Pegem Journal of Education and Instruction, Vol. 14, No. 4, 2024 (pp. 235 -244)

RESEARCH ARTICLE WWW.PEGEGOG.NET

The Effects of Collaborative Writing on L2 Individual Writing Performance in Terms of Complexity, Accuracy, Fluency and Functional Adequacy

Serpil ÖZDEMİR[1]*

Bartın University, Faculty of Education, Bartın/Turkey

Abstract

In collaborative writing research, writing performance has generally been examined in terms of complexity, accuracy, fluency. Complexity, accuracy and fluency analyzes describe the linguistic structure of the text, but do not provide a sufficient conclusion about its semantic structure. In order to see the semantic structure, the functional adequacy of the texts must also be examined. The effect of collaborative writing on complexity, accuracy, fluency and functional adequacy has been examined in very few studies. Similarly, the effect of collaborative writing on individual writing performance has been examined in several studies. Based on this gap in the literature, the current research was aimed to determine how collaborative writing activities affect L2 individual writing performance in terms of complexity, accuracy, fluency and functional adequacy. The study was conducted in a one group pre-test post-test model. The research was completed in 8 weeks. 21 students at B2 level who were learning L2 Turkish in a preparatory class at a state university participated in the research. As a result of the research, significant differences were seen in the post-test in terms of average sentence length, error word rate, and functional adequacy. However, no significant difference was seen in complex linguistic structures. For both linguistic and semantic development, meaning- and form-focused writing activities should be included in a balanced manner in collaborative studies. In future studies, how individual writing development occurs in students from different language families through collaborative writing can be examined comparatively.

Keywords: Collaborative writing, individual writing performance, CAF, functional adequacy.

Introduction

According to social constructivist theory, learning is a social activity (Vygotsky, 1978) and language is learned not alone but through social interactions (Mitchell & Myles, 1988). The Action-Oriented Approach adopted in the Common European Framework of Reference for Languages sees foreign language learners as social actors who perform an action and function in interaction (Council of Europe, 2001). One of the effective methods for developing writing skills in interaction is collaborative writing (CW). CW is a writing process in which students interact, negotiate meaning, make joint decisions throughout the writing process, and ultimately produce a common text (Storch,2013).

There are many studies showing that CW activities are more effective than individual writing (IW) on writing performance (Avcı Akbaş, 2017; Aldossary, 2021; Chen, 2019; Khatib & Meihami, 2015; Shehadeh, 2011; Storch, 2005; Wigglesworth & Storch, 2009; Winarti & Cahyono, 2020; Zabihi & Bayan, 2020). Furthermore, it is seen that CW is effective in grammar (Chen, 2019; Storch, 2005) and vocabulary learning (Fernández-Dobao, 2014), in improving listening, speaking (McDonough, 2004; Storch, 2005) and communication skills (Vorobel & Kim, 2017), and in increasing attitude, motivation andself-confidence towards L2 learning (Shehadeh, 2011). What is effective in the successful completion of CW studies has been discussed in many studies to date. The interaction process, whether the task is form or meaning focused, task mode, L1 or L2 use, group size, students' prior knowledge about writing, task complexity, and students' attitude towards CW have been considered among the factors that may affect the outcome in CW research. How these elements affect the quality of collaborative writing products is examined in the literature review section.

Corresponding Author e-mail: serpilozdemir34@gmail.com

https://orcid.org/0000-0002-8063-8690

How to cite this ar ticle: ÖZDEMİR S (2024). The Effects of Collaborative Writing on L2 Individual Writing Performance in Terms of Complexity, Accuracy, Fluency and Functional Adequacy. Pegem Journal of Education and Instruction, Vol. 14, No. 4, 2024, 235-244.

Source of support: Nil

Conflict of interest: None

DOI: 10.47750/pegegog.14.04.20

Received: 26.03.2024

Accepted : 07.06.2024 **Publised :** 01.09.2024

Literature Review

Factors effecting in CW performance

Interaction seems to be a very important feature affecting the CW product. Mutual interaction enables students to create ideas about the writing topic and to form ideas together. These processes, in which differences of opinion are discussed and the best opinions to be written are determined, contribute to the creation of a common text. In the studies on interaction/writing processes, language-related episodes were analyzed (Chen & Yu, 2019; Chen, & Hapgood, 2021; Kim & McDonough, 2011; Li, & Kim, 2016; Watanabe & Swain, 2007). The analysis of language-related episodes illuminates the patterns of pair interaction, how the negotiation process is conducted, how the joint text is constructed, how learners support each other, and the outcomes related to language learning. By analyzing peer talk, Storch (2002) defined four different interaction models based on the concepts of "equality", which refers to the level of students' contribution and control over the task, and "reciprocity", which refers to the interaction of peers with each other's contributions: collaborative (high equality and high reciprocity), expert-novice (low equality and high reciprocity), dominant-dominant (high equality and low reciprocity) and dominant-passive (low equality and low reciprocity). In studies based on this interaction model, it has been determined that the frequency of language-related episodes in collaborative groups is high (Watanabe & Swain, 2007), peers at different levels can provide learning opportunities when they work collaboratively (Watanabe, 2008), and there is a moderate relationship between pre-writing discussions and students' written texts (Neumann & McDonough, 2015). In the development of writing performance, it was determined that highly collaborative groups produced better texts in terms of content and coherence than less collaborative groups (Abrams, 2019; Li & Zhu, 2017). In addition, the individual texts written by students who participated in CW activities were found to be more successful than those of students who did not participate in collaborative activities (Chen, 2019; McDonough & De Vleeschauwer, 2019).

