English Teaching, Vol. 78, No. 1, Spring 2023, pp. 225-245
DOI: https://doi.org/10.15858/engtea.78.1.202303.225
http://journal.kate.or.kr

The Effects of Languaging in Response to Automated Written Correction for Grammar Accuracy Improvement

Wenwen Xu and Ji-Hyun Kim *

Xu, Wenwen, & Kim, Ji-Hyun. (2023). The effects of languaging in response to automated written correction for grammar accuracy improvement. *English Education, 78*(1), 225-245.

This study explored the role of written languaging (WL) in response to automated written corrective feedback (AWCF) in L2 accuracy improvement in English classrooms at a university in China. A total of 254 freshmen enrolled in intermediate composition classes participated, and they wrote 4 essays and received AWCF. A half of them engaged in WL after receiving correction, while the other half did not. To measure whether WL could contribute to L2 accuracy improvement, error correction tests were conducted three times. AWCF targeted all types of errors, but the study focused on investigating six language features (nouns, conjunctions, subject-verb agreement, sentence constituents, articles, and prepositions) known as the most frequent errors intermediate students make. The results proved that WL had a positive effect on accuracy improvement overall. However, when individual language features were considered separately, the effects were not the same. The inherent complexity of the features along with L1 influence appeared to affect the effects of WL for accuracy improvement.

Key words: written languaging, automated written corrective feedback, accuracy improvement

*First Author: Wenwen Xu, Instructor, Anhui University of Finance & Economics, China; Email: fruitwenwen@163.com

Corresponding Author: Ji-Hyun Kim, Professor, Department of English Education, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu; Email: jhk2024@kmu.ac.kr

Received 31 December 2022; Reviewed 15 January 2023; Accepted 15 March 2023

1. INTRODUCTION

'Languaging' may sound unfamiliar, but it is a day-to-day practice. We often find ourselves talking aloud when we need to clarify a problem. This practice often helps us process our thoughts into object for further reflection. The notion of 'languaging' is rooted in Vygotsky's sociocultural theory. According to sociocultural theory, language can be used as a cognitive tool to regulate learners' thinking and to mediate their mental activities (Lantolf & Thorne, 2006; Swain, 2006). Informed by Vygotsky's sociocultural theory, Swain (2006) coins the term 'languaging' and defines it as "the process of making meaning and shaping knowledge and experience through language" (p. 89) and "a dynamic, never-ending process of using language to make meaning" (p. 96). In other words, languaging is the act of producing language in order to think through a problem. According to Swain (2006), learners benefit from both the process of languaging (i.e., externalizing their thoughts) and the product of languaging (i.e., reflecting on them). Languaging seems to produce an ideal condition for noticing (Schmidt, 2001) and reflection, which are two functions of output hypothesized to facilitate L2 development (Swain, 2000).

Languaging can be practiced using two different modes, oral languaging (OL) and written languaging (WL). In the case of OL, it has been widely explored in L2 studies with the name of 'metatalk' (Storch, 2008; Swain, 1998), 'verbalization' (Swain, 2000), 'collaborative dialogue' (Swain & Lapkin, 1998), or 'private speech' (Ohta, 2001). These studies have proved that OL has a positive effect on processing L2 information and development. In contrast, the effect of WL on L2 learning is less explored. Compared to speaking, writing usually provides learners with more time and allows them to engage in a "much slower, repeated mediating process of analysis and synthesis" (Luria, 1999, p. 103). Put differently, writing offers learners more time to reflect on their output; therefore, they are also more likely to consider their thoughts, and this may accordingly generate optimal circumstances for learning.

Considering the benefits of writing over speaking, it can be hypothesized that WL and OL are comparable, or WL can be even more beneficial. Although some L2 studies in WL (Suzuki, 2012, 2017) have been conducted, the number of studies is still scant. Furthermore, they mainly focused on the role WL plays in learners' revision process after receiving teachers' indirect or direct written corrective feedback (Kim, 2021; Niu & You, 2020). It has been pointed out that the correction L2 learners produce during revision processes could be interpreted as a sign of L2 improvement, but it is not equivalent to L2 accuracy development and learning (Truscott, 1996; Truscott & Hsu, 2008).

The present study aimed to expand and fill the gaps in the previous studies of WL in L2 learning by complementing the methodological limitations related to a tool of accuracy measurement, length of research, pedagogical relevance, and so on. In addition, the study

investigated the role of WL in response to automated written corrective feedback (AWCF), which has been increasingly used in a large scale of L2 writing classrooms (Ranalli, 2018).

2. REVIEW OF THE LITERATURE

2.1. Written Languaging, Corrective Feedback, and L2 Learning

The contribution of written corrective feedback (WCF) to L2 learning has long been discussed and investigated from various perspectives in the field of L2 research. One of the research strands which attracted L2 researchers is the exploration of the process whereby L2 learners understand WCF. Languaging, especially oral languaging, has been widely employed to examine and/or trigger L2 learners' process of WCF (Swain & Lapkin, 2002). Although WL has been relatively less explored in L2 WCF research compared to OL, there have been some WCF studies focusing on the role of WL. The studies in WL in response to teacher WCF can be largely classified into two types according to whether they examined direct WCF or indirect WCF.

