Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
40(4), Fall 2021, pp. 121-160
DOI: 10.22099/JTLS.2021.39939.2954

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

The Effect of Dictogloss vs. Debating on L2 Writing Proficiency: A Mixed-Methods Study

Ghasem Modarresi *

Department of English, Quchan Branch, Islamic Azad University, Quchan, Iran

Abstract

The use of task-based collaborative output activities in enhancing EFL students' writing proficiency has been underrated in the Iranian context. To foster students' writing ability, the present study, employing a mixed-methods design, aimed to evaluate the effect of innovative tools, dictogloss and debating, on the writing development of English-major university students in terms of complexity, accuracy, and fluency (CAF). The study involved two experimental extracurricular classes and consisted of 11 sessions during the regular academic term. Having analyzed the writing tasks produced by the students' performance on pre-test and post-test, the researcher found that the students of both groups significantly enhanced their writing performance, representing an increase in a number of indices of CAF measures following the intervention. More specifically, the results of a paired-samples t-test confirmed that the students in the dictogloss group showed significant improvement in six indices of CAF measure, and the students in the debate group displayed significant improvement in seven indices of CAF measures. Moreover, the results of ANCOVA confirmed that the debate group improved more than the dictogloss group in terms of CAF measures. The results of the interviews with the students regarding the role of task-based collaborative output activities in their written performance yielded several commonalities, which were coded into 11 codes for dictogloss and 16 codes for debating, taking inter-coder

Received: 28/02/2021 Accepted: 07/07/2021

* Assistant Professor, Email: qasem.modarresi@gmail.com

reliability and agreement into account. In the end, the study offers some practical implications for L2 learners and teachers.

***Keywords**:* Output-based Instruction, Collaborative Activities, Complexity, Accuracy, Fluency

Although the English language is acknowledged as the language for international communication, it is the written English that is considered a leading media for much of this communication mainly due to the fact that most of the communication through the internet, as the most powerful communication tool, is written (Kroll, 2003). According to Crowhurst (1991), the prevalent perspective in a Second Language (L2) context is that writing mostly serves to manifest patterns of oral language use, syntax, and semantics. Indeed, successful creative writing is taken as a required 21st-century skill for academic accomplishment. Meanwhile, Persky et al. (2003) remarked that writing is commonly used in many meaningful communicative activities like producing academic essays, writing business or newspaper reports, or web pages and e-mails. Likewise, considering the vast expansion of globalization worldwide communication, Naghdipour (2016) reaffirmed the need to be proficient in English writing skills.

Nevertheless, according to Qin and Uccelli (2016), writing is a highly challenging task that requires cognitive processing and sociolinguistic knowledge of written communication. For this reason, Kanakri (2016) declared that students should gain mastery over the linguistic and rhetorical schemes of writing in order to develop their writing ability. Actually, EFL teachers perform a significant role here since students want their help to acquire and apply innovative instructions and techniques to their writing so that they can gradually gain communicative competence (Ferris & Hedgcock,

2004). In this regard, Polio and Park (2016) call for novel instructions for L2 writing pedagogy to guide students through the writing process more effectively.

Meanwhile, to make writing classes more creative and communicative, previous relevant research acknowledges the need to substitute the traditional modes of teaching such as presentation, practice, and production (PPP) with task-based language teaching (TBLT) method (e.g., Dobao, 2012; Ellis, 2008). However, Golparvar and Rashidi (2021) argue that much of the related literature in task-based instruction has paid particular attention to task characteristics in oral production rather than written production. Moreover, according to Allen (2018), L2 task-based writing instruction has not adequately considered the growing importance of writing in English. Hence, a fruitful step forward seems to be the collaboration between collaborative task-based instruction and L2 writing development. The present study explores this synergy by examining the impact of collaborative task-based instruction on writing proficiency. In this regard, of the most salient task-based instructions, this study opts for two collaborative task-based instructions, including dictogloss and debate. According to Benati (2017), dictogloss, as a task-based collaborative output activity, could enable learners to develop their L2 writing proficiency and help students to utilize their grammar resources to reconstruct a text.

Actually, dictogloss is considered an instructional technique that draws students' attention to the function of grammar at the level of discourse (Nassaji, 2016; Wajnryb, 1990). Moreover, as another task-based collaborative output activity, debate-based instruction is considered an effective tool for L2 learning (Lustigova, 2011). It seems that both dictogloss and debate interventions subscribe to process-oriented instruction rather than

product-oriented instruction, during which the learners are required to go through some specific steps before producing a written product. Process writing provides learners with an enriched learning experience (Yong, 2010), and task-based collaborative output activities enable learners "to go through the process of writing, feedback processing and revision" (el Majidi et al., 2020, p. 806). According to Swain (2005), learners attend to both meaning and forms in collaborative output activities.