One of the factors that can affect the quality of CW products is whether students benefit from L1 or L2 throughout the process. Comparing L1 and L2 use, it has been determined that collaboration in L1 led to higher syntactic complexity, but there is no difference in accuracy, fluency and text quality (Zhang, 2018).

The number of students in CW groups has been addressed as a factor that can affect writing performance. In terms of group size, the performances of students working in pairs and in groups of 3 or 4 students were compared in the CW process. However, the results on this issue are complex. Studies have found that both pairs and groups are useful (Fernandez-Dobao & Blum, 2013), groups have a significant effect on the quality of the written product (Fernández-Dobao, 2014), and the writing performance of pairs is more successful than groups (Winarti et al., 2021). However, studies are generally carried out with pairs (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen, 2019; Villarreal & Martínez-Sánchez, 2023).

Students' prior knowledge about CW affects the quality of the texts they produce. It has been determined that students who receive explicit training in CW produce more fluent, accurate and high-quality texts than students who do not. At the same time, CW knowledge enables benefiting from metacognitive strategies such as planning, monitoring and evaluation. Using metacognitive strategies effects the quality of essays prepared collaboratively (Chen & Ren, 2022).

The effect of task complexity on CW performance has been addressed in research. It was determined that there was no significant difference in terms of complexity, accuracy and fluency, but couples achieved higher scores than individuals in terms of functional adequacy (Zhang, 2022).

One of the factors that affect the quality of CW outcomes is students' attitudes towards CW. Attitude towards CW affects both the CW process and its products. Positive attitudes create motivation to solve language problems, and this motivation increases the quality of writing (Chen & Ren, 2022). Students find CW interesting, entertaining and especially useful in terms of developing grammar, vocabulary and writing skills (Aldossary, 2021; Pham 2021; Shehadeh, 2011). Positive attitudes increase student participation and affect their learning (Chen & Yu, 2019; Fernández-Dobao & Blum, 2013; Storch, 2005). Studies examining student perceptions show that L2 students' attitudes towards CW are generally positive. Language level, task type and mode, learners' beliefs and learning experiences and group dynamics are factors that affect attitudes towards CW (Chen & Yu, 2019; Fernández-Dobao & Blum, 2013; Vorobel, & Kim, 2017). Studies also show that some students have reservations CW work. The reasons for these reservations are worrying about criticizing their peers' writings and hurting their feelings, deficiencies in their second language proficiency (Vorobel, & Kim, 2017; Storch, 2005), some of the group members remaining passive, and some students thinking that they can progress more successfully and faster individually (Storch, 2005).

Collaborative Writing Products and Complexity, Accuracy and Fluency

In studies analyzing CW products, students' linguistic development is usually examined in terms of complexity, accuracy and fluency (CAF). The complexity criterion aims

to determine how much the learner has developed in using linguistically complex structures. It is assessed by criteria such as sentence length, T-units length, and more dependent clauses (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen 2019). Accuracy refers to the accuracy rate of the sentences and T-units in the text, and f luency refers to the total number of words in the text (Storch, 2005). The results of collaborative studies in terms of CAF are quite complex. There are studies showing that CW increases complexity (Bueno-Alastuey & Rodero Albaiceta, 2019; Soleimani & Rahmanian, 2014; Storch, 2005; Villarreal & Martínez-Sánchez, 2023; Zhang, 2018) as well as studies showing that it does not (Dobao 2012; Mozaffari 2017; Wigglesworth & Storch, 2009). There are many studies showing that seeing how peers use language in different situations in the CW process increases accuracy (Chen, 2019; Fernández-Dobao, 2012; Soleimani & Rahmanian, 2014; Storch, 2005; Villarreal & Martínez-Sánchez, 2023), but there are also studies showing that accuracy does not increase (Bueno-Alastuey & Rodero Albaiceta, 2019, Shehadeh, 2011). In terms of fluency, the results are also mixed. Studies showing that CW increase vocabulary learning and thus fluency (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen, 2019, Mozaffari, 2017, Pham, 2021; Storch, 2005; Villarreal & Martínez-Sánchez, 2023) as well as studies showing that it does not (Fernández-Dobao, 2012; Kim, 2008; Shehadeh, 2011; Storch, 2005; Swain & Lapkin, 2001; Wigglesworth & Storch, 2009; Villarreal & Martínez-Sánchez, 2023; Zhang, 2018).