For instance, Suzuki (2012) investigated the effectiveness of WL in response to direct WCF (i.e., indicating the location of errors and giving the correct forms) on developing writing accuracy over text revision tasks. The study took a three-stage writing task (writing an original draft, receiving feedback with languaging, and revising the original draft). Suzuki described the purpose of the study: it was "of an exploratory nature, rather than a carefully controlled experimental study," (Suzuki, 2012, p. 1128). The study reported that conducting WL was facilitative, helping learners successfully revise errors in the immediate writing revision stage. However, because Suzuki did not include a control group to distinguish the effectiveness of languaging from the effect of WCF, the study fell short of claiming the genuine effects of WL.

To have a clear understanding of the effects of WL on the grammatical accuracy of writing, Moradian, Miri, and Hossein-Nasab (2017) improved Suzuki's (2012) study by adding a control group. In the experiment, both groups took a three-step writing process (writing a draft, receiving feedback with/without languaging, and revising the draft). When the WL group verbalized the reasons behind their errors, the control group only read their writings and counted the errors. The original writing was used as the pretest, and the revision text was used as the posttest. The WL group outperformed the control group, suggesting that WL had a favorable impact on grammatical accuracy in writing.

While the facilitative role of WL in response to direct WCF in L2 learning has been reported, the studies into WL and indirect WCF resulted in complicated outcomes. Suzuki (2009) investigated the effects of WL in response to indirect WCF (underlining) on the

improvement of L2 writing accuracy. Twenty-four Japanese EFL learners engaged in a three-stage-writing task. Data analyses showed that the average number of errors notably reduced from the first draft to its revision, and 'languaged' errors were revised more often than non-languaged errors. This result proved the effect of WL. However, Suzuki (2009) failed to distinguish the effectiveness of WL from one of indirect WCF. The study did not provide clear evidence in support of the effects of WL on L2 development.

Acknowledging this limitation, Niu and You (2020) separated the effect of WL from one of indirect WCF. In order to explore the effect of WL, the study compared a group that received both indirect WCF and WL with a group that received indirect WCF only. The study analyzed learners' writing and revisions, respectively used as pre-tests and posttests, to measure accuracy improvement. The analyses revealed that both groups showed a significant improvement in accuracy measured by immediate and delayed revisions, and found no difference between the groups. In other words, they did not find any specific effect of WL on improving accuracy.

Fukuta, Tamura, and Kawaguchi (2019) reported a more complicated result. The study explored whether indirect WCF enhanced L2 learners' linguistic gains (complexity, accuracy, and fluency) through WL. The cases where leaners received indirect WCF with WL and the cases where they only engaged in WL without indirect WCF were compared. Forty learners engaged in a three-stage writing task (writing a first draft, revision with/without feedback, and writing on the second new writing). It was found that WL was effective in improving fluency and accuracy, but not complexity, regardless of the existence of indirect WCF. They also noted that "indirect feedback drew learners' attention to grammar and induced successful error correction, but the directed attention did not necessarily lead to any improvement" (p. 10). In other words, learners were able to correct their own errors using metalinguistic knowledge in the revision stage, but they could not produce grammatically correct sentences in subsequent writing.

Fukuta et al. (2019) has made a vital point, the limitation of measuring accuracy improvement through the revision process. All of the aforementioned studies have examined the effects of WL on accuracy improvement while they engage in revisions after they received WCF. The corrections learners make in response to WCF can be a first step of L2 learning (i.e., noticing of errors or gaps). However, the ability to correct the errors while they are revising cannot offer sufficient evidence that WCF has a lasting effect beyond the revision stage; in other words, it cannot prove that L2 learning has taken place (Guénette, 2007; Truscott, 1999).

As a matter of fact, the accuracy improvement reported in WCF studies has been controversial. As noted above, many of the WCF studies considered accuracy gains reported in revised texts. Even the studies which examined accuracy improvement through a more robust tool such as a test or other than revision task investigated focused WCF: WCF only

targeted certain linguistic structures such as the acquisition of the English article and the simple past tense (Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008). These studies have been criticized for their lack of pedagogical relevance because teacher WCF in L2 classrooms targets various errors, not just one error. Thus, it is questionable whether the efficacy reported in these studies can be generalized into real L2 classroom contexts. By acknowledging the limitations in the previous WCF and WL studies, the present study explored the efficacy of WL when learners received WCF (more specifically automated WCF) targeting a wide range of linguistic features in real L2 classrooms. In addition, learners' accuracy improvement of the features was measured by a test (not via a revision text), which will be explained in detail in the Methods section.

2.2. Automated Written Corrective Feedback (AWCF)

AWCF refers to written corrective feedback given by automated writing evaluation tools. With the help of artificial intelligence, natural language processing, and statistical methods, AWCF provides students with both immediate essay ratings and immediate automated feedback, vastly accelerating the practice-feedback loop (Kellogg, Whiteford, & Quinlan, 2010). The pros and cons of AWCF have been pointed out. With regard to the advantageous features, first, it detects a variety of linguistic errors and provides real-time feedback on them, bringing errors to students' attention and encouraging them to revise quickly (Jeon & Kaya, 2006; Lee, 2017). Second, it contains consistent metalinguistic explanations that help students deal with errors (Woodworth & Barkaoui, 2020). However, AWCF lacks in pedagogical considerations since what needs to be corrected is determined by technological capacities rather learners' psychological and linguistical readiness for accepting corrective feedback. In other words, the one-size-fits-all nature of AWCF may render correction ineffective.

