SELL Journal, February 2024
VOL. 9 NO. 1, 29-64
e-ISSN 2580-8400, p-ISSN 2528-6595

Exploring Students' Writing Performance through Corrective Feedback

Masrul[1], Bayu Hendro Wicaksono[2]
masrulm25@gmail.com[1], bayu_hw@umm.ac.id[2]
Universitas Pahlawan Tuanku Tambusai[1], Universitas Muhammadiyah Malang[2]

Abstract:

Prior studies on corrective feedback have highlighted the effect of feedback on students writing accuracy. In this present study, the impacts of corrective feedback on students' writing performance; grammatical accuracy, complexity, fluency, content and organization, appropriateness, argumentation, interpersonal, textual, lexical were examined. Participants were 96 students who were assigned into different groups to accept written corrective feedback, content and organization of ideas, multilateral feedback (i.e., on grammatical accuracy, complexity, fluency, appropriateness, argumentation, interpersonal, textual, lexical), and control group that did not accept any feedback during learning sessions. Pre-test, post-test and delayed post-test and weekly writing revisions were carried out. The results showed that the four experiment groups showed significant improvements. The results of this study showed that students had language learning potential which can be enhanced when teachers provide corrective feedbacks. This study suggested that students' writing skills can be improved by providing sustained feedback.

Keywords: Corrective Feedback, Exploring Writing, Writing Performance

Introduction

Over the past two decades, the impact of teacher feedback toward Second Language learners writing has intrigued researchers around the world to explore deeper. The impacts of writing feedback and teacher feedback in L2 classes are also popular topics. These two fields of study differ in terms of the writing assessment. Teacher feedback assesses content, organization, and accuracy, while quantitative research focuses on feedbacks on grammatical accuracy. A study with major focus on the impact of feedback on grammar, accuracy and content was carried out by (Zhang, 2015; Nicolás–Conesa et al., 2019; Kim & Emeliyanova, 2019). The effect of feedback needs to be investigated in an experimental study (Rouhi & Azizian, 2013).

This quantitative study mainly focused on investigating the influences of writing aspects. This study also highlighted how feedback affected the accuracy of grammar, content and organization, complexity, fluency, appropriateness, argumentation, interpersonal, textual, lexical. In addition, the combined effects of both direct and indirect feedback were also examined. Direct feedback involves identifying errors and providing specific correction, as described by (Bitchener et al., 2005a). Stefanou & Révész (2015) found direct written feedback more beneficial than no feedback, but the evidence for the benefits of providing metalinguistic information was not clear. The study also revealed that those with higher grammatical sensitivity and metalinguistic knowledge were at an advantage when receiving immediate feedback. On the other hand, indirect feedback presents the existence of errors without directly providing the correct form, as described by (Ferris et al., 2011).

Review of Literature

The literature review of this study emphasizes on several relevant key components. The first component is the impact of feedback on writing accuracy, specifically for L2 learners, including feedback on both content and grammar accuracy. The second component is the examination of language use feedback and its impact on grammatical accuracy in writing. These findings contribute to the overall understanding of the effects of WCF on the writing of L2 students, both directly and indirectly. The significance of feedback as also examined by Rouhi & Azizian (2013) based on the theories of process writing and collaborative learning. The study found that feedback provider had stronger role than the recipient, although the recipients also showed more significant role than the comparison group. Both the feedback provider and the recipient play significant role in the evaluation of the WCF effectiveness in writing.

The theory of Direct and indirect Written Corrective Feedback has been employed by Hosseiny (2014) and Youjin Kim & Choi (2020) with varying findings. Hosseiny found that the group receiving written corrective feedback showed greater improvement in their grammar knowledge compared to the control group. Whereas, Youjin Kim and Choi discovered that although direct feedback was more successful in improving students' writing accuracy,

both direct and indirect feedback were effective in promoting the learning of new linguistic elements through collaborative writing. However, there was no difference in students' perceptions between the two feedback settings. Youjin Kim and Kim (2020) also employed the same theoretical approach and found that task repetition enhanced writing fluency, but synchronous feedback had a negative impact on fluency during collaborative writing. Instead, the approach improved students' writing accuracy yet it had no substantial effect on the acquisition of target grammatical properties.

Lee (2020) conducted a study with main focus on the evaluation of the scope of written corrective feedback (WCF) in the context of English as a Second Language (ESL) writing courses. He attempted to address three major issues in the field of WCF research (WCF; Bitchener, 2012), including the vagueness of related terms, limitations of laboratory-style classroom conditions, and the narrow scope of previous studies. He found that these limitations can be tackled by making WCF research more relevant to real-life classrooms. This results of his study contribute to the body of knowledge in the area of WCF by exploring its role in providing specific, targeted, and personalized grammatical feedback to help L2 learners acquire certain grammatical forms and structures. As a teacher, the author also provides written corrective feedback to L2 learners to correct mistakes in their writing.

Feedback on Complexity, Accuracy, and Fluency

Studies on the impact of written corrective feedback (CF) on second language writing mainly focused on the determination of the most feasible way to improve L2 writing accuracy through direct or indirect feedback for various types of errors. One study group that focused on CF investigated the effect of CF on one or two grammatical structures, where students received direct feedback, including the correct form, with or without oral or written linguistic explanations. Some researchers, including (Sheen, 2007), found that after treatment, the treatment group showed improvement in the targeted structure (such as articles in English) compared to the control group. In another study, You Jin Kim & Emeliyanova (2019) compared the types of written corrective feedback (WCF) and students' revision behaviors in the classroom during peer-tasks and individual tasks. The results indicated that

the group that corrected their errors in pairs made more accurate corrections than the self-correction group. Both groups improved significantly in writing accuracy after receiving feedbacks. The effectiveness of different types of written feedback also depends on the complexity of the structure being targeted. Ferris suggests that errors in verb tense and form are easier to fix since they follow a patterned and rule-based way (Gao et al., 2000). Similar findings were also found in oral CF and written CF (Yang & Lyster, 2010).