Collaborative Writing Products and Functional Adequacy

The analyses in the CAF framework are based on linguistic criteria, which are useful to see the linguistic development of L2/FL learners, but linguistic accuracy does not mean that the text fulfills all the requirements in terms of meaning. Pallotti (2009) suggested that proficiency should also be measured to determine the extent to which performance fulfills the intended task objectives. Kuiken and Vedder (2017) addressed the issue with the concept of functional adequacy (FA). FA focuses on the quantity, relevance, style and quality of the message that the writer conveys to the reader. Kuiken and Vedder developed an analytical rubric for measuring FA in second language writing. This rubric includes content, task fulfillment, comprehensibility and coherence/consistency criteria. The content criterion is related to the number of information units (ideas or concepts) in the text being sufficient and relevant to the topic. The task fulfillment criterion considers whether the task fulfills all questions and requirements. Comprehensibility measures how much effort the reader has to make to understand the author's purpose and the ideas expressed. The coherence and cohesion criterion is concerned with the text's ability to provide grammatical coherence and cohesion in meaning (Kuiken & Vedder, 2017, p. 326-327). A study examining CW products in terms of both CAF and FA was conducted by Zhang (2022) with 9th grade students. Zhang (2022) examined how individuals and pairs were affected by the role of task complexity in their written products. The texts of students who completed two tasks, one simple (writing a letter with scaffolding) and one difficult (writing a letter without scaffolding), through CW were evaluated in terms of CAF and FA. As a result of the study,it was determined that the collaborative writers produced more functionally adequate and longer texts in the simple task, and they achieved superiority in terms of linguistic accuracy and fluency in the difficult task.

The gap in the literature and the original value of the research

As mentioned above, CW products have been analyzed in various studies. The analyses were generally based on CAF criteria. In previous studies, individual and group outputs were generally examined through experimental (Aldossary, 2021; Khatib & Meihami, 2015; Shehadeh, 2011; Winarti & Cahyono, 2020) and descriptive (e.g., Storch, 2005; Wigglesworth & Storch, 2009; Yang & Polin, 2023; Zabihi & Bayan, 2020) methods. In a limited number of studies, individual texts written by students who engaged in CW activities were examined (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen, 2019; McDonough & De Vleeschauwer, 2019; Villarreal & Martínez-Sánchez, 2023). Bueno-Alastuey and Rodero Albaiceta (2019) analyzed the texts written individually by 14-15 year-old students after CW activities in terms of CAF and lexical diversity. In this study, only one CW was conducted. Chen (2019) analyzed students' individual texts written one week later and 8 weeks later in terms of CAF and quality. In this study, collaborative writing was done 3 times. McDonough and De Vleeschauwer (2019) examined the texts written individually after 3 weeks by students who participated in CW activities 5 times in terms of ratings, accuracy, coordination and subordination and compared them with the writing development of students who did not participate in collaborative activities. Villarreal & Martínez-Sánchez (2023) investigated the immediate and long-term effects of CW on primary school students' texts. In this study, students wrote collaboratively once.

In previous research, CW studies have generally been addressed in English as a foreign language or L2, with pairs and based on a small number of CW activities (Bueno-

Alastuey & Rodero Albaiceta, 2019; Chen, 2019; Villarreal & Martínez-Sánchez, 2023). The current study was conducted with groups of 3, 6 CW tasks were carried out and the participants were learners of L2 Turkish. This research, which deals with CW as treatment, is unique in terms of revealing the effects of intensive CW done by groups in different sociocultural environments.

FA of texts cannot be determined in CAF analyses. Therefore, the FA criterion should also be taken into consideration (Li & Zhang, 2023). Considering these two criteria together allows the CW products to be examined both in the linguistic dimension and in the dimension of fulfilling the communicative function. Thus, the effect of CW can be seen more clearly. The only study that deals with the subject from this perspective has been done by Zhang (2022). One of the unique aspects of the study is that it contributes to the CAF and FA research in collaborative writing, which is very few in the literature.

Aim of the study

The aim of the study is to determine how collaborative writing activities affect foreign language learners' individual writing products in terms of complexity, accuracy and fluency and functional adequacy. The questions guiding the research are as follows:

RQ1. Does collaborative writing affect students' individual writing performances in terms of complexity, accuracy and fluency?

RQ2. Does collaborative writing affect the functional adequacy level of individual writing products?

Method

Research Design

The study was conducted in a single group pre-test and post-test model. In this model, participants are selected in a non-random way and the effect of the program is measured before and after the training. This study was conducted with a group of students who had attended Turkish courses at B2 level and failed, and therefore repeated the same course. It was analyzed whether the students could improve in terms of language use and content design through CW method. In TOMERs, those who come from a lower course and those who repeat the same course usually learn the language in the same class. Since there was not another group repeating the course, a single group pre- and post-test model was selected. The research lasted 8 weeks. A course period is 6 weeks, a pre-test was administered the week before the course started and a post-test was administered the week after the course ended. The research process and data collection tools are presented in detail below.

Procedure and Data Collection Tools

Before starting research ethics committee permission was obtained from Bartın University Social and Human Sciences Ethics Committee (Ethics committee application number 2024-SBB-0037).

