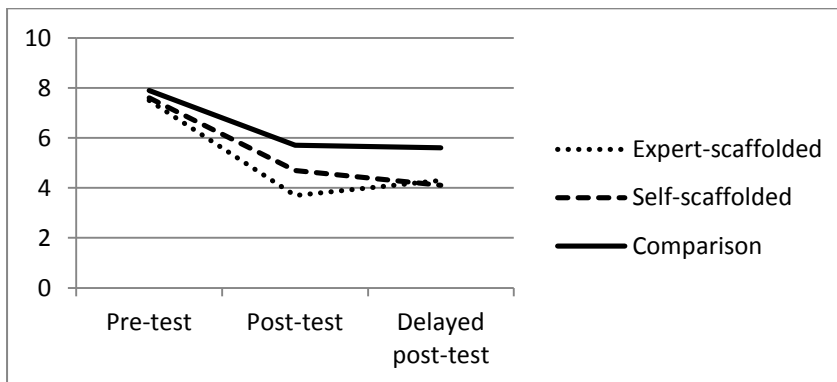***Figure 2 Mean words written at each time point***

Figure 3 Mean targeted feature errors at each time point

RQ1: What is the differential effect of languaging (a) as scaffolding and (b) as self-scaffolding compared to (c) unscaffolded direct CF on accuracy of targeted grammatical features, as measured by error ratios?

The data for each group in the study were evaluated using SPSS, version 23. A mixed ANOVA was run with *time* as the within-subjects factor and *group* as the between-subjects factor. A log10 transformation corrected moderate skewness, and all other assumptions were met.

The results show a significant main effect of time $F(2, 212) = 37.856$, $p < .001$, partial $\eta^2 = .263$, indicating improvement on the whole, and a significant main effect of group $F(2, 106) = 3.767$, $p < .026$, partial $\eta^2 = .066$, indicating some differences overall between groups (found to be between the comparison group and the self-scaffolded group). However, there was no statistically significant interaction between time and group, $F(4, 212) = 2.125$, $p = .079$, partial $\eta^2 = .039$, indicating that there were no significant differences between groups regarding the degree to which the error ratio had decreased. The ANOVA provides information about the groups in relation to each other; to address possible concern about

17

whether each group had improved their grammatical accuracy, paired-samples t-tests were run. All three groups showed significant improvement from pre-test to post-test, and the improvement was seen to be durable at the delayed post-test. Transforming the data was unnecessary, as paired-samples t-tests are robust to violations of normality. Regardless, tests with transformed and untransformed data produced similar and significant results. Table 4 provides the results.

	Expert-scaffolded	Self-scaffolded	Comparison
Pre-test – Post-test	$t(36) = -5.440, p <.001$	$t(33) = -5.451, p <.001$	$t(37) = -2.784, p =.008$
Pre-test – Delayed post-test	$t(36) = -4.742, p <.001$	$t(33) = -5.830, p <.001$	$t(37) = -3.414, p = .002$

Table 4 Paired-samples t-test results for the three groups

RQ2: Does the quality of metalinguistic reflections predict an increase in accuracy of targeted grammatical features in new writing?

As explained, only the data from the self-scaffolded group could be analysed. Thus, a linear regression was used to see whether improvement in pre-test to post-test gain scores for improvement in accuracy of targeted features (outcome variable) could be predicted by the worksheet quality score (predictor variable). Table 5 provides worksheet quality scores.

	Mean	SD
Worksheet 1	70.2	23.4
Worksheet 2	72.8	25.6

Table 5 Mean worksheet quality scores (%)

The score from Worksheet 2, shaded in Table 5, was used in the analysis as it was thought this would best reflect participants' languaging ability, as participants would have had some practice with the first worksheet. Regardless, the first and second worksheet scores are similar. It was found that worksheet quality score did not predict improvement in accuracy of targeted features[1], F(1, 32) = .396, p = .534.

---

[1] An anonymous reviewer notes that it cannot be known exactly what the self-scaffolding participants actually did while completing the worksheets. The participants had 75 minutes to complete the worksheets and their revisions, and then they were required to submit all work, so there was little time for them to try to do more. Nonetheless, it is acknowledged that the possibility exists that for all groups there might have been cognitive engagement that the study could not capture.