This present longitudinal study was performed to examine the effect of written feedback on grammatical accuracy, content, and organization, where students received sustained feedback, attended to the feedback, and engaged with it as they did the revision. These concept aligns with cognitive theories and hypotheses in second language acquisition, particularly the noticing hypothesis (Schmidt, 1990) and the skill acquisition theory (DeKeyser, 1997). These theories are significant in L2 development through the interface hypothesis of second language instruction which emphasizes the role of explicit information processing in SLA through providing opportunities for attention and guided output practice (Ellis, 2011).

Organization and Content in Writing

The term "content" in literature refers to the ideas or meanings conveyed in a piece of writing. The content of writing can encompass various elements of writing such as vocabulary, coherence, grammar, and text length, yet its relationship to these elements has not yet been critically examined. Content is evaluated based on the quality of ideas expressed, including their relevance, completeness, originality, development, and logical consistency. Bae & Lee (2011) stated that writing content is evaluated based on the relevance to the given task, as well as the ingenuity and completeness of the thoughts. Mei (2001) defined high-quality content as extremely innovative, engaging, and fully developed content. On the other hand, Bae et al (2016) described that high-quality content is characterized by greater originality, interpretative brilliance, and a depth of understanding.

In the other hand, Organization was operationalized in four constructs: paragraph structure, logical flow of ideas, cohesiveness, and paragraphing (Hsu & Yuan, 2018). The content and organization were assessed in an eight-

point scale. Content was evaluated based on the thesis development, topic coverage, importance of details, and conclusion of primary arguments, whereas organization was evaluated based on fluency of expression (i.e. how easily concepts were related) as well as organization and sequencing of ideas. Language complexity was measured based on the average clauses per T-unit, mean length of T-unit, lexical diversity (Guiraud index), and lexical sophistication. Whereas, weighted clause ratio was utilized to assess language accuracy (WCR).

Method

This study was conducted at a Senior High School in Bangkinang, Indonesia in more than two consecutive periods. In the first week of the semester, the researcher came into the target class (intermediate level) to explain about the research and to ask for students' consent to participate in this study. Students were instructed to write one composition per week. There were 96 students participating (66 female and 30 male) in 12 sessions. All of the participants were Indonesian speaker who had learned English as foreign language for six years at school.

There were fifteen recount-writing topics used in this study. The pre-test (meeting 1), post-test (meeting 10) and delayed post-test (meeting 16) employed the same topic: "explain how you feel when you just won something (game, price, etc) Fourteen others prompt, "e.g. write about a funny thing that happened to you" were used for meeting 2-16.

In the final week of the study, a post-test was conducted. The participants were given 45 minutes to write a composition on a given theme, and their scripts were then collected, scanned, and analyzed. Data analysis was performed to investigate the impact of the different types of feedback on the students' grammatical accuracy, content, and organization in their writing. The results of this study provide insights into the effectiveness of different types of feedback on L2 writing improvement. The results of this study also contributed to the current understanding of the role of feedback in second language acquisition. Practical implications have been proposed for the practice of English as a Foreign Language in Indonesia, where feedback is an essential aspect of language instruction.

Data Collection Technique

The outcomes of pre-test, post-test, and delayed test scores of each group were measured to examine the impacts of comments on the writing outcomes. Furthermore, the results of the experiment were compared to those of the control group to determine if the feedback had a significant impact on the participants' writing performance. The data of this study were analyzed using inferential statistics, including ANOVA and post-hoc tests, to determine the significant differences between the groups. The results were then presented in the forms of mean scores and standard deviations. There were 96 participants and each of them wrote 31 scripts, resulting in a total script of 3038 scripts. The scripts contained one pre-test script, one post-test script, one delayed test, and twenty-eight scripts with the revision (script 2-15). The total number of pre-test script, post-test script, and delayed test was 294 that were assessed based on ten points; grammatical accuracy, complexity, fluency, content, organization, appropriateness, argumentation, interpersonal, textual and lexical.

Accuracy was measured by calculating the number of errors divided with the total word multiplied by 100. The percentage of error free clause was measured by dividing the number of error free clause and multiplied it by 100. After that, the complexity was ,measured and the number total of word with the total of clause and fluency was measured by paying attention to the number of words.

Two independent scales were used to measure the content and organization. Content was assessed based on five elements: writer's position, relevance, idea, adequacy, and clarity. Organization was assessed with 4 sets: paragraph structure, logical flow of ideas, cohesion and paragraphing. Additionally, the organization scored 4-16 and 5-20 for content.

Data Analysis Technique

The Repeated Measures MANOVA was used in this study as it allowed the analysis of multiple dependent variables (e.g. grammatical accuracy, complexity, fluency, content, and organization) that were measured repeatedly over time (e.g. from pre-test to post-test and delayed test). This analysis provided information on how the scores of these variables changed over time

and how the different types of feedback (e.g. constructive criticism, content feedback, organization feedback, or no feedback) affected the changes. This analysis also controlled the potential influence of individual differences among participants.

Results and Discussions

Complexity

Descriptive Statistics

Table 1. Descriptive Statistics on Complexity

Descriptive Statistics				
	Group	Mean	Std. Deviation	N
Pre-test	1	3.0050	1.69683	8
	2	3.6663	1.56246	8
	3	3.5750	1.56339	8
	4	3.3575	1.72443	8
	Total	3.4009	1.57852	32
Post-test	1	3.8738	1.54580	8
	2	4.0975	1.57748	8
	3	4.6488	0.97792	8
	4	4.5275	1.12247	8
	Total	4.2869	1.30535	32
Delayed post-test	1	4.7225	1.18691	8
	2	4.8988	0.75118	8
	3	4.9688	0.95890	8
	4	4.9125	1.22699	8
	Total	4.8756	1.00094	32

The following table displays the mean scores of pre-test, post-test, and delayed post-test for four different groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.005, 3.6666, 3.575, and 3.358 respectively. The

post-test mean scores for Group 1, 2, 3, and 4 are 3.874, 4.098, 4.649, and 4.528 respectively. Finally, the delayed post-test mean scores for Group 1, 2, 3, and 4 are 4.723, 4.899, 4.969, and 4.913 respectively.