In the first week, students were informed about the research. The study was conducted with the students who agreed to participate in the study. The texts written by the students in the final exam of the B2 course, which they failed, were used as a pre-test. The texts written by the students in the B2 proficiency exam were used as a pre-test. In the proficiency exams, a similar question is asked to the topics in the textbook used in the teaching process. In this exam, the subject "The effects of technology on human life" was given. They were asked to explain the subject within the framework of the following questions: What are the benefits of technology? What are the harms of technology? What is the conscious use of technology?" Students wrote advantages and disadvantagesessay.

For the next 6 weeks, CW activities were carried out on the writing topics in the textbook used in the course. Since it is aimed to develop students in terms of writing in different genres and subjects, the writing activities in the textbook also vary. The topics and text types written in the CW process are as follows:

1. Introduce a place you have seen for the student magazine in your school (Descriptive Essay).
2. What do you think friendship is? (DefinitionEssay)
3. What are the benefits and disadvantages of using social media? (Advantage and Disadvantage Essay)
4. Write a critical essay about a culture and art event you attended. (CriticalEssay)
5. Is technology useful or harmful for humans? (Argumentative Essay)
6. Explain the causes and effects of climate change. (Cause and EffectEssay)

The week immediately after the completion of the CW activities (week 8), the B2 proficiency exam was held. The texts written in the B2 proficiency exam were analyzed as a post-test. The same subject was given again in the exam. As a result, the study was completed in 8 weeks.

Study Group

The study was conducted with 21 B2 level Turkish learners at a state university. The students are continuing their preparatory education as they have not passed the language proficiency exam, which is compulsory to start university education. Seven of the students came from various countries

in Asia and 14 from Africa. Their mother tongues are Arabic, Russian, Kazakh, Tajik, and various African languages. Their ages are between 18 and 26.

The students in the study group have successfully completed levels A1, A2, B1. They have taken courses at level B2, but failed in the proficiency exam. Courses at each level lasted 6 weeks. The students' Turkish learning history is 24 weeks in total. Since there were no students who were successful in B1 and continued to B2, there were only students in the group who took the same course again. Since this situation is very rare, it was not possible to compare the results of the research with another group with similar characteristics.

The students participating in the study were informed about the research and participated voluntarily. Ethics committee permission was obtained from Bartın University Social and Human Sciences Ethics Committee (Ethics committee application number 2024-SBB-0037).

Data Analysis

The data were analyzed in terms of CAF and FA. For the analyses in terms of CAF, firstly the number of words with and without errors, T-units, dependent clauses, and total clauses were determined. Complexity was analyzed at the sentence level. Mean clause length, mean T-unit length, and clauses per T-unit were used to determine complexity. In determining accuracy, the proportion of error words, the proportion of error-free clauses (error-free clauses/total clauses) and the proportion of error-free T-units (error-free T-units/total T-units) were analyzed. The number of words was taken into account in determining fluency.

In determining FA, the FA scale, whose reliability studies were conducted by Kuiken and Vedder (2017), was used. The scale is Likert-type and consists of four dimensions. The dimensions of the scale are content, task fulfillment, comprehensibility, consistency and coherence. Each dimension is scored between 1-6 points. Reliability studies of the scale were conducted with university students learning L2 Dutch and L2 Italian. The reliability scores for Dutch are between .824-.940 for all dimensions and .725-.901 for Italian. In the reliability analyses for Turkish in this study, correlations ranging from .838 to .945 were determined. Correlations for all criteria are presented in Table 1 under the heading of reliability of the study.

Whether the CAF and FA scores showed significant differences in the pre-test and post-test was analyzed with paired groups t-test. When the normality of the data set was examined, it was seen that the CAF data were not normally distributed, while the FA data were normally distributed. Based on this, Wilcoxon Signed Ranks Test was used for CAF, and Paired Sample t test, one of the parametric tests, was used for FA.

Reliability

Reliability Study for CAF: After the texts were coded bythe researcher, they were also coded by another field expert. Miles and Huberman (1994) formula [Agreed errors / (agreed errors + disagreed errors) x100] was used to determine the level of agreement between the coding of the two experts. Agreement was 98% for the word with error, 96% for T-units, 90% for dependent clauses, and 100% for total clauses. Disagreements were discussed and a consensus was reached.

Reliability Study for FA: FA scoring was done by both the researcher and another expert in the whole data set. The reliability of the inter-expert scoring was determined by Pearson Correlation. As a result of the analysis, the correlation coefficients determined in the pre-test and post-test are presented in Table 1.

Table 1. *Inter-rater Correlation of FA Scores*

Dimension	*Pre-test*	*Post-test*
Content	.838**	.842**
Task fulfillment	.930**	.945**
Comprehensibility	.862**	.870**
Coherence and cohesion	.905**	.920**

**p<.0

As seen in Table 1, there are high and significant (p<.01) relationships between the scores of the two experts in all dimensions, ranging from .838 to .945.

Findings

In this section, the results obtained are presented in line with the research questions.

RQ1. Does CW affect students' IW performances in terms of complexity, accuracy and fluency?

Wilcoxon Signed Ranks Test was conducted to determine whether students' IW performances showed significant differences in terms of complexity, accuracy and fluency in the pre-test and post-test. The analysis of complexity is presented in Table 2, accuracy in Table 3 and fluency in Table 4.