Then, the question is whether AWCF actually contributes to L2 learning. Research into the effect of AWCF on L2 writing has yielded conflicting findings. Some studies reported that AWCF helped accuracy improvements in revision texts and learners had a positive perception for AWCF (Grimes & Warschauer, 2010; Li, Link, & Hegelheimer, 2015). In contrast, other studies reported that learners did not use the revision functions (Attali, 2004), and even when they revised initial texts, about half of AWCF were ignored in their revision (Chapelle, Cotos, & Lee, 2015). Chen and Cheng (2008) and Lai (2010) also discovered that learners considered the feedback too general and impersonal, so they often ignored AWCF.

Although there have been some studies in AWCF, it has not been extensively investigated. Therefore, it might be too early to reach any conclusion about its efficacy in the L2 writing class. However, based on what has been found thus far, its main weakness is that learners often ignore corrections AWCF offers. There might be many different reasons why they do

not attend to the corrections and metalinguistic information AWCF provides. The current study hypothesized that if learners were required to reflect and think over the metalinguistic information regarding the errors they had made, the effect of AWCF would be enhanced and eventually contribute to L2 learning. In this sense, WL would lead learners to pay attention to the information given in AWCF, and this would be attributed to L2 accuracy improvement. Thus, the current study aimed to explore the following research questions:

1) Is written languaging in response to automated written corrective feedback effective for grammar accuracy improvement?
2) If so, does the effect sustained across different structures vary?

3. METHODOLOGY

3.1. Participants

A total of 254 freshmen in four intact classes at X University in China participated in the study. They were taught by different teachers but followed the same curriculum offered by the university. They were English learners majoring in finance and economics and placed into the intermediate level on the basis of their English scores in the National College Entrance Examination organized by the Chinese National Education Examinations Authority. Their mean score ranged from 80 to 145 (out of 150), with an average of 110.70. The classes were randomly assigned to two (+) WL groups or two (-) WL groups.

3.2. Writing Tasks

The classes required that students wrote six essays throughout the semester, and AWCF were used as a tool for error correction. In order to provide students with the time to be familiarized for essay writing tasks and AWCF, the current study included the data from the third to the sixth essays (a total of 4 essays). The writing class in X University did not include a revision process, which is quite common in the Chinese college English teaching context. Thus, unlike the typical process-oriented essay writing procedures which includes the revision process (drafting-feedback-revision), the classes participating in the study consisted of two phases (drafting-feedback).

3.3. Automated Written Corrective Feedback (AWCF)

AWCF in the present study was provided by *Pigai* (*http://www.pigai.org/*). *Pigai* is

designed for Chinese English learners and feedback is offered in Chinese. *Pigai* provides a holistic score ranging from 1 to 100, general feedback, and sentence-based corrective feedback. The holistic score is given by comparing the submitted text's quantitative differences (vocabulary, sentence, structure and organization, and content relevance) with texts of standard English in its corpus, consisting of students' English essays and English textbooks. General feedback addresses vocabulary, sentences, structure and organization, and content relevance, accompanied by a bar graph indicating the relative strength of a submitted essay in each of these areas. Sentence-based feedback includes corrective feedback on punctuation, spelling, grammar, word choices, collocations, and so on. *Pigai* offers metalinguistic explanations, pointing out an error without providing a correct form. *Pigai* provides unfocused feedback, and some of the examples of corrections collected in the present study are presented in Table 1.

TABLE 1

Examples of Corrective Feedback Given by *Pigai*

Error Type	Example	Feedback Given by *Pigai*
Noun Errors	Different individual hold different view.	Please check *view*; usually the plural form is used here.
Article Errors	Secondly, you can do what you like in a small town, don't have to start at a bottom of a big company.	Please check *a*, and confirm the right article is used.
Subject-verb Disagreement	Secondly, competition in small cities are less than those in big cities.	Please check *are* and confirm whether the verb agrees with the subject.

3.4. Written Languaging (WL)

As noted previously, languaging refers to the act of producing language in the form of speaking and writing in order to think through a problem (Swain, 2006). In the current study, WL was operationalized as the students' production in the form of written language while they were dealing with the corrections made to their errors, which would lead them to go through a series of cognitive processes.

The students in the (+) WL group received a paper and were instructed to write down what they thought about the errors and the given metalinguistic feedback (Suzuki, 2009). WL was conducted in L1 to prevent the case where English proficiency affected their ability to verbalize their thoughts. They were given 20 minutes for WL and were allowed to write "I don't have any idea" when they had nothing to say. Then, the papers were collected. A training session was conducted before the data were collected to help them feel comfortable with the process of WL.

3.5. Target Features

Pigai provided corrections to a wide range of errors, and the students were asked to do WL for all the corrections. However, the present study focused on examining the accuracy improvement of six language features for the following reasons. First, these six target structures were very difficult and accounted for most of the errors in writing in the Report on English Writing Ability of Chinese College Students 2020. Furthermore, up to date, published written corrective feedback (WCF) studies have mainly focused on one or two target structures, such as definite and indefinite articles (Bitchener, 2008; Sheen, 2007), the simple past tense and prepositions (Bitchener, Young, & Cameron, 2005). These studies provided a better understanding of the impact of WCF on the mentioned structures but could not offer evidence for the overall effect on a wider category of linguistic errors. However, the studies were conducted in different research contexts using different assessment tools, so it's impossible to compare the effects of WCF across different grammar structures. To fill the research gap, the present study chose six target grammar structures.