Discussion

There were three main findings. a) Expert-scaffolded, self-scaffolded, and comparison groups all demonstrated similar, significant, and durable increased accuracy in the use of targeted features. b) Apparent reluctance of participants to actively engage in conferences meant that there were insufficient data to investigate possible effect of quality of MLRs, and possible reasons for this are explored next. c) While self-scaffolded participants demonstrated willingness to engage in languaging on the worksheets, increase of grammatical accuracy was seen, perhaps surprisingly, regardless of the quality of MLRs. These points are discussed next.

RQ1: What is the differential effect of using languaging (a) as scaffolding and (b) as self-scaffolding compared to (c) unscaffolded direct CF on accuracy of targeted grammatical features in new writing, as measured by error ratios between groups (RQ1)

The data do not provide immediate support for the hypothesis, informed by information processing theory (McLaughlin et al., 1983), that languaging could result in greater learning, as evidenced by increased grammatical accuracy. However, that effects were not immediately visible does not necessarily mean that greater learning did not occur; learning might be evidenced at a later point, as 'any SLA notion of progress has always been non-linear, gradual, unevenly paced, and often proceeding through interim nontargetlike (but developmentally helpful) solutions' (Ortega, 2012, p. 408). It is possible that a single university term is insufficient time for learning to be evidenced.

There are several possible explanations for why the three groups in the current study performed similarly. It is possible that, at least for lower-proficient learners, perhaps the form of CF is less important than the fact that it is provided (Bitchener & Knoch 2009b). Participants receiving CF in Bitchener and Knoch's studies (2008; 2009a; 2009b; 2010), including indirect CF; direct CF; oral/ written metalinguistic CF; or combinations of these, significantly outperformed groups receiving no CF, but performed similarly to each other. A lack of time pressure in the current study might also contribute to the observed similarities between groups (Santos et al., 2010). Perhaps with sufficient time to write and reflect, the combination of (a) explicit instruction of targeted features, (b) provision of CF in *some* form, and (c) required revisions were sufficient to aid the development of accuracy, and that scaffolding CF use for this population may not add further benefit. Finally, a plausible explanation is that the inclusion of multiple grammatical features might have influenced the results (Shintani et al., 2014), placing too high a demand on information processing resources.

While this possibility cannot be discounted, a strength of the current study is its ecological validity; teachers and students alike may resist spending an entire term focusing on one or two grammar points.

RQ2: Does the quality of metalinguistic reflections predict an increase in accuracy of targeted grammatical features in new writing?

As explained, there was insufficient participant input for analyses to be conducted on the data from the expert-scaffolded group. However, it was unexpected that self-scaffolded group participants who produced higher-quality written MLRs did not demonstrate greater increases in accuracy compared to participants who wrote lower-quality MLRs. While benefits from information processing's *deeper analysis* might not be seen if the analysis is too challenging (Baralt, 2013), this seems not to have been the case in the current study. Self-scaffolded languaging seems to have provided an appropriate level of challenge, as demonstrated by the acceptable scores of worksheet quality (Table 5). The possibility exists that the task design could not prevent participants from completing the worksheets with minimal engagement (Shintani & Ellis, 2015). Even though participants in the self-scaffolded group in the current study were required not only to copy the corrections but also to produce MLRs, it is possible that they did so with insufficient levels of cognitive engagement.

If the results seen in all three groups can be considered the minimal expected improvement under the conditions, an examination of data from the background survey and interviews could provide possible explanations. As discussed earlier, the participants in the current study reported very little experience writing in Korean L1, and it was seen that secondary school science curricula can allow little room for English L2 writing. These participants are not alone in their lack of writing experience, as discussed in the introduction. A low valuation of writing could affect students' engagement. If participants come into an L2 writing class with little or no experience with L1 or L2 writing, then they likely have correspondingly little experience receiving CF in either language; it is perhaps unsurprising that they might struggle to use it, even with scaffolding. While there is evidence that this de-emphasis on writing is changing in the region (Rinnert & Kobayashi, 2009), the general lack of writing experience is consistent with the current study and other reports in the region (Williams, 2017).