Table 2. *Pre-test and Post-test for Complexity*

Comp-lexity	Pretest-Posttest	N	Mean Rank	Sum of Ranks	Z	p
Mean clause length	Negative Ranks	5	9.80	49	-2.311	021
	Positive Ranks	16	11.38	182		
	No difference	0				
Mean T-unit length	Negative Ranks	8	9.26	66,50	-.078	.938
	Positive Ranks	8	8.11	69,50		
	No difference	0				
Clauses per T-unit	Negative Ranks	5	4.60	23	-.460	.646
	Positive Ranks	5	6.40	32		
	No difference	6				

As seen in Table 2, in terms of complexity, there is an increase in the average sentence length in 16 students and a decrease in 5 students in the post-test, the difference is statistically significant ($z = 2.311$, $p < .05$). Average T-unit length decreased in 8 students in the post-test, while it increased in the other 8 students, the difference was not significant ($z = .078$, $p > .05$). In terms of clauses per T-unit 5 students decreased, 5 students increased, and 6 students maintained the same level in the post-test. The difference between the scores is not statistically significant ($z = 460$, $p > .646$).

Table 3. *Pre-test and Post-test for Accuracy*

Accuracy	Pretest-Posttest	N	Mean Rank	Sum of Ranks	Z	p
Rate of incorrect words	Negative	17	11.06	188	-2.520	.012
	Positive	4	10.75	43		
	No difference	0				
Error-free clauses/ total clauses	Negative	12	10	120	-.156	.876
	Positive	9	12.33	111		
	No difference	0				
Error-free T-units/ total T-units	Negative	9	7.50	67.5	-.950	.342
	Positive	5	7.50	37.5		
	No difference	2				

In terms of accuracy, the rate of incorrect words decreased for 17 students and increased for 4 students in the post-test. The difference is significant in favor of the post-test ($z = 2.52$, $p < .05$). The rate of error-free clauses decreased for 12 students and increased for 9 students in the post-test. However, the difference between pre and post-test averages is not significant ($z = .156$, $p > .05$). Error-free T-units ratio decreased in 9 students, increased in 5 students and remained the same in 2 students. There is no significant difference between the averages ($z = .950$, $p > .05$).

Table 4. *Pre-test and Post-test for Fluency*

Fluency	Pretest-Posttest	N	Mean Rank	Sum of Ranks	Z	p
Text length	Negative	5	7.80	39	-2.659	.008
	Positive	16	12	192		
	No difference	0				

In the post-test, text length decreased for 5 students and increased for 16 students. Considering the rank average and total of the difference scores, the difference is significant in favor of positive ranks ($z = 2.659$, $p < .05$).

As a result, in terms of CAF criteria, it was seen that CW was only effective in increasing the average sentence length in terms of complexity, decreasing the number of wrong words in terms of accuracy, and increasing in terms of fluency, but was not effective in other criteria.

RQ2. Does CW affect the functional adequacy level of IW products?

The *t* test conducted to determine whether the FA level of the individual products of the students participating in CW showed a significant difference in the pre-test and post-test is presented in Table 5.

Table 5. *FA Difference in Pre-Test and Post-Test*

	Pretest-Posttest	*N*	*Mean*	*S*	*df*	*t*	*p*
Content	Pre-test	21	4.05	1.16	20	-2.23	.038
	Post-test	21	4.52	1.17			
Task fulfillment	Pre-test	21	4.23	1.09	20	.498	.624
	Post-test	21	4.10	1.34			
Comprehensibility	Pre-test	21	3.29	.902	20	-4.81	.000
	Post-test	21	3.91	.889			
Coherence/cohesion	Pre-test	21	2.81	1.12	20	-4.66	.000
	Post-test	21	3.71	1.10			
Composite score	Pre-test	21	14.38	3.50	20	-2.74	.013
	Post-test	21	16.24	3.62			

In terms of content ($t20 = 2.23$, $p < .05$), comprehensibility ($t20 = 4.81$, $p < .05$), and coherence / consistency ($t20 = 4.66$, $p < .05$), the means are high in the post-test and thedifferences are significant in favor of the post-tests. However, in terms of task fulfillment, the mean is low in the post-test and the difference between the means is not significant ($t20 = .498$, $p > .05$). There is a significant difference in the composition score, which is the sum of all criteria, in favor of the post-test ($t20 = 2.74$, $p < .05$). This result shows that CW is effective in terms of FA.

DISCUSSION

The aim of this study was to determine the effects of CW activities on IW performance. IW performance was evaluated in terms of language use and FA of texts. Language use was analyzed in terms of complexity, accuracy and fluency, and FA was analyzed in terms of content, task fulfillment, comprehensibility and coherence/consistency. As a result of the study, it was determined that there was a significant difference in the post-test only in the average sentence length among the complexity criteria in terms of language use, while there were no significant differences in the average T-unit and sentence per T-unit ratios, which are the other criteria of complexity. In the accuracy dimension of language use, a significant difference was observed only in the error-free word rate criterion, while no significant differences were observed in the error-free clauses and error-free T-units rates. In the fluency dimension, there was a significant difference in favor of the post-test. In terms of FA, there were significant differences in favor of the post-test in the content, comprehensibility and coherence/consistency criteria, but there was no significant difference in the task fulfillment criterion. However, when all criteria were considered together, the difference in FA scores was significant in favor of the post-test. These results are discussed below.