The study employed a pretest-posttest-delayed-posttest design to prove the effects of WL on L2 accuracy improvement. In order to generate tests, it was necessary to decide language features in advance. Therefore, the study chose six language features, articles, subject-verb agreement (S-V agreement), nouns, conjunctions, sentence constituents and prepositions because these six structures have been considered the most difficult English language features according to the Report on English Writing Ability of Chinese College Student 2020. Furthermore, the six linguistic structures account for 81.13% of errors made by students enrolled in intermediate writing classes at X university according to the data from *Pigai*'s writing corpus, consisting of 681 essays from 427 intermediate learners.

3.6. Error Correction Test (ECT)

ECT was used to measure participants' grammar accuracy improvement. Participants were required to determine whether each sentence was grammatically correct and, if not, to make the needed modifications. ECTs in the present study were developed by the researchers adapted from Ellis (2005). After developing the test items, the items were reviewed by language teachers with more than fifteen years of English teaching experience at X University. According to their advice, changes were made to make sure whether the tests were proper for the participants. Care was especially taken in the vocabulary choices offered on the test to avoid hindering the students from answering the questions due to unknown or difficult words. Finally, the tests were piloted before being employed in the present study, and only ones which were statistically satisfactory were retained.

The ECT was scored on a discrete item basis. For each item, the learners were asked to

locate an error and correct it (see Appendix for examples). The students received 2 points when they successfully located an error and corrected the error. When they located an error but failed to correct the error, they received 1 point. Three versions of ECTs (ECT 1, ECT 2, and ECT 3) were used for a pretest, a posttest, and a delayed posttest. The three tests examined the same grammatical structures and were comparable in terms of length. But vocabulary differed. The internal consistency reliability coefficients for the three versions were found to be acceptable (Cronbach's alpha: .78, .80, .74).

3.7. Procedures

The study employed a pretest-posttest-delayed posttest design and conducted in participants' regular classrooms. The study lasted 14 weeks (see Table 2 for a detailed schedule).

TABLE 2

Sequence of the Research Procedure

Week	(+) WL Group	(-) WL Group
Week 1	Background questionnaire + ECT 1 (Pretest) WL orientation	Background questionnaire ECT 1 (Pretest)
Week 2	Essay 1 Reflecting on AWCF + conducting WL	Essay 1 Reflecting on AWCF
Week 4	Essay 2 Reflecting on AWCF + conducting WL	Essay 2 Reflecting on AWCF
Week 6	Essay 3 Reflecting on AWCF + conducting WL	Essay 3 Reflecting on AWCF
Week 8	Essay 4 Reflecting on AWCF + conducting WL	Essay 4 Reflecting on AWCF
Week 9	ECT 2 (posttest 1)	ECT 2 (posttest 1)
Week 14	ECT 3 (posttest 2)	ECT 3 (posttest 2)

In Week 1, the students were first given 5 minutes to fill out a background questionnaire administered in Chinese. Then, they took Error Correction Test (ECT) 1 as a pretest (20 mins.) through a Chinese online learning application named *Xuexitong*. Then, the (+) WL group received a 20-minute orientation for WL. In order to help the (+) WL group understand WL process, the teachers explained WL and offered several demonstrations. In Week 2, the students wrote an essay in response to a writing prompt. The essay length was restricted to 120-180 words. When students finished the writing task, they submitted essays to *Pigai*'s AWCF program. While the (+) WL group engaged in WL, the (-) WL group was asked to look through the feedback they received. The teachers did not direct or specify the participants' focus in the languaging procedures. WL papers were collected by teachers. The

students repeated the same procedures in Week 4, Week 6, and Week 8. In Week 10, the first posttest was employed followed by the second posttest in Week 14.

4. RESULT

4.1. The Effects of WL on Grammar Accuracy Improvement

Table 3 shows the results of the descriptive statistics of both groups' ECTs. Judging from the mean scores, both groups improved from the pretest to the delayed posttest. A repeated measures ANOVA confirmed that the improvement exhibited in both groups is significant: for the (+) WL group, F (2, 240) = 26.86, p < .001, η^2 = .16), and for the (-) WL group, F (2, 264) = 9.68, p < .001 η^2 = .07.

In order to examine the group difference, first, a t-test was performed on the pretest. The test showed that the (+) WL outperformed the (-) WL in the pretest (t = 3.32, p <.001), and this suggested that the two groups were not comparable. Therefore, a one-way analysis of covariance (ANCOVA) was conducted to measure the accuracy improvement the two groups showed in the posttest 1 (ECT 2) while controlling for the difference that existed in the pretest, ECT 1. The result showed that there was a significant difference in ECT 2, F (1, 251) = 5.20, p = .02, η^2 = .02. Following the similar procedure, another ANCOVA was conducted to examine whether the two groups differed in ECT 3, and it reported a significant group difference in ECT 3, F (1, 251) = 13.08, p < .001, η^2 = .05. This suggests that both groups improved in grammar accuracy, but the (+) WL group's scores improved more than the (-) WL group.