As for the importance of L2 proficiency, previous research has associated this concept with the trial notions of CAF as an indicator of L2 development (Ellis, 2008) and of writing development (Johnson, 2017). According to Housen and Kuiken (2009), these factors can be used "as performance descriptors for the oral and written assessment" (p. 462). Similarly, a number of professionals in SLA conceptualize writing performance in terms of CAF (e.g., Frear & Bitchener, 2015; Johnson, 2017). In this regard, Pourdana et al. (2011) found that TBLT could enhance intermediate Iranian EFL learners' writing ability in terms of accuracy and complexity. Moreover, writing development in terms of CAF is related to emotioncy. It is held that background knowledge, senses, and emotions can impact cognition and understanding, and senses can relativize cognition (Pishghadam, Jajarmi, & Shayesteh, 2016). For instance, according to Pishghadam, Adamson and Shayesteh (2013), words towards which learners express higher levels of emotions can be learned easier. For this reason, sense-induced emotions could enrich learners' lexical complexity.

However, a number of researchers argue that achieving L2 writing proficiency in terms of CAF is a highly demanding task (e.g., Leki, Cumming, & Silva, 2010; Skehan, 2013). Meanwhile, Polio and Park (2016) highlight the role of intervention in enhancing writing ability in terms of CAF. Actually,

to the best knowledge of the researcher, there is a severe lack of research to examine the effect of task-based collaborative output activities on writing proficiency in terms of CAF in our classroom context. Actually, delving into the indices of CAF measures in written performance has been underrated in the Iranian context. In this study, writing proficiency is defined as the students' mastery over aspects of writing development entailing syntactic complexity, lexical complexity, accuracy, and fluency.

Literature Review

Task-based Instruction and Writing Skills

Previous research investigating the factors associated with task-based writing instruction has mainly focused on task complexity and task types (Skehan, 2003; Robinson, 2011; Zalbidea, 2017). In line with Skehan's (2003) Limited Attentional Capacity Model and Robinson's (2005) Cognition Hypothesis, Alexopoulou et al. (2017) concluded that the cognitive complexity of a task affects writing proficiency. Moreover, task type has been documented to have a determining effect in writing development (Lu, 2011; Yoon & Polio, 2016). However, Zohrabi and Hassanpoor (2020) investigated the effectiveness of open vs. closed tasks on enhancing EFL learners' oral performance, and they found no significant difference between open and closed tasks in improving learners' oral ability. Yoon and Polio (2016) found that the functional differences between task types are more noticeable in comparison to the differences in cognitive complexity.

In the Iranian context, Ganji and Ketabi (2015) examined the effect of collaborative output tasks on the learning of English lexical collocations and the obtained results indicated that accomplishing the tasks collaboratively could extend knowledge of collocation more than completing the tasks

individually. Talebinezhad and Esmaeili (2012) examined the impact of three types of tasks, including dictation tasks, individual reconstruction tasks, and collaborative tasks on the acquisition of gerunds and infinitives and they found that the group exposed to collaborative tasks outperformed the two other groups with respect to the grammatical structures. More recently, Golparvar and Rashidi (2021) found that task complexity substantially impacted some measures of syntactic complexity, lexical diversity, and causal cohesion.

A large number of published studies acknowledge the value of collaborative tasks in fostering learning development (e.g., Dörnyei, 2019; Storch & Wigglesworth, 2007). For instance, Lee (2001), employing collaborative output tasks, concluded that learners find collaborative tasks satisfactory. To be more specific, Collins (2007) found Dictogloss, as a task-based collaborative output technique, useful in dealing with verb tenses in a Japanese classroom. Similarly, Kuiken and Vedder (2002) confirmed the effectiveness of dictogloss tasks in learning passive forms. Although Dictogloss has been proved to be effective in L2 research (Malmqvist, 2005), the number of studies that have explored its applicability with respect to proficiency level is not many (García Mayo, 2002; Fortune, 2005). Recently, Gallego (2014) carried out a multilevel analysis of the learners' perceptions on the usefulness of dictogloss. She concluded that most of the students, especially advanced-low level students, found it both effective and useful. More recently, Dehghan and Mohammadi Amiri (2017) investigated the effect of two types of tasks, including dictogloss and text reconstruction editing, on learning English comparative adjectives, and they found that the text reconstruction editing group had a better performance than the dictogloss group in learning the English comparative adjectives.

Previous studies have also documented that there is a positive association between debating and language development (Steward, 2003; el Majidi et al., 2018). Moreover, the related literature acknowledges that L2 learners have a positive attitude towards debating as a pedagogical tool (Lustigova, 2011; Doody & Condon, 2012). For example, Lustigova (2011) found that more than half of the learners were willing to take part in the debate course in the upcoming terms. In this regard, Doody and Condon (2012) stated that instructional strategies that stir up active engagement offer significant advantages for the learners. Snider and Schnurer (2006) focused on the dual function of debate as a performance and a method to express ideas. They concluded that debate is a communication event including both oral and written modes of operation. In this respect, Cho (2017) emphasized the role that these oral interactions and negotiations could have in writing development.

CAF: Dimensions and Experimental Research

According to Housen and Kuiken (2009), important studies in SLA have focused on the association between the triad of CAF and language proficiency and