Tests of within Subject Effect

Table 2. Tests of within Subject Effect on Complexity
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	35.266	1.562	22.582	37.593	.000

Interpretation:

Based on result above, significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected), indicating the presence of a different syntactic complexity over time. The pairwise comparison analysis was then performed.

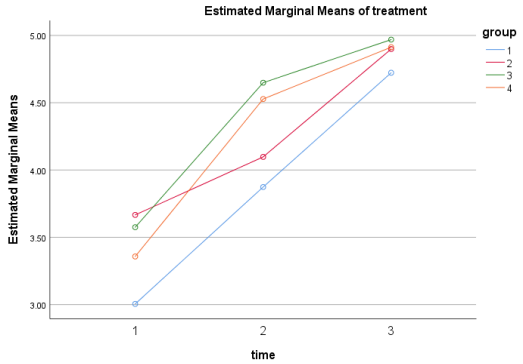
Table 3. Pairwise Comparisons Complexity
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.886*	.169	.000	-1.315	-.457
	3	-1.475*	.208	.000	-2.004	-.945
2	1	.886*	.169	.000	.457	1.315
	3	-.589*	.128	.000	-.914	-.263
3	1	1.475*	.208	.000	.945	2.004
	2	.589*	.128	.000	.263	.914

Based on table above, there is difference value of complexity on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 3 performed the best as it obtained the highest syntactic complexity value among other groups.

Fluency

Descriptive Statistic

Table 4. Descriptive Statistic on Fluency

Descriptive Statistics				
	Group	Mean	Std. Deviation	N
Pre-test	1	3.4238	1.21103	8
	2	3.7238	1.30535	8
	3	4.1213	1.16886	8
	4	3.4875	1.43676	8
	Total	3.6891	1.25209	32
Post-test	1	4.6775	1.39396	8
	2	4.6063	1.34642	8
	3	5.0550	1.07939	8
	4	4.0825	1.52289	8
	Total	4.6053	1.32626	32

Delayed post-test	1	5.3350	0.85192	8
	2	5.0875	1.10317	8
	3	5.5112	0.82466	8
	4	4.7625	1.13471	8
	Total	5.1741	0.98217	32

The following table displays the mean scores of pre-test, post-test, and delayed post-test for four different groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.424, 3.724, 4.121, and 3.488 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.678, 4.606, 5.055, and 4.083 respectively. Finally, the delayed post-test mean scores for Group 1, 2, 3, and 4 are 5.335, 5.088, 5.511, and 4.762 respectively.

Tests of within Subject Effect

Table 5. Tests of within Subject Effect on Fluency
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	35.928	1.659	21.656	48.280	.000

The significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected) shows the difference in the fluency over time. Then, pairwise comparison analysis was performed.

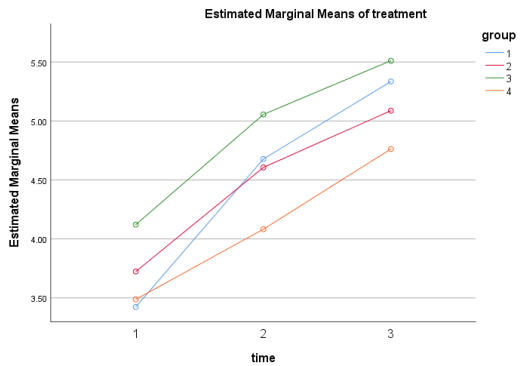
Table 6. Pairwise Comparisons Fluency
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.916*	.170	.000	-1.350	-.482
	3	-1.485*	.167	.000	-1.911	-1.059
2	1	.916*	.170	.000	.482	1.350
	3	-.569*	.113	.000	-.856	-.281
3	1	1.485*	.167	.000	1.059	1.911
	2	.569*	.113	.000	.281	.856

As seen in the Table, the score for fluency on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 3 performed the best as shown by the highest fluency score.

Content

Descriptive Statistics

Table 7. Descriptive Statistics for Content

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.6450	1.55619	8
	2	3.3375	1.79131	8
	3	2.9188	1.52627	8
	4	3.0788	1.38744	8
	Total	3.2450	1.51993	32
Post-test	1	4.4275	1.56225	8
	2	4.8112	0.77459	8
	3	3.7100	1.55115	8
	4	3.8263	1.50165	8
	Total	4.1938	1.39528	32
Delayed post-test	1	4.7550	1.28178	8
	2	5.3925	0.62751	8
	3	4.8487	0.88113	8
	4	4.6088	1.29342	8
	Total	4.9013	1.05051	32

The following table displays the mean scores of pre-test, post-test, and delayed post-test for four different groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.645, 3.338, 2.919, and 3.079 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.428, 4.811, 3.710, and 3.826 respectively. Finally, the delayed post-test mean scores for Group 1, 2, 3, and 4 are 4.755, 5.393, 4.849, and 4.609 respectively.

Tests of within Subject Effect

Table 8. Tests of within Subject Effect on Content
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Sphericity Assumed	44.201	2	22.101	35.247	.000

The significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected), indicating the gap in fluency score over time. Pairwise comparison analysis was then conducted.