TABLE 3

Descriptive ECTs Scores of ECTs

Group	Test	*M*	*SD*	95% *CI*
(+) WL (n = 121)	ECT 1	37.72	6.47	[36.56, 38.88]
	ECT 2	39.17	5.98	[38.10, 40.25]
	ECT 3	41.20	5.96	[40.12, 42.26]
(-) WL (n = 133)	ECT 1	34.80	7.46	[33.52, 36.08]
	ECT 2	36.05	7.44	[34.77, 37.32]
	ECT 3	37.77	9.42	[36.16, 39.39]

4.2. The Effects of WL for Different Language Forms

The study further explored whether WL had a different effect on the accuracy improvement for each target structure. Table 4 shows the descriptive statistics of ECT scores,

and it shows mixed findings. In the case of nouns and conjunctions, both groups' ECT scores increased over time. However, in the case of S-V agreement and prepositions, both groups' scores increased in ECT 2, but the improvement was not sustained in ECT 3. The cases of sentence constituents and articles showed complicated results: both groups' scores decreased in ECT 2 but bounced back in ECT 3.

TABLE 3

Descriptive Statistics of Six Target Structures' Accuracy

Structure	Group	Test	*M*	*SD*	95% *CI*
Nouns	(+) WL	ECT 1	1.75	0.26	[1.71, 1.80]
		ECT 2	1.90	0.21	[1.86, 1.93]
		ECT 3	1.92	0.21	[1.88, 1.96]
	(-) WL	ECT 1	1.68	0.34	[1.62, 1.74]
		ECT 2	1.81	0.32	[1.76, 1.87]
		ECT 3	1.83	0.28	[1.78, 1.88]
S-V agreement	(+) WL	ECT 1	1.72	0.40	[1.64, 1.79]
		ECT 2	1.80	0.32	[1.74, 1.85]
		ECT 3	1.69	0.41	[1.61, 1.76]
	(-) WL	ECT 1	1.55	0.50	[1.46, 1.64]
		ECT 2	1.65	0.46	[1.57, 1.73]
		ECT 3	1.55	0.49	[1.47, 1.63]
Conjunctions	(+) WL	ECT 1	1.19	0.55	[1.09, 1.29]
		ECT 2	1.55	0.44	[1.47, 1.62]
		ECT 3	1.61	0.45	[1.53, 1.69]
	(-) WL	ECT 1	1.07	0.56	[0.97, 1.17]
		ECT 2	1.26	0.47	[1.18, 1.34]
		ECT 3	1.31	0.50	[1.23, 1.40]
Sentence Constituents	(+) WL	ECT 1	1.86	0.25	[1.82, 1.91]
		ECT 2	1.66	0.33	[1.60, 1.72]
		ECT 3	1.86	0.29	[1.80, 1.91]
	(-) WL	ECT 1	1.77	0.37	[1.71, 1.83]
		ECT 2	1.70	0.40	[1.63, 1.76]
		ECT 3	1.76	0.35	[1.70, 1.82]
Prepositions	(+) WL	ECT 1	1.35	0.56	[1.25, 1.45]
		ECT 2	1.38	0.49	[1.29, 1.47]
		ECT 3	1.34	0.50	[1.25, 1.43]
	(-) WL	ECT 1	1.17	0.52	[1.08, 1.25]
		ECT 2	1.27	0.53	[1.18, 1.36]
		ECT 3	1.17	0.51	[1.08, 1.25]
Articles	(+) WL	ECT 1	1.25	0.57	[1.15, 1.36]
		ECT 2	1.13	0.48	[1.04, 1.22]
		ECT 3	1.46	0.45	[1.38, 1.54]
	(-) WL	ECT 1	1.20	0.50	[1.11, 1.28]
		ECT 2	1.02	0.56	[0.92, 1.11]
		ECT 3	1.30	0.54	[1.20, 1.39]

In order to examine whether the accuracy improvement found in nouns and conjunctions

was statistically significant or not, one-way repeated ANOVAs and post-hoc analyses were conducted. The analyses showed that the accuracy gains were significant in both groups: for nouns, F (2, 240) = 23.67, p < .001 in the (+) WL; F (2, 264) = 14.06, p < .001 in the (-) WL, and for conjunctions, F (2, 240) = 45.17, p < .001 in the (+) WL; F (2, 264) = 12.38, p < .001 in the (-) WL. A following ANCOVA showed that the (+) WL group gained more accuracy than the (-) WL group in both structures in ECT 2 and ECT 3: for nouns, F (1, 251) = 4.44, p = .04, η^2 = .02 in ECT 2; F (1, 251) = 4.84, p = .03, η^2 = .02 in ECT 3, and for conjunctions, F (1, 251) = 22.02, p < .001, η^2 = .08 in ECT 2; F (1, 251) = 21.96, p < .001, η^2 = .08 in ECT 3. These results revealed that WL in response to AWCF benefited accuracy improvement in the case of nouns and conjunctions

For the cases of S-V agreement and prepositions, one-way repeated ANOVAs and post-hoc analyses showed that the gains observed in ECT 2 in both groups were not significant: for S-V agreement, F (2, 240) = 2.98, p = .05 in the (+) WL; F (2, 264) = 2.23, p = .11 in the (-) WL. Since the analyses indicated that AWCF did not contribute to the accuracy gains of these two structures in both groups, a group comparison was not meaningful. Thus, no further analyses were conducted.