A further finding from this dataset is that participants' lack of prior writing experience and their English educational background might have influenced their views of what it means to learn or know a grammatical feature. There are two forms of evidence for this in the data. The first is that comments in post-study interviews indicated that participants considered the targeted features to be easy, elementary level forms; however, interviewees who mentioned that the features were easy to understand also admitted that they were difficult to use.

Participant   [The grammar] is easy.

Researcher   Was it easy to use or easy to understand or both?

Participant   Easy to understand.

Researcher   How about using it?

Participant   Ah, uhm…. (laughs, shakes head).

It is true that the secondary school English L2 curriculum focuses on more advanced grammatical features, if only receptively. If participants viewed receptive knowledge of language as being sufficient, then perceiving the targeted features as *too easy* might have prompted participants to be dismissive of the instruction and the CF.

The second way the data indicate possible influence of a lack of writing experience is that interviewees reported feeling surprise at the discovery that their work contained errors. To experts accustomed to analysing and correcting language-learning errors, this surprise might seem bewildering, as errors are generally accepted as part of language learning. However, the educational background of the participants in this study tends to be (necessarily) multiple-choice test-focused (Kwon, Lee, & Shin, 2017; Seth, 2002). Participants at the competitive, highly ranked university where the study took place likely excelled on such tests in their secondary schools. Being accustomed to scoring highly on multiple-choice tests of grammar, participants might fail to realise the gap between the ability to identify the correct grammatical feature in a provided list and the ability to use that same grammatical feature productively and spontaneously. Awareness-raising activities could help learners identify the gap between their receptive and productive knowledge.

Further contributing to difficulty participants might have accurately gauging their own ability could be a lack of interaction with L2 users of participants' L1 (Korean). It seems encountering speakers of Korean as a second language might be quite rare. Anecdotally,

people who do not appear to be of East Asian descent report being stared at, pointed at, and receiving shocked reactions when out in public in Korea, which the Korean government has tried to address (KTO, 2012). This indicates the novelty of such encounters. Thus, while English L1 users may encounter English L2 speakers, thereby developing a tacit understanding that non-standard linguistic expressions by L2 speakers can occur, Korean L1 users in Korea might not have a sufficient amount of this experience. A non-expert's unawareness of their lack of knowledge can negatively impact on their ability to increase that knowledge (Kruger & Dunning, 1999). Given the participants' (a) probable lack of interaction with L2 users of Korean and corresponding lack of awareness of what language learning 'sounds like'; (b) prior English L2 educational focus on receptive and test-taking skills; and (c) English L2 proficiency level, the Dunning-Kruger framework might help explain participants' descriptions of the targeted features as 'easy', despite continuing to struggle to use them inaccurately, and the reported feelings of surprise about the presence of errors in their work.

Conclusions

While grammatical accuracy was the focus of this study, it was not the focus of the class. Nonetheless, while a singular focus on accuracy might be undesirable, the value of linguistic accuracy cannot be entirely dismissed, as L2 writers can experience institutional and societal pressures to write with linguistic accuracy. The results of this study indicate a complex relationship between CF and the development of L2 linguistic accuracy. Data on participants' L2 learning backgrounds and educational histories can offer contextual information about this development, providing researchers with nuanced understanding of study results. Though more research is needed, the findings of the current study indicate that while scaffolding can result in significant improvement, it might not provide additional benefit to some learners. Direct written CF seems to produce similar results in less time. Practitioners might be pleased that less-resource intensive CF may impact on grammatical accuracy as significantly as spending valuable class time scaffolding learners' use of CF. A further recommendation is that teachers could benefit from gathering data on learners' prior experiences in order to better understand learners' engagement in classroom practices. This could be particularly relevant to teachers in contexts where they might have less familiarity with the learners' L1 and culture, as is common in some regions of East Asia (Williams, C. 2017, p. 1–3).