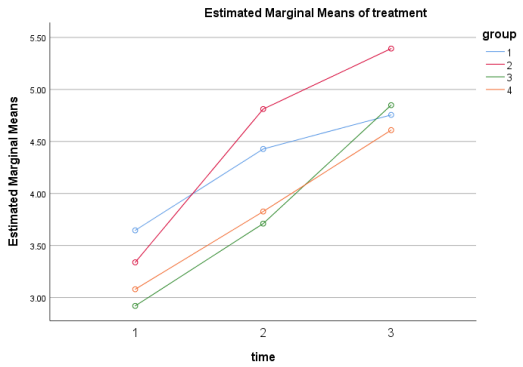
Table 9. Pairwise Comparisons on Content
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.949*	.206	.000	-1.473	-.425
	3	-1.656*	.227	.000	-2.235	-1.077
2	1	.949*	.206	.000	.425	1.473
	3	-.707*	.154	.000	-1.098	-.317
3	1	1.656*	.227	.000	1.077	2.235
	2	.707*	.154	.000	.317	1.098

Significant difference was found in the content on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 2 has the highest score among other groups.

Organization

Descriptive Statistics

Table 10. Descriptive Statistics on Organization

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.1100	1.33181	8
	2	2.6900	1.26854	8
	3	3.0663	1.13176	8
	4	3.3263	1.31041	8
	Total	3.0481	1.22270	32
Post-test	1	4.3025	1.17909	8
	2	3.8025	1.30721	8
	3	4.2725	0.86392	8
	4	4.4925	0.91484	8
	Total	4.2175	1.06015	32
Delayed post-test	1	4.9150	0.86298	8
	2	4.7175	1.05673	8
	3	4.8750	0.73975	8
	4	4.8600	0.87271	8
	Total	4.8419	0.84948	32

Table 10 showcases the mean scores for four groups in pre-test, post-test, and delayed post-test. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.110, 2.690, 3.066, and 3.326 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.303, 3.803, 4.273, and 4.493 respectively. Finally, the delayed post-test mean scores for Group 1, 2, 3, and 4 are 4.915, 4.718, 4.875, and 4.860 respectively.

Tests of within Subject Effect

Table 11. Tests of within Subject Effect on Organization
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	53.065	1.343	39.510	55.603	.000

Significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected) shows the gap in organization over time. Hence, pairwise comparison analysis was performed.

Table 12. Pairwise Comparisons on Organization
Pairwise Comparisons

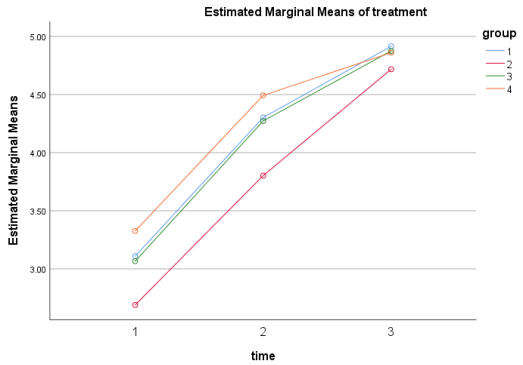
Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.169*	.214	.000	-1.714	-.625
	3	-1.794*	.183	.000	-2.261	-1.327
2	1	1.169*	.214	.000	.625	1.714
	3	-.624*	.101	.000	-.881	-.368

3	1	1.794*	.183	.000	1.327	2.261
	2	.624*	.101	.000	.368	.881

Difference gap in the scores of organization on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 4 appeared as the best group for obtaining the highest score among other groups.

Grammatical Accuracy

Descriptive Statistics

Table 13. Descriptive Statistics of Grammatical Accuracy

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.0863	1.40136	8
	2	3.1613	1.33636	8
	3	2.5588	1.29764	8
	4	3.3513	1.31145	8
	Total	3.0394	1.30553	32
Post-test	1	4.3375	0.84166	8
	2	3.9887	1.39076	8

	3	4.4937	0.97974	8
	4	4.1325	1.20833	8
	Total	4.2381	1.08697	32
Delayed post-test	1	4.8625	0.86217	8
	2	4.6387	0.89662	8
	3	5.2513	1.09836	8
	4	5.2638	0.91860	8
	Total	5.0041	0.94081	32

Table 13 presents the mean scores of pre-test, post-test, and delayed post-test for four groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.086, 3.161, 2.559, and 3.351 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.338, 3.989, 4.494, and 4.133 respectively. The delayed post-test mean scores for Group 1, 2, 3, and 4 are 4.863, 4.639, 5.251, and 5.264 respectively.

Tests of within Subject Effect

Table 14. Tests of within Subject Effect on Grammatical Accuracy
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	62.759	1.533	40.951	53.125	.000

A significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected) shows different grammatical accuracy over time. Then, the pairwise comparison analysis was performed.

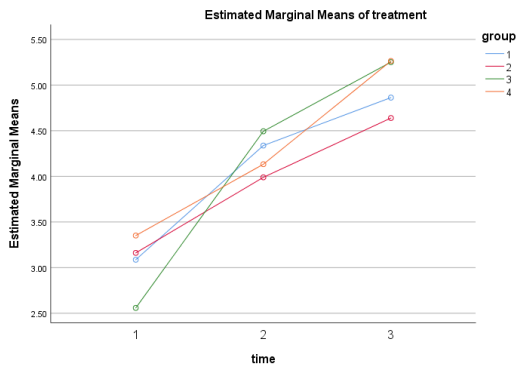
Table 15. Pairwise Comparisons on Grammatical Accuracy
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.199*	.210	.000	-1.735	-.663
	3	-1.965*	.223	.000	-2.532	-1.397
2	1	1.199*	.210	.000	.663	1.735
	3	-.766*	.129	.000	-1.096	-.436
3	1	1.965*	.223	.000	1.397	2.532
	2	.766*	.129	.000	.436	1.096

There are differences in the scores of grammatical accuracy on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 3 outperformed other groups with the highest score on grammatical accuracy.