As noted, in the case of sentence constituents and articles, both groups' scores decreased in ECT 2 but bounced back in ECT 3. However, they showed a different pattern. In the case of sentence constituents, in both groups, the scores in ECT 3 were almost equal to ECT 1. While in the case of articles, the scores of ECT 3 in both groups were greater than ECT 1. In other words, there was no accuracy improvement between ECT 1 and ECT 3 in sentence constituents, but there was in articles. Thus, in order to examine whether or not such improvement was significant for articles, a one-way repeated ANOVA and a post-hoc analysis were carried out. The analyses reported a significant improvement in both groups, and the improvement occurred between ECT 1 and ECT 3: F (2, 240) = 16.16, p < .001 in the (+) WL; F (2, 264) = 16.26, p < .001 in the (-) WL. In order to examine whether or not the gains of both groups in ECT 3 differed, an ANCOVA was conducted. The outcome showed the group difference, suggesting that the (+) WL group gained more accuracy than the (-) WL group (F (1, 251) = 5.95, p = .02, η^2 = .02). Effect size for the ANCOVAs were calculated as eta-squared (η^2) with values of .01, .06, and .14 indicating small, moderate and large effects, respectively (Cohen, 1988).

5. DISCUSSION

5.1. The Effects of WL on Grammar Accuracy Improvement

The study found that the students in both groups improved the accuracy of the target

features when they were considered together. Due to the absence of a control group that did not receive AWCF, it is not feasible to argue that their improvement was solely attributed to AWCF. However, the finding that the (+) WL group improved more accuracy than the (-) WL group suggests that WL played a facilitative role in accuracy improvement. This result is consistent with previous studies in WL (Ishikawa, 2018; Ishikawa & Suzuki, 2016). The facilitative effect may be explained by drawing on the following theoretical claims.

First, WL entails writing down the reflections and producing output, and this serves as an attention-getting device that facilitates noticing (Ishikawa & Révész, 2020; Swain & Lapkin, 1995). In the current study, the (+) WL group had a chance to verbalize their reflections in response to AWCF, and this process might enable them to pay more attention to the AWCF compared to the (-) WL group. Furthermore, WL process might cause a deeper level of processing of AWCF. As Craik and Lockhart (1972) pointed out, WL can elicit profounder processing, resulting in stronger long term memory representations. In a similar vein, Moradian, Hossein-Nasab, and Miri (2020) asserted that WL can enhance learners' awareness level to notice and transform this into comprehension. In the present study, WL in response to AWCF might encourage learners to process the AWCF more deeply, helping them improve their awareness level from noticing the feedback to understanding the feedback.

Second, engaging in WL in response to AWCF might result in a generation effect (Slamecka & Graf, 1978). The generation effect takes place when learners remember information produced by themselves better than information that they receive from an outside source. In the current study, WL encouraged the learners to produce information by themselves, and this process might have helped them keep in mind their reflection (Ishikawa & Révész, 2020).

Lastly, the students may have used their WL as an instrument of thought in locating and formulating a solution to an issue (Vygotsky, 1986). Although it is assumed that the students in the (-) WL group might conduct languaging silently in their mind, the better outcomes for the (+) WL group suggest that externalizing one's thoughts through WL could accelerate the development of grammar accuracy.

5.2. Differential Effects of WL for Different Language Forms

The first research finding seems to confirm our assumption that WL might play a positive role in enhancing learner attention to metalinguistic information in AWCF, and this brought about accuracy improvement. However, when the effect of WL on each individual target feature was examined, it turned out that the results were more complicated.

For nouns and conjunctions, WL contributed to accuracy improvement. In contrast, for S-V agreement, sentence constituent, and prepositions, neither (+) WL nor (-) WL groups show

any significant accuracy improvement, and this suggests WL did not have an impact. What seems arguable is why WL did not help the students process AWCF for these three language features when it did for nouns and conjunctions. The complexity inherent in these three structures could be a factor. In addition, L1 influence could make it more difficult for the students to process and digest the information in AWCF although they had enough time and chance to reflect their errors and corrections.

The accurate use of S-V agreement and sentence constituent require understanding the complex relations among several constituent parts such as person, gender, verb tense, and phrases. According to Jarvie (1993), S-V agreement is a rule that ensures the harmonizing of different grammatical units. Truscott (1996) also claims that syntactic structure is a complicated system in which knowledge of one component is connected to knowledge of other components; therefore, errors related to these structures may be more challenging to correct. This complexity might hinder the students from fully processing the corrections (i.e., metalinguistic information).

Furthermore, the students' L1 (Chinese) seems to make it more difficult to grasp the metalinguistic information. In Chinese, since verbs do not take morphological inflections to represent the person, gender, number (singular or plural), or time. S-V agreement, using morphological inflections is definitely difficult for Chinese learners of English (Chen, 2011; Mohamed, Lian, & Eliza, 2004; Tseng & Liou, 2006). In addition, Chinese has topic prominence construction while English has subject prominence construction. Influenced by topic prominence construction in Chinese, Chinese learners of English often make sentences like the followings:

> **This kind of article*, he think it is hard to write.
> **Going to America for study*, the China government has regulation early.
> **There are five apples* include three bad apples.
> (Chen, 2011; asterisk (*) refers to ungrammaticality.)

In these sentences, a topic in the form of a noun phrase, a verb phrase, and a clause serves a subject. Furthermore, sentences omitting subjects are acceptable and frequently used because of the relatively weak subject consciousness. The differences in L1-L2 sentence constituent rules might even increase the complexity level of the language forms, S-V agreement, and sentence constituent.