Strengths and limitations

Ecological validity was a goal and a strength of the study: (a) it examines the development of linguistic accuracy over the course of a term, as in a typical classroom; (b) it provides instruction on multiple grammatical features and then focuses on the effectiveness of CF on those structures; (c) it uses new pieces of participant-directed writing in academic genre. However, there are corresponding limitations. While the lack of a true control group might be unproblematic from a pedagogical perspective, it limits contribution to theory and to the conclusions which can be drawn from the results of the analyses. Additionally, the study is limited in scope; it does not examine teacher and student behaviours or interactions, and it took place during just one university term with two CF treatments. Longer studies looking at more variables are needed to gauge language development. Furthermore, as the participants in this study were STEM majors from test-centric, science backgrounds, other populations might engage differently with the interventions.

Future research

There seemed to be a gap between participants' perceptions of their knowledge of grammatical features and their ability to use these features accurately. Future research might investigate ways of narrowing this gap, investigating how being unaware of one's lack of knowledge (Kruger & Dunning, 1999) interacts with CF provision. For example, research might engage participants in awareness-raising activities, such as getting them involved with L2 learners of the participants' L1. Future research might also explore effects of CF training for learners whose educational backgrounds have included little writing and feedback experience. The results of the current study highlight the need for increased research focus on contextual factors in CF studies, such as the influences of prior educational experiences on CF use. Research on training learners in these contexts to use CF could inform how much time would need to be spent, and what kind of pedagogical activities might be most beneficial.

insightful suggestions for improvement. I especially thank the two teachers who generously implemented this study and the students who volunteered their data for analysis.

References

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal, 78*(4), 465-483. doi:10.2307/328585

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*(04), 543-574. doi:10.1017/S0272263106060268

Author (2018). A quantitative investigation of effectiveness of corrective feedback. *Manuscript Submitted for Publication*

Author (2019). Metalinguistic reflections elicited via conferences or reflective worksheets: An examination of South Korean university STEM learners. *Manuscript in preparation*

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition, 35*(4), 689-725. doi:10.1017/S0272263113000429

Bickhard, M. (1992a). How does the environment affect the person? In L. Winegar, & J. Valsiner (Eds.), *Children's Development within Social Context, V.1: Metatheory and Theory* (pp. 63-92). New Jersey: Lawrence Erlbaum Associates.

Bickhard, M. (1992b). Scaffolding and self-scaffolding: Central aspects of development. In L. Winegar, & J. Valsiner (Eds.), *Children's Development within Social Context, V2: Research and Methodology* (pp. 33-52). New Jersey: Lawrence Erlbaum Associates.

Bickhard, M. (2005). Functional scaffolding and self-scaffolding. *New Ideas in Psychology, 23*(3), 166-173. doi:10.1016/j.newideapsych.2006.04.001

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102-118. doi:10.1016/j.jslw.2007.11.004

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*(4), 348-363. doi:10.1016/j.jslw.2012.09.006

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217. doi:10.1016/j.jslw.2010.10.002

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*(3), 409-431. doi:10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2009a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*(2), 193-214. doi:10.1093/applin/amp016

Bitchener, J., & Knoch, U. (2009b). The relative effectiveness of different types of direct written corrective feedback. *System, 37*(2), 322-329. doi:10.1016/j.system.2008.12.006

Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development* (Kindle ed.). Bristol: Multilingual Matters.

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*(3), 267-296. doi:10.1016/S1060-3743(03)00038-9

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (Seventh ed.). New York: Routledge.

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. doi:10.1016/j.system.2008.02.001

Esteban, N., & Roca de Larios, J. (2010). The use of models as a form of written corrective feedback to secondary school pupils of English. *International Journal of English Studies, 10*(2), 143-170.

Ferris, D. (1995). Teaching ESL composition students to become independent self- editors. *TESOL Journal, 4*(4), 18-22.

Ferris, D. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime …?). *Journal of Second Language Writing, 13*(1), 49-62. doi:10.1016/j.jslw.2004.04.005

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition, 32*(02), 181-201. doi:10.1017/S0272263109990490

Gaies, S. J. (1980). T-Unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly, 14*(1), 53-60. doi:10.2307/3586808

Goldstein, L. (2016). Making use of teacher written feedback. In R. Manchón, & P. K. Matsuda (Eds.), *Handbook of Second and Foreign Language Writing* (Electronic ed., pp. 407-430). Berlin: De Gruyter.