CAF Results

In terms of CAF criteria, it was seen that CW had very limited effects on IW perfo0rmance. In terms of complexity, the average sentence length increased in the post-test, but only in simple sentences. There were no significant differences in the average T-unit and the ratio of sentences per T-unit. A T-unit is defined as an independent clause and all its attached or embedded dependent clauses (Storch, 2005, p. 171). In order to create T-units, students need to use more complex structures of the language. In studies examining individual texts written immediately after CW activities, it has seen that CW is effective in producing complex texts (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen, 2019; Villarreal & Martínez-Sánchez, 2023). However, L2/FL learning history of the students in the study group of these studies is longer than the participants of the current study. In these studies, the English learning history of the participants varies between 2 and 7 years. English is widely demanded in the world, so EFL teaching starts from primary school in many countries. However, the participants in this study were exposed to Turkish at university level for the first time. Turkish learners' lack of development in using complex structures may be the result of their late exposure to the language. The result of this study coincides with the research conducted by McDonough and De Vleeschauwer (2019) in the Thai EFL context. The commonality between this study and the current study is that learners from different L1 families have difficulty in making progress in complex linguistic structures.

In terms of accuracy, a significant difference was observed only in the error-free word rate in the post-test, but no significant differences were observed in the error-free clauses and error-free T-units rates. The decrease in the number of incorrect words shows that the ability to choose the correct word has improved. However, students could not use grammatical elements correctly in complex sentences. This result is in contrast to many studies showing that accuracy increases in IW activities immediately following CW practices (Chen, 2019; McDonough & De Vleeschauwer, 2019; Villarreal & Martínez-Sánchez, 2023). In error analyses, incorrect arrangement of words in sentences, missing words, and incorrect writing of suffixes are taken into consideration. Turkish is an agglutinative language. Articles, prepositions and verb tenses in English are expressed with suffixes in Turkish and verb conjugation varies depending on the subject. Some linking elements that are separate words in English, such as "because", can be expressed in Turkish both with a word with a conjunction function (çünkü) and with a suffix + word form (-DIğı için, -DIğIndAn dolayı). The Turkish learning history of the participants in the study was 24 weeks at the beginning of the experimental process and reached a total of 30 weeks at the end of the B2 course. Students taking 6 weeks course at each level and 24 hours lessons per week. It should be taken into consideration that this time may be insufficient to learn a language from a different language family at B2 level. In addition, the fact that the study group consisted of students repeating the B2 level also shows that students are not sufficient in language learning.

In terms of fluency, a significant difference was observed in favor of the post-test in the texts written individually by the students who participated in CW activities. Fluency was measured by the number of words used in the text. The increase in the average sentence length in terms of complexity and the decrease in the rate of incorrect words in terms of accuracy are related to the new words learnt. When these results are evaluated together in terms of CAF, it is seen that CW supports vocabulary learning. This result is in parallel with previous research (Bueno-Alastuey & Rodero Albaiceta, 2019; Chen, 2019; Mozaffari, 2017; Pham, 2021; Storch, 2005; Villarreal & Martínez-Sánchez, 2023).

A satisfactory result could not be reached from a linguistic point of view. This result may be due to the fact that the experimental process was conducted with a focus on meaning. A form-focused CW process may be more effective in providing linguistic development. In form-focused

collaborative writing, students will focus more on language, collaborate to solve language-relate problems (Fernández-Dobao, 2014) and transfer their individual knowledge to each other (Garcia Mayo,2002).

FA Results

In terms of FA, the differences in all criteria (content, comprehensibility and coherence/consistency) except the criterion of task fulfillment and in the total composition score were found significant in favor of the post-test. This result shows that CW activities positively affect FA in individual writing. In this study, an extra task structure was presented both in the experimental process and in the pre- and post-tests, that is, writing activities were carried out with simple tasks. Zhang (2022) compared the texts of students writing collaboratively and individually in terms of FA in simple and complex tasks. As a result of the study, it was determined that the overall FA, content, and comprehensibility scores of the pairs were significantly different in the simple task, while there was no difference in any criterion in the complex task. This study supports the conclusion that FA increases in simple tasks.

Villarreal and Martínez-Sánchez (2023) found that CW did not affect the overall quality of elementary school students' individual texts. In the mentioned study, only one CW study was conducted. In current research, students worked collaboratively six times. In the repeated process, they had the opportunity to see different strategies from each other on how to structure the text. This result may be related to the repetition of collaborative work for a long time. In addition, the age of the students may also be effective. This study was conducted with university students. Adult students have more prior knowledge about how to organize content semantically. Chen (2019), who conducted a study with university students, found that the organization of students who wrote individually immediately after the collaborative work significantly outperformed the organization of students who did not do collaborative work.