Appropriateness

Descriptive Statistics

Table 16. Descriptive Statistics on Appropriateness

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	4.1600	1.01940	8
	2	4.0425	1.25694	8
	3	3.4950	1.44704	8
	4	3.2788	1.64801	8
	Total	3.7441	1.34797	32
Post-test	1	4.6638	0.91384	8
	2	4.7000	1.10965	8
	3	4.3238	1.45257	8
	4	4.2788	1.02956	8
	Total	4.4916	1.10464	32
Delayed post-test	1	5.0688	1.02673	8
	2	5.0813	1.07773	8
	3	4.6988	1.31493	8
	4	4.7463	0.79558	8
	Total	4.8987	1.03249	32

Table 16 displays the mean scores of pre-test, post-test, and delayed post-test for four groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 4.160, 4.043, 3.495, and 3.279 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.664, 4.700, 4.324, and 4.279 respectively. The delayed post-test mean scores for Group 1, 2, 3, and 4 are 5.069, 5.081, 4.699, and 4.746 respectively.

Tests of within Subject Effect

Table 17. Tests of within Subject Effect on Appropriateness
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	21.951	1.167	18.803	27.338	.000

Significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected) presents the gaps in appropriateness over time. The pairwise comparison analysis was then performed.

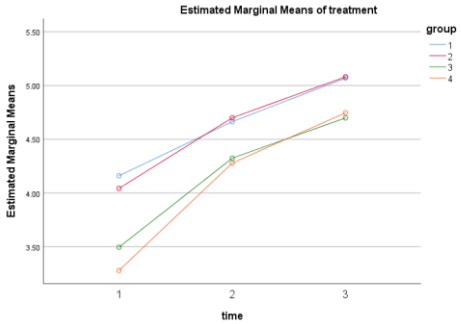
Table 18. Pairwise Comparisons Appropriateness
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.747*	.191	.002	-1.235	-.260
	3	-1.155*	.186	.000	-1.629	-.680
2	1	.747*	.191	.002	.260	1.235
	3	-.407*	.063	.000	-.567	-.248
3	1	1.155*	.186	.000	.680	1.629
	2	.407*	.063	.000	.248	.567

Gaps are present in the scores of appropriateness on pre-test and post-test (sig. 0.002), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



The mean scores of appropriateness for all group have increased over time. Group 1 obtained the highest score in appropriateness.

Argumentation

Descriptive Statistics

Table 19. Descriptive Statistics on Argumentation

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.7863	1.46541	8
	2	3.2987	1.50936	8
	3	3.1863	1.81090	8
	4	3.6438	1.49637	8
	Total	3.4788	1.51898	32
Post-test	1	5.0875	0.49690	8
	2	4.4163	1.12199	8
	3	4.4888	1.10951	8
	4	4.5787	1.19163	8
	Total	4.6428	1.00503	32
Delayed post-test	1	5.4100	0.44023	8
	2	4.6850	1.05740	8
	3	4.8363	1.01492	8
	4	4.7688	1.12454	8
	Total	4.9250	0.94777	32

Table 19 presents the mean scores of pre-test, post-test, and delayed post-test for four groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.786, 3.299, 3.186, and 3.644 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 5.088, 4.416, 4.489, and 4.579 respectively. The delayed post-test mean scores for Group 1, 2, 3, and 4 are 5.410, 4.685, 4.836, and 4.769 respectively.

Tests of within Subject Effect

Table 20. Tests of within Subject Effect on Argumentation
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	37.614	1.032	36.437	41.972	.000

The significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected) shows the presence of gaps in the argumentation over time. Pairwise comparison analysis was then performed.

Table 21. Pairwise Comparisons on Argumentation
Pairwise Comparisons

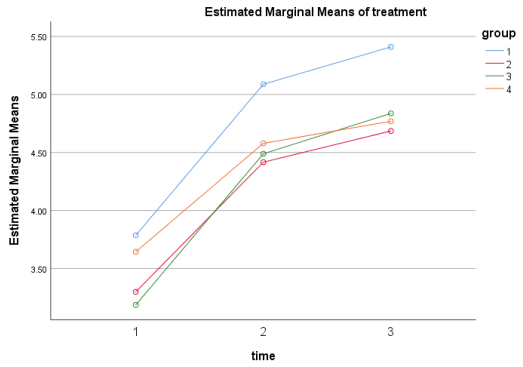
Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.164*	.207	.000	-1.691	-.637
	3	-1.446*	.201	.000	-1.957	-.935
2	1	1.164*	.207	.000	.637	1.691
	3	-.282*	.031	.000	-.360	-.204

3	1	1.446*	.201	.000	.935	1.957
	2	.282*	.031	.000	.204	.360

Gaps in the scores of argumentation on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000) were identified.

Profile Plot



,

Group 1 was the best group for obtaining the highest score on argumentation.

Interpersonal

Descriptive Statistics

Table 22. Descriptive Statistics on Interpersonal

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.1163	1.28149	8
	2	3.7125	1.38737	8
	3	3.4150	1.64392	8
	4	3.0212	1.70991	8
	Total	3.3163	1.46687	32
Post-test	1	4.3525	0.94314	8
	2	4.9013	0.94151	8

	3	4.5388	0.96252	8
	4	4.1975	1.19640	8
	Total	4.4975	1.00229	32
Delayed post-test	1	4.8263	1.12144	8
	2	5.3488	0.79112	8
	3	5.2025	0.70807	8
	4	5.1313	0.89030	8
	Total	5.1272	0.86891	32

Table 22 shows the mean scores of pre-test, post-test, and delayed post-test for four groups. The pre-test mean scores for Group 1, 2, 3, and 4 are 3.116, 3.713, 3.415, and 3.021 respectively. The post-test mean scores for Group 1, 2, 3, and 4 are 4.353, 4.901, 4.539, and 4.198 respectively. The delayed post-test mean scores for Group 1, 2, 3, and 4 are 4.826, 5.349, 5.203, and 5.131 respectively.

Tests of within Subject Effect

Table 23. Tests of within Subject Effect on Interpersonal
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	54.094	1.566	34.552	40.513	.000

The significant value (p-value) = 0.000 or $p - value < \alpha$ (H_0 rejected), implying the presence of gaps in the argumentation over time. The pairwise comparison analysis was then carried out.