In addition, because prepositions are not discrete items, they are related to other constituents (Jarvie, 1993). For instances, words like *on, at, to* and *before* show the relation of a noun or noun equivalent to the rest of a sentence; thus, these are very complicated language forms to master even for advanced learners of English. In addition, because Chinese has no frame of reference for English prepositions, it feels more challenging to learn

English prepositions. Richards (1974) notes that students often use what they have learned for new contexts by drawing analogies, which is a primary cause of prepositional errors. For instance, after learning, *on foot* and *stand on one's own (two) feet*, they may have linked *on* with *foot/feet*. Thus, they make a sentence like **I jumped on my feet*, which seems to be correct to them. In addition, Chinese learners of English tend to overuse prepositions by inserting a preposition one where it is not required, such as **My mother was comforting at him*. Most Chinese structures do not require preposition usages, so learners rely on English structures as a guide. After being exposed to the sentences like *My mother was looking at him* and *My mother was shouting at him*, *My mother was comforting him* may seem incorrect for them. Thus, they insert the preposition after the verb, *comforting at him*. In contrast, learners sometimes omit prepositions. After learning that a verb such as *saw* does not require a preposition (e.g., I saw an accident), learners compose sentences such as, *She heard the accident*. The L1-L2 difference may have heightened the intricacy of prepositions, and WL does not seem effective enough to help the students reconstruct their knowledge and understanding of the usage of prepositions.

In the case of articles, both (+) WL and (-) WL groups did not show any improvement in the posttest 1, but the (+) WL group showed a significant improvement between the pretest and the posttest 2 while the (-) WL did not. First, it is noteworthy that WL in response to AWCF resulted in accuracy improvement of articles in the posttest 2. Like prepositions, articles are considered one of the most difficult language forms for learners of English, especially, when their L1 does not have an equivalent concept to or a system different from English articles (Li, 2009). Considering the fact that Chinese does not have the article system, the finding that the (+) WL group improved its accuracy scores in the posttest 2, seems promising for the positive role of WL. However, two questions need to be considered in this regard. The first is why the students improved accuracy in articles but not in prepositions; the second is why the improvement was observed in the posttest 2 and not in the posttest 1. The use of articles requires learners to consider many different factors including referentiality (i.e., specific references and hearer's knowledge), idiomatic expressions, and conventional uses (Huebner, 1983). Although grasping the uses related to idiomatic/conventional expressions and exceptional cases is complicated for learners, some article usages regulated by certain rules (e.g., the difference between *a* and *the*) seem straightforward and less complicated. However, the use of prepositions is mostly based on semantic. Thus, explicit metalinguistic information may not be effective for learners to develop an understanding of the use of prepositions. In the current study, all of the students had learned English in the rule-oriented traditional class. In other words, they might have a certain degree of declarative knowledge of English articles, especially rule-based parts, and such prior knowledge might enable learners to use the opportunity of WL effectively. However, as articles are still complicated for the students, it seemed to take more time for

them to internalize the rule and to reconstruct their interlanguage, which partly accounted for the reason why they showed improvement in the posttest 2 not in the posttest 1.

6. CONCLUSION

The present study investigated the effect of WL in response to AWCF on grammar accuracy improvement, focusing on the six language forms (i.e., nouns, conjunctions, S-V agreement, sentence constituent, prepositions, and articles). When the forms were measured as a whole, both (+) WL and (-) WL groups gained grammar accuracy improvement and sustained the improvement, but the (+) WL group improved more than the (-) WL group. However, when the six structures were measured individually, the effects varied. WL was facilitative for three structures, nouns, conjunctions, and articles, but not for prepositions, sentence constituent, and S-V agreement. The result indicates that WL in response to AWCF is effective in grammar accuracy improvement, but it does not work for all language forms.

Based on the findings of this study, a few pedagogical suggestions can be made. First, the study proved that WL could be used as a facilitative tool that triggers learner attention to metalinguistic information in AWCF. Instead of teacher correction, AWCF has been used in large L2 writing class for practical reasons. However, AWCF is often neglected because students do not feel compelled to pay attention to it, especially, when revision is not required in the class. In this context, using WL in response to AWCF might be a good solution for students to process AWCF, which may consequently result in accuracy improvement. Second, nevertheless, it should be kept in mind that WL seems to have a selective effect. In some language structures, taking time to reflect and verbalize thoughts is not enough to internalize the metalinguistic information. The difficulty caused by the inherent complexity along with L1 influence seems to be a significant mediating factor for the efficacy of WL. Thus, teachers should take a different approach for these features. Third, teachers need to pay attention to learners' WL because it provides them with valuable information regarding how learners interpret AWCF. This information can be a precious resource for teachers to decide what to prepare for further lessons. Lastly, in order for WL to be effective, students need to be familiar with the WL processes. Thus, training and practicing time would enable students to take more advantage of WL.