As described in the literature review, there are many variables that can affect the qualities of writing products in collaborative work. One of these is the intra-group interaction model. Having a high equality and high reciprocity model of intra-group interaction, in other words, high cooperation, increases content and consistency scores (Abrams, 2019; Li & Zhu, 2017). During the experimental process of this research, students were encouraged for high cooperation. Individual and group scores were given after CW and students with the highest scores were rewarded. As a result, students tried to develop ideas and organize their ideas to write better each time. It is thought that high cooperation had an impact on the positive result obtained in terms of semantic organization of the text.

CONCLUSION

The results of the research show that CW provides lexical development in individually written texts and supports FA immediately after collaborative writing. However, there is no significant difference in terms of using complex linguistic structures.

The results of the research highlight some important points from a pedagogical perspective and for future research. The motivation for this research was the desire to see the effect of CW on improving the writing skills of students who failed at the B2 level. However, a satisfactory result could not be achieved in terms of language use. The experimental process of the research was meaning-orientated. In order for linguistic and semantic development to go together, meaning-focused and form-focused writing activities should be included in a balanced manner when doing collaborative works in L2/FL classes. In addition, it was seen that working with writing topics provided with scaffolding was effective in structuring meaning in the development of FA. For students who have difficulty in writing, the subject should not be presented only as a topic sentence, but should be supported with scaffolding. Individual and group scores were given in each study in order to prevent students from being passive in CW studies and to ensure high cooperation. For this purpose, individual and group evaluation rubrics were created. The best students and groups were rewarded. It was observed that this practice increased motivation and participation. For this reason, it is recommended to make observations through rubrics and give awards to increase the performance of students in the teaching process.

It is suggested that the experimental process of future research on CW should be longer. In studies conducted on this subject, texts written individually as a result of a very small number of CW practices have generally been examined. It is anticipated that students can benefit more from peer learning in the process. The results of short and long-term CW practices should be examined comparatively. In this study, the lack of sufficient linguistic development was associated with the fact that the students' mother tongue and the foreign language they learned were from different languagefamilies. However, this is one of the issues that need to be examined comparatively.

It is anticipated that students will benefit more from peer learning during the process. The results of short and long-term CW studies should be investigated comparatively. In this study, the lack of sufficient linguistic development was associated with the fact that the students' native languages and foreign language were from different language families. However, this should also be examinedcomparatively.

REFERENCES

Abrams, Zs. I. (2019). Collaborative writing and text quality in Google Docs. *Language Learning & Technology, 23*(2), 22-42. https://doi.org/10125/44681

Aldossary, K. S. (2021). The impact of collaborative writing on EFL learners' writing development: A longitudinal classroom-based study in Saudi Arabia. *Arab World English Journal, 12*(3), 174-185. https://dx.doi.org/10.24093/awej/vol12no3.12

Avcı Akbaş, T. (2017). *The effects of collaborative work on students' writing performance and students' perceptions of collaborative writing*. (Publication No. 452519). [Ubpublished Master Thesis, Abant Izzet Baysal Unıversıty]. Council of Higher Education Thesis Center.

Bueno-Alastuey, M. C., & Rodero Albaiceta, S. (2019). The effects of using collaborative writing vs. peer review treatments on subsequent individual writings. *Huarte De San Juan. Philology and Didactics of Language, 19*(2019), 32-61. Retrieved from http:// revista-hsj-filologia.unavarra.es/article/view/2664

Chen, W. (2019). An exploratory study on the role of L2 collaborative writing on learners' subsequent individually composed texts. *The Asia-Pacific Education Researcher, 28*(6), 563-573. https://doi.org/10.1007/s40299-019-00455-3

Chen, W., & Ren, W. (2022). Educating L2 learners about collaborative writing: exploring the relationship between knowledge of collaborative writing and writing products. *Language Awareness, 31*(3), 371-391. https://doi.org/10.1080/09658416.2021.1969403

Chen, W., & Hapgood, S. (2021). Understanding knowledge, participation and learning in L2 collaborative writing: A metacognitive theory perspective. *Language Teaching Research, 25*(2), 256-281.https://doi.org/10.1177/1362168819837560

Chen, W., & Yu, S. (2019). A longitudinal case study of changes in students' attitudes, participation, and learning in collaborative writing. *System, 82*(2019), 83-96. https://doi.org/10.1016/j.system.2019.03.005

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Retrived from https://rm.coe.int/1680459f97

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering L2 development and mastery of writing conventions. *Language Learning & Technology, 14*(3), 51-71. http://dx.doi.org/10125/44226

Fernández-Dobao, A. (2014). Attention to form in collaborative writing tasks: Comparing pair and small group interaction. *Canadian Modern Language Review, 70*, 158–187.

Fernández-Dobao, A. (2020). Collaborative writing in mixed classes: What do heritage and second language learners think? *Foreign Language Annals, 53*(1), 48-68. https://doi.org/10.1111/ flan.12446

Fernández-Dobao, A., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System, 41*(2), 365-378. https://doi.org/10.1016/j.system.2013.02.002

Garcia Mayo, M. P. (2002). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics, 12*(2), 156-175. http://dx.doi.org/10.1111/1473- 4192.t01-1-00029

Khatib, M., & Meihami, H. (2015). Languaging and writing skill: The effect of collaborative writing on EFL students' writing perfor- mance. *Advances in Language and Literary Studies, 6*(1), 203-211.