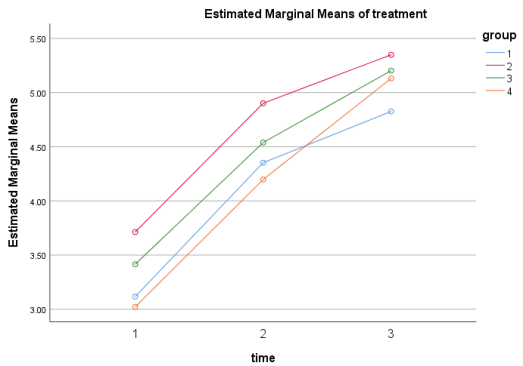
Table 24. Pairwise Comparisons Interpersonal
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.181*	.201	.000	-1.693	-.670
	3	-1.811*	.248	.000	-2.442	-1.180
2	1	1.181*	.201	.000	.670	1.693
	3	-.630*	.153	.001	-1.019	-.240
3	1	1.811*	.248	.000	1.180	2.442
	2	.630*	.153	.001	.240	1.019

Differences are found in the scores of interpersonal on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.001)

Profile Plot



Group 2 performed the best as shown by the highest score for interpersonal among other groups.

Textual

Descriptive Statistics

Table 25. Descriptive Statistics on Textual

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.5237	1.39438	8
	2	4.0350	1.29902	8
	3	3.8425	1.34102	8
	4	2.9288	1.58259	8
	Total	3.5825	1.40469	32
Post-test	1	4.9712	0.59012	8
	2	4.6750	0.67521	8
	3	4.7575	1.25127	8
	4	4.1325	1.20201	8
	Total	4.6341	0.97980	32
Delayed post-test	1	5.5425	0.53245	8
	2	5.4613	0.71820	8
	3	5.4987	0.50865	8
	4	4.5675	1.14820	8
	Total	5.2675	0.84028	32

According to Table 25, the mean pre-test scores for group 1, 2, 3, and 4 were 3.524, 4.035, 3.843, and 2.929, respectively. The mean post-test scores for group 1, 2, 3, and 4 were 4.971, 4.675, 4.758, and 4.133, respectively. The mean delayed post-test scores for group 1, 2, 3, and 4 were 5.543, 5.461, 5.499, and 4.568, respectively.

Tests of within Subject Effect

Table 26. Tests of within Subject Effect on Textual
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	46.360	1.542	30.060	42.594	.000

p-value = 0.000 or $p - value < \alpha$ (H_0 rejected) shows the presence of difference in textual over time. The pairwise comparison analysis was conducted.

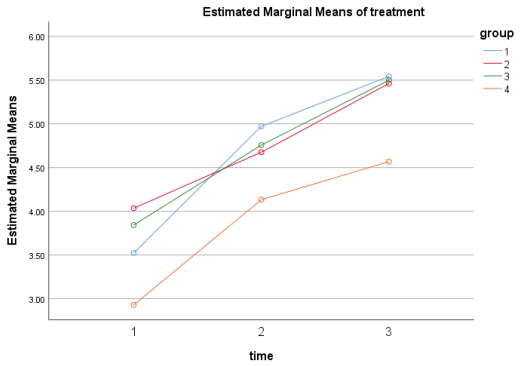
Table 27. Pairwise Comparisons on Textual
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.052*	.187	.000	-1.527	-.576
	3	-1.685*	.223	.000	-2.252	-1.118
2	1	1.052*	.187	.000	.576	1.527
	3	-.633*	.132	.000	-.971	-.296
3	1	1.685*	.223	.000	1.118	2.252
	2	.633*	.132	.000	.296	.971

Gaps in the scores of textual on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000) were found.

Profile Plot



Group 1 appeared as the best group with the highest score on textual.

Lexical

Descriptive Statistics

Table 28. Descriptive Statistics on Lexical

Descriptive Statistics				
	group	Mean	Std. Deviation	N
Pre-test	1	3.9038	1.07723	8
	2	3.2025	1.51620	8
	3	4.1088	1.42178	8
	4	3.5662	1.64312	8
	Total	3.6953	1.40340	32
Post-test	1	4.7788	1.08435	8
	2	4.4788	0.85753	8
	3	4.7925	1.26075	8
	4	4.2450	1.41987	8
	Total	4.5737	1.13979	32
Delayed post-test	1	5.2425	0.96952	8
	2	5.0463	0.74379	8
	3	5.0963	1.13785	8
	4	5.1038	0.96422	8
	Total	5.1222	0.91921	32

Based on table 28 above, pre-test at group 1 (3.904), group 2 (3.203), group 3 (4.109) group 4 (3.566). Mean of post-test at group 1 (4.779), group 2 (4.479), group 3 (4.793) group 4 (4.245). Mean of delayed post-test at group 1 (5.243), group 2 (5.046), group 3 (5.096) group 4 (5.104)

Tests of within Subject Effect

Table 29. Tests of within Subject Effect on Lexical
Tests of Within-Subjects Effects

Measure: treatment

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Greenhouse-Geisser	33.156	1.483	22.350	41.102	.000

p-value = 0.000 or $p - value < \alpha$ (H_0 rejected) shows the gap in the lexical over time. Pairwise comparison analysis was performed afterwards.