The present study shed light on the research in WL and AWCF because it filled the gaps in the previous studies by adopting the error correction tests (not revisions) to access accuracy improvement, conducting the research in intact classrooms throughout a semester, and examining the differential effects of WL on various language forms. However, some of the main limitations need to be pointed for future research. First of all, one limitation of the design is that although (-) WL group was given time to reflect on the corrections AWCF

offered while (+) WL group engaged in languaging, the former case solely relied on their willingness to participate in this process. In other words, in the situation where no tasks were required, (-) WL group might not have paid attention to the corrections, and eventually, this different condition might have affected the results of the present study. In a follow-up study, this problem could be dealt with by leading (-) WL group to involve in a more demanding task (e.g., lecture) while (+) WL group is doing languaging. Another design-related shortcoming concerns the fact that there was not a true control group which did not receive AWCF, so the study was limited to find out the effect of AWCF itself. Future studies are encouraged to include a control group to arrive at more rigorous evidence regarding the effect of WL in response to AWCF on L2 learning. Third, the study only measured accuracy improvement using the error correction test. In other words, it did not test how accurately the students could use the corrected forms in a new piece of writing, which is definitely a worthwhile object of study for future research. Lastly, the study did not take learner individual differences into consideration. Considering learner individual differences, such as proficiency levels, language aptitude, and attitudes can be a mediating factor for learner internalization of L2 input, exploring whether or not these factors are related to the effect of WL would be a thrilling topic for future research.

Applicable level: Tertiary

REFERENCES

Attali, Y. (2004, April). *Exploring the feedback and revision features of Criterion*. Paper presented at the 2004 National Council on Measurement in Education Conference, San Diego, CA.

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

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14*(3), 191-205.

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385-405.

Chen, C. (2011). *The influence of Chinese prominence construction on English acquisition*. Unpublished master's thesis, Krisianstad University, Sweden.

Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, *12*(2), 94-112.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, *12*(4), 425-434.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141-172.

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.

Fukuta, J., Tamura, Y., & Kawaguchi, Y. (2019). Written languaging with indirect feedback in writing revision: Is feedback always effective? *Language Awareness*, *28*(1), 1-14.

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6), 1-43.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, *16*(1), 40-53.

Huebner, T. (1983). Linguistic systems and linguistic change in an interlanguage. *Studies in Second Language Acquisition*, *6*(1), 33-53.

Ishikawa, M. (2018). Written languaging, learners' proficiency levels and L2 grammar learning. *System*, *74*, 50-61.

Ishikawa, M., & Révész, A. (2020). L2 learning and the frequency and quality of written languaging. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 219-240). Amsterdam: John Benjamins Publishing Company.

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

Jarvie, G. (1993). *Bloomsbury grammar guide*. London: Bloomsbury Publishing Ltd.

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165-211). Amsterdam: John Benjamins.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, *42*(2), 173-196.

Kim, J. (2021). Examining the role of WL in ehnacing the effect of indirect CF on accuracy improvement. *Journal of Research in Curriculum & Instruction, 25*(4), 1-10.

Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, *41*(3), 432-454.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford, England: Oxford University Press.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer.

Li, D. C. (2009). Learning English for academis purposes: Why Chinese learners find EAP do difficult to master. *Revista Canadian de Estudio Ingleses, 59*, 35-48.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1-18.

Luria, A. R. (1999). Speech development and the formation of mental processes. In P. Lloyd & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments* (Vol. 2) (pp. 84-122). London: Taylor & Francis.

Mohamed, A., Lian, G., & Eliza, W. (2004). English errors and Chinses learners. *Sunway College Journal, 1*, 83-97.

Moradian, M. R., Hossein-Nasab, M., & Miri, M. (2020). Effects of written languaging in response to direct and indirect corrective feedback on developing writing accuracy. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 267-286). Amsterdam: John Benjamins Publishing Company.

Moradian, M. R., Miri, M., & Hossein-Nasab, M. (2017). Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied Linguistics, 27*(2), 406-426.

Niu, R., & You, X. (2020). Effects of indirect corrective feedback with and without written languaging on L2 written accuracy: A multitask intervention study. *Asia-Pacific Edu Res, 29*(4), 343-351.

Ohta, A. S. (2001). *Second language acquisition processes in the classroom learning Japanese*. Mahwah, NJ: Lawrence Erlbaum.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653-674.

Richards, J. C. (Ed.). (1974). *Error analysis*. Singapore: Longman.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, England: Cambridge University Presss.

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.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592-604.

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

Suzuki, W. (2009). Improving Japanese university students' second language writing accuracy: Effects of languaging. *ARELE: Annual Review of English Language Education in Japan*, *20*, 81-90.

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning*, *62*(4), 1110-1133.

Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing & Pedagogy*, *8*(3), 461-482.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). New York: Cambridge University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). New York: Oxford University Press.

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

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371-391.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320-337.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, *37*(3-4), 285-304.

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

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

Truscott, J., & Hsu, A. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*(4), 292-305.

Tseng, Y. C., & Liou, H. C. (2006). The effects of online conjunction materials on college EFL students' writing. *System*, *34*(2), 270-283.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: The MIT Press.

Woodworth, J., & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL Classroom. *TESL Canada Journal, 37*(2), 234-247.

APPENDIX

Error Correction Test [Sample]

Each of the following sentences contains an error. Please locate the error and correct it.

(1 point for each correct answer, the total score is 50.)
1. She has studied at Peking University since two years.
Error location_____ Correction_____

2. You had better choose another gift instead of a umbrella.
Error location_____ Correction_____

3. One of my classmates go to the museum every week.
Error location_____ Correction_____

4. The weather is nice we can go out for a picnic.
Error location_____ Correction_____

5. He could not afford for a new TV.
Error location_____ Correction_____