González-Becerra, I. (2017). Language learning among STEM students: motivational profile and attitudes of undergraduates in a UK institution. *The Language Learning Journal, 47(3).* 385-401. doi:10.1080/09571736.2017.1320421

Granott, N. (2005). Scaffolding dynamically toward change: Previous and new perspectives. *New Ideas in Psychology, 23*(3), 140-151. doi:10.1016/j.newideapsych.2006.07.002

Gravetter, F., & Forzano, L. (2016). *Research Methods for the Behavioral Sciences* (5th ed.). Stamford: Cengage Learning.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi:10.3102/003465430298487

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal, 80*(3), 287-308. doi:10.1111/j.1540-4781.1996.tb01612.x

Hyland, K. (2013). Student perceptions of hidden messages in teacher written feedback. *Studies in Educational Evaluation, 39*(3), 180-187. doi:10.1016/j.stueduc.2013.06.003

Ishikawa, M. (2013). Examining the effect of written languaging: the role of metanotes as a mediator of second language learning. *Language Awareness, 22*(3), 220-233. doi:10.1080/09658416.2012.683435

Ishikawa, M. (2018). Written languaging, learners' proficiency levels and L2 grammar learning. *System, 74*, 50-61. doi:10.1016/j.system.2018.02.017

Ishikawa, M., & Suzuki, W. (2016). The effect of written languaging on learning the hypothetical conditional in English. *System, 58*, 97-111. doi:10.1016/j.system.2016.02.008

Knouzi, I., Swain, M., Lapkin, S., & Brooks, L. (2010). Self- scaffolding mediated by languaging: microgenetic analysis of high and low performers. *International Journal of Applied Linguistics, 20*(1), 23-49. doi:10.1111/j.1473-4192.2009.00227.x

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology, 77*(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121

KTO. (2012). 한국방문의 해 위원회 - 숨기편. Retrieved from

http://youtu.be/tT60Gw7wkuA

Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in*

Education: Principles, Policy & Practice, 24(1), 60-77.
doi:10.1080/0969594X.2015.1074540

Lepper, M., Drake, M., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert
human tutors. In K. Hogan, & M. Pressley (Eds.), *Scaffolding Student Learning:
Instructional approaches and issues* (pp. 108-144). Cambridge, MA: Brookline
Books/Lumen Editions.

Li, S. (2009). The differential effects of implicit and explicit feedback on second language
(L2) learners at different proficiency levels. *Applied Language Learning, 19*(1,2), 53-78.

Lightbown, P., & Spada, N. (2013). *How Languages are Learned* (4th ed.). Oxford: Oxford
University Press.

Littleton, K. (2013). Adaptation and authority in scaffolding and teacher–student
relationships: Commentary on the Special Issue 'Conceptualising and grounding
scaffolding in complex educational contexts'. *Learning, Culture and Social Interaction,
2*(1), 52-55. doi:10.1016/j.lcsi.2013.01.003

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of
corrective feedback in L2 writing. *Journal of Second Language Writing, 30*, 66-81.
doi:10.1016/j.jslw.2015.08.011

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language
classrooms. *Language Teaching, 46*(1), 1-40. doi:10.1017/S0261444812000365

Mascolo, M. F. (2005). Change processes in development: The concept of coactive
scaffolding. *New Ideas in Psychology, 23*(3), 185-196.
doi:10.1016/j.newideapsych.2006.05.002

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33*(2), 135-158. doi:10.1111/j.1467-1770.1983.tb00532.x

Ney, J. (1966). Review of *Differences in Grammatical Structures Written at Three Grade Levels*. *Language Learning, 16*(3-4), 230-235. doi:10.1111/j.1467-1770.1966.tb00822.x

Ortega, L. (2012). Epilogue: Exploring L2 writing—SLA interfaces. *Journal of Second Language Writing, 21*(4), 404-415. doi:10.1016/j.jslw.2012.09.002

OUP. (2014). oxfordenglishtesting.com. Retrieved from https://www.oxfordenglishtesting.com/defaultmrarticle.aspx?id=3074

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1-12. doi:10.1207/s15326985ep4001_1

Rao, Z., & Liu, F. (2011). Effect of academic major on students' use of language learning strategies: a diary study in a Chinese context. *The Language Learning Journal, 39*(1), 43-55. doi:10.1080/09571731003653565

Reichelt, M. (2009). Situated writing practices in foreign language settings: The role of previous experience and instruction. In R. Manchón (Ed.), *Writing in Foreign Language Contexts* (pp. 183-206). Bristol: Multilingual Matters.