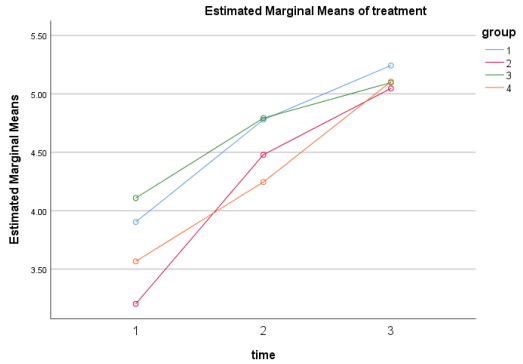
Table 30. Pairwise Comparisons on Lexical
Pairwise Comparisons

Measure: treatment

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.878*	.165	.000	-1.300	-.457
	3	-1.427*	.192	.000	-1.916	-.938
2	1	.878*	.165	.000	.457	1.300
	3	-.548*	.107	.000	-.820	-.277
3	1	1.427*	.192	.000	.938	1.916
	2	.548*	.107	.000	.277	.820

Table 30 shows the gap in the lexical scores on pre-test and post-test (sig. 0.000), pre-test and delayed post-test (sig. 0.000), post-test and delayed post-test (sig. 0.000)

Profile Plot



Group 1 obtained the highest score in lexical

Discussions

Improving Grammatical Accuracy through Corrective Feedback

Corrective feedback on grammar, content and organization can improve students' writing accuracy. The group that received written corrective feedback showed significant improvement on accuracy. Students also obtained significant improvement in linguistic accuracy (Evans et al., 2011). Similarly, Students who got corrective feedback outperformed those that did not (Sarré et al., 2019) because written corrective feedback targeting single linguistic features improved learners' accuracy (Bitchener & Knoch, 2008). On the other side, feedback is often undertaken at the surface level, with an emphasis on meta-linguistic qualities, often at the expense of content-related concerns. This might occur due to the emphasis on exam preparation and the requirement to produce evidence of progress (Hardman & Bell, 2018). Therefore, it is crucial for the students to understand linguistic information provided through corrective feedback (Simard et al., 2015).

Improvements in Syntactic Complexity and Fluency

This research aligns with previous findings suggesting that attention to grammatical errors and feedback for accuracy improve the accuracy in writing

and language learning. (Nicolás–Conesa, Manchón, & Cerezo, 2019) and Kim et al. (2020) also highlighted the roles of task repetition and synchronous written corrective feedback (WCF) to improve writing fluency and accuracy. The importance of considering the type of feedback and its timing in language learning contexts is emphasized (Y. J. Kim et al., 2020).

Improvements in Content and Organization

According to the findings, only recipients who received feedback on content and organization saw a significant improvement in post-test quality. This finding agree with (Jacobs et al., 1998) revealed feedback improved ESL components such as content, organization, language use, and vocabulary significantly; however, the study did not produce the similar benefits for students' punctuation use (Gharehbagh et al., 2019). (R. Ellis et al., 2008) added Written CF (corrective feedback) is effective, at least in the context of English articles, which strengthens the case for teachers to provide written CF because Written CF focusing on a particular feature improves learner accuracy, especially when metalinguistic feedback is provided and the learners have a high level of language analytic ability (Sheen, 2007).

Improvements in Appropriateness, Argumentation, Interpersonal, Textual and Lexical

Empirical evidence regarding the significant gaps in the scpres pf appropriateness, argumentation, interpersonal, textual and lexical in differnet groups. Similarly, (Zhao, 2018) revealed that for Chinese EFL teachers, the perceived appropriateness of peer evaluation is significantly influenced by their understanding of peer assessment, the role of examinations and teachers in the existing learning culture, and teachers' and learners' preparedness to accept and embrace peer review. Meanwhile, (Chen et al., 2020) showed that argumentative writing helped students develop conceptual knowledge. Students in the treatment group 7 outperformed the control group as shown by higher imrpovement from pre- to post-test. The intervention of AW can predict the improvement in students' conceptual performance. In a textual setting, the results showed that the experimental group outperformed the control group in terms of referential cohesion as well as emotive, challenge,

and qualified attitudes. GWE might have improved students' textual cohesion and writing attitude (Chang et al., 2019). Evidences were also found for lexical but not semantic processes in both writing to dictation and copying from memory, as well as for both common nonwords and function words. They also demonstrated that lexical writing can be fluent without any prior practice with the lexical objects in question (Phillips & Goodall, 2007).

Conclusion

In conclusion, the findings of this study have important implications for EFL teachers and students. Sustained feedback on students' writing significantly improved writing performance. Hence, this type of feedback should be provided on a regular basis. To note, that the results of this study are specific to the context of this study and the findings are not generalizable to other classroom contexts. This study also did not focus on a specific approach or a larger number of targeted structures. Groups that did not receive feedback at the beginning of the study eventually were given feedback at the end of the study.

References

Bae, J., Bentler, P. M., & Lee, Y. (2016). On the Role of Content in Writing Assessment. *Language Assessment Quarterly*, *13*(4), 302–328. https://doi.org/10.1080/15434303.2016.1246552.

Bae, J., & Lee, Y. (2011). *Language Testing Assessment of Language Skills*. https://doi.org/10.1177/0265532210382446.

Bitchener, J. (2012). Written Corrective Feedback for L2 Development: Current Knowledge and Future Research. *TESOL Quarterly*, *46*(4), 855–860. https://doi.org/10.1002/tesq.62.

Bitchener, J., & Knoch, U. (2008). The Value of Written Corrective Feedback for Migrant and International Students. *Language Teaching Research*, *12*(3), 409–431. https://doi.org/10.1177/1362168808089924.

Bitchener, J., Young, S., & Cameron, D. (2005a). The Effect of Different Types of Corrective Feedback on ESL Students' Writing. *Journal of Second Language Writing*, *14*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001.

Chang, W., Liao, C., & Chan, T. (2019). Improving Children Textual Cohesion and Writing Attitude in a Game-based Writing Environment. *Computer Assisted Language Learning*, *0*(0), 1–26. https://doi.org/10.1080/09588221.2019.1671459.

Chen, Y., Aguirre-mendez, C., Terada, T., Chen, Y., & Aguirre-mendez, C. (2020). Argumentative Writing as a Tool to Develop Conceptual and Epistemic Knowledge in a College Chemistry Course Designed for Non-science Majors. *International Journal of Science Education*, *0*(0), 1–34. https://doi.org/10.1080/09500693.2020.1837990.

DeKeyser, R. M. (1997). Beyond Explicit Rule Learning. *Studies in Second Language Acquisition*, *19*(2), 195–221. https://doi.org/10.1017/s0272263197002040.