Kim, Y., & McDonough, K. (2011). Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research, 15*(2) 183-199. https://doi. org/10.1177/1362168810388711

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*(3), 321-336. https://doi.org/10.1177/0265532216663991

Li, M., & Kim, D. (2016). One wiki, two groups: Dynamic interactions across ESL collaborative writing tasks. *Journal of Second Language Writing, 31*(2016) 25-42. https://doi.org/10.1016/j. jslw.2016.01.002

Li, M., & Zhang, M. (2023). Collaborative writing in L2 classrooms: A research agenda. *Language Teaching, 56*(1), 94–112. https:// doi.org/10.1017/S0261444821000318

Li, M., & Zhu, W. (2017). Good or bad collaborative wiki writing: Exploring links between group interactions and writing prod- ucts. *Journal of Second Language Writing, 35*(2017), 38-53. https://doi.org/10.1016/j.jslw.2017.01.003

McDonough, K. & De Vleeschauwer, J. (2019). Comparing the ef- fect of collaborative and individual prewriting on EFL learn- ers' writing development. *Journal of Second Language Writing, 44*(2019), 123-130. https://doi.org/10.1016/j.jslw.2019.04.003

McDonough, K. (2004). Learner-learner interaction during pair and small group activities in a Thai EFL context. *System, 32*(2004), 207-224. https://doi.org/10.1016/j.system.2004.01.003

Miles, M. B., & Huberman, A. M. (1994). Q*ualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.

Mitchell, H., & Myles, F. (1988). S*econd language learning theories*. Arnold.

Neumann, H., & McDonough, K. (2015). Exploring student interaction during collaborative prewriting discussions and its relationship to L2 writing. J*ournal of Second Language Writing, 27*(2015) 84-104. https://doi.org/10.1016/j.jslw.2014.09.009

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590-601. https://doi. org/10.1093/applin/amp045

Pham, V. P. H. (2021). The effects of collaborative writing on students' writing fluency: An efficient framework for collaborative writing. *SAGE Open, 11*(1), 1-11. https://doi. org/10.1177/2158244021998363

Shehadeh, A. (2011). Effects and student perceptions of collaborative writing in L2. *Journal of Second Language Writing, 20*, 286- 305. https://doi.org/10.1016/j.jslw.2011.05.010

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning, 52*(1), 119-158. https://doi.org/10.1111/1467-9922.00179

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*(2005) 153-173. https://doi.org/10.1016/j.jslw.2005.05.002

Storch, N. (2013). *Collaborative writing in L2 clasrooms.* Multilingual Matters.

Villarreal, I., & Martínez-Sánchez, A. (2023). Exploringimmediate and prolonged effects of collaborative writing on young learners' texts: L2 versus FL. *International Review of Applied Linguistics in Language Teaching.* https://doi.org/10.1515/iral-2022-0062

Vorobel,O.,&Kim,D.(2017).AdolescentELLs'collaborativewriting practices in face-to-face and online contexts: From perceptions to action. *System, 65*(2017), 78-89. https://doi.org/10.1016/j.system.2017.01.008

Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*. Harvard University Press.

Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research, 11*, 121-42. https://doi.org/10.1177/1362168806074599

Watanabe, Y. (2008). Peer-peer interaction between L2 learners of different proficiency levels: Their interactions and reflections. *The Canadian Modern Language Review, 64*(4), 605-663. https://doi.org/10.3138/cmlr.64.4.605

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445-466. https://doi.org/10.1177/0265532209104670

Winarti, W., & Cahyono, B. Y. (2020). Collaborative writing and process writing approach: The effect and students perception. *Journal of English Educator Society, 5*(2), 163-169. https://doi. org/10.21070/jees.v5i2.773

Winarti., Cahyono, B. Y., Mukminatien, N., & El Khoiri, N. (2021). Collaborative writing using process writing approach: The effect of group size and personality types. *International Journal of Instruction, 14*(4), 391-410. Retrieved from https://files.eric. ed.gov/fulltext/EJ1319104.pdf

Yang, L., & Polin, L. G. (2023). Exploring the learning benefits of collaborative writing in L2 Chinese: a product-oriented perspective. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2023-0034

Zabihi, R., & Bayan, M. (2020). Are two voices better than one? Comparing aspects of text quality and authorial voice in paired and independent L2 writing. *Written Communication, 37*(4), 512-535. https://doi.org/10.1177/0741088320939542

Zhang, M. (2018). Collaborative writing in the EFL classroom: The effects of L1 and L2 use. *System, 76*(2018), 1-12. https://doi. org/10.1016/j.system.2018.04.009

Zhang, M. (2022). Collaborative writing in an EFL secondary setting: the role of task complexity. *International Review of Applied Linguistics in Language Teaching,* https://doi.org/10.1515/iral-2022-002