Rinnert, C., & Kobayashi, H. (2009). Situated writing practices in foreign language settings: The role of previous experience and instruction. In R. Manchón (Ed.), *Writing in Foreign Language Contexts* (pp. 23-48). Bristol: Multilingual Matters.

Ro, J. (2017) A journey from the classroom to the world of educational reform: A study of

    three Korean teachers' practitioner inquiry. *Asia Pacific Journal of Education,* 37 (1),

    pp. 28-41. doi:10.1080/02188791.2016.1142423

Ruegg, R. (2016). The effect of assessment of process after receiving teacher feedback.

    Assessment & Evaluation in Higher Education, 41(2), 199-212.

    doi:10.1080/02602938.2014.998168

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing

    revision task. *Studies in Second Language Acquisition, 29*(1), 67-100.

    doi:10.1017/S0272263107070039

Santos, M., López-Serrano, S., & Manchón, R. (2010). The differential effect of two types of

    direct written corrective feedback on noticing and uptake: Reformulation vs. error

    correction. *International Journal of English Studies, 10*(1), 131-154.

    doi:10.6018/ijes.10.1.114011

Seth, M. J. (2002). *Education Fever: Society, politics, and the pursuit of schooling in South*

    Korea. Honolulu: University of Hawaii Press.

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.

    doi:10.2307/40264353

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. doi:10.1111/lang.12029

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. doi:10.1016/j.jslw.2013.03.011

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. doi:10.1016/j.system.2015.01.006

Simard, D., French, L., & Fortier, V. (2007). Elicited metalinguistic reflection and second language learning: Is there a link? *System, 35*(4), 509-522. doi:10.1016/j.system.2007.06.004

Song, J. (2011). English as an official language in South Korea: Global English or social malady? *Language Problems & Language Planning, 35*(1), 33-55. doi:10.1075/lplp.35.1.03son

Stone, A. (1998). Should we salvage the scaffolding metaphor? *Journal of Learning Disabilities, 31*(4), 409-413. doi:10.1177/002221949803100411

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness, 17*(2), 95-114. doi:10.1080/09658410802146644

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition, 32*, 303-334. doi:10.1017/S0272263109990532

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning, 62*(4), 1110-1133. doi:10.1111/j.1467-9922.2012.00720.x

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced Language Learning: The Contribution of Halliday and Vygotsky* (pp. 95-108). London: Continuum International Publishing.

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal, 93*(1), 5-29. doi:10.1111/j.1540-4781.2009.00825.x

Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education, 26*(6), 605-614. doi:10.1080/02602930120093922

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327-369. doi:10.1111/j.1467-1770.1996.tb01238.x

Vardi, I. (2009). The relationship between feedback and change in tertiary student writing in the disciplines. *International Journal on Teaching and Learning in Higher Education, 20*(3), 350-361.

Vogt, W., Gardner, D. & Haeffele, L. (2012) *When to Use What Research Design.* London: The Guilford Press.

Williams, C. (2017). *Teaching English in East Asia: A teacher's guide to Chinese, Japanese, and Korean learners.* Singapore: Springer.

Williams, J. & Kane, D. (2009). Assessment and feedback: Institutional experiences of student feedback, 1996 to 2007. *Higher Education Quarterly, 63*(3), 264-286. doi:10.1111/j.1468-2273.2009.00430.x

WISET. (2015). *2015 Report on Korean Women in Science, Engineering and Technology.* (). Seoul - http://www.eng.wiset.or.kr: WISET: Women in Science, Engineering and Technology.

Wong, L. L. C., & Nunan, D. (2011). The learning styles and strategies of effective language learners. *System, 39*(2), 144-163. doi:10.1016/j.system.2011.05.004

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi:10.1111/j.1469-7610.1976.tb00381.x