Ellis, N. C. (2011). *Implicit and Explicit SLA and Their Interface.*

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. https://doi.org/10.1016/j.system.2008.02.001.

Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The Efficacy of Dynamic Written Corrective Feedback for University-Matriculated ESL Learners. *System*, *39*(2), 229–239. https://doi.org/10.1016/j.system.2011.04.012.

Ferris, D., Brown, J., Liu, H., Eugenia, M., & Stine, A. (2011). Responding to L2 Students in College Writing Classes: Teacher Perspectives. *TESOL Quarterly*, *45*(2), 207–234. https://doi.org/10.5054/tq.2011.247706.

Gao, C. Z., Celce-Murcia, M., & Larsen-Freeman, D. (2000). The Grammar Book: An ESL/EFL Teacher's Course. *TESOL Quarterly*, *34*(2), 372. https://doi.org/10.2307/3587964.

Gharehbagh, M. J., Stapa, S. H., & Darus, S. (2019). *The Effects of Written Corrective Feedback Using Wikis among ESL Learners*. *25*(1), 1–10.

Hardman, W., & Bell, H. (2018). 'More Fronted Adverbials than Ever Before'. Writing Feedback Practices and Grammatical Metalanguage in an English Primary School. *Language and Education*, *0*(0), 1–16. https://doi.org/10.1080/09500782.2018.1488864.

Hosseiny, M. (2014). The Role of Direct and Indirect Written Corrective Feedback in Improving Iranian EFL Students' Writing Skill. *Procedia - Social and Behavioral Sciences*, *98*, 668–674. https://doi.org/10.1016/j.sbspro.2014.03.466.

Hsu, H., & Yuan, C. (2018). *Using Wiki-mediated Collaboration to Foster L2 Writing Performance*. *22*(3), 103–123.

Jacobs, G. M., Curtis, A., Braine, G., & Huang, S. Y. (1998). Feedback on Student Writing: Taking the Middle Path. *Journal of Second Language Writing*, *7*(3), 307–317. https://doi.org/10.1016/S1060-3743(98)90019-4.

Kim, Y. J., Choi, B., Yun, H., Kim, B., & Choi, S. (2020). Task Repetition, Synchronous Written Corrective Feedback and The Learning of Korean Grammar: A Classroom-based Study. *Language Teaching Research*. https://doi.org/10.1177/1362168820912354.

Kim, Youjin, & Choi, B. (2020). *Comparing the Effects of Direct and Indirect Synchronous Written Corrective Feedback : Learning Outcomes and Students' Perceptions*. *January*. https://doi.org/10.1111/flan.12443.

Kim, Youjin, & Emeliyanova, L. (2019). *The Effects of Written Corrective Feedback on The Accuracy of L2 Writing: Comparing Collaborative and Individual Revision Behavior*. https://doi.org/10.1177/1362168819831406.

Kim, Youjin, & Kim, B. (2020). *Task Repetition, Synchronous Written Corrective Feedback and The Learning of Korean Grammar: A Classroom-based Study*. https://doi.org/10.1177/1362168820912354.

Lee, I. (2020). Utility of Focused/Comprehensive Written Corrective Feedback Research for Authentic L2 Writing Classrooms. *Journal of Second Language Writing*, *49*(May), 100734. https://doi.org/10.1016/j.jslw.2020.100734.

Mei, W. S. (2001). *Investigating Raters ' Use of Analytic Descriptors in Assessing Writing*. *9*(2), 69–104.

Nicolás–Conesa, F., Manchón, R. M., & Cerezo, L. (2019). The Effect of Unfocused Direct and Indirect Written Corrective Feedback on Rewritten Texts and New Texts: Looking into Feedback for Accuracy and Feedback for Acquisition. *Modern Language Journal*, *103*(4), 848–873. https://doi.org/10.1111/modl.12592.

Phillips, W. A., & Goodall, W. C. (2007). *Lexical Writing can be Non-semantic and It can be Fluent Without Practice*. *March 2015*, 37–41. https://doi.org/10.1080/02643299508251994.

Rouhi, A., & Azizian, E. (2013). Peer Review: Is Giving Corrective Feedback Better than Receiving it in L2 Writing? *Procedia - Social and Behavioral Sciences*, *93*, 1349–1354. https://doi.org/10.1016/j.sbspro.2013.10.042.

Sarré, C., Grosbois, M., & Brudermann, C. (2019). Fostering Accuracy in L2 Writing: Impact of Different Types of Corrective Feedback in an Experimental Blended Learning EFL Course. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2019.1635164.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning1. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129.

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. https://doi.org/10.1002/j.1545-7249.2007.tb00059.

Simard, D., Guénette, D., & Bergeron, A. (2015). *L2 Learners' Interpretation and Understanding of Written Corrective Feedback: Insights from Their Metalinguistic Reflections.8416*(November).https://doi.org/10.1080/09658416.2015.107

6432.

Stefanou, C., & Revesz, A. (2015). Direct Written Corrective Feedback, Learner Differences, and the Acquisition of Second Language Article Use for Generic and Specific Plural Reference. *The Modern Language Journal*, *99*(2), 263–282. https://doi.org/10.1111/modl.12212.

Yang, Y., & Lyster, R. (2010). Effects of Form-focused Practice and Feedback on Chinese EFL Learners Acquisition of Regular and Irregular Past Tense Forms. *Studies in Second Language Acquisition*, *32*(2), 235–263. https://doi.org/10.1017/S0272263109990519.

Zhang, X. (2015). *Reading-Writing Integrated Tasks, Comprehensive Corrective Feedback, and EFL Writing Development*. https://doi.org/10.1177/1362168815623291.

Zhao, H. (2018). Assessment & Evaluation in Higher Education Exploring Tertiary English as a Foreign Language Writing Tutors' Perceptions of The Appropriateness of Peer Assessment for Writing. *Assessment & Evaluation in Higher Education*, *2938*, 1–13. https://doi.org/10.1080/02602938.2018.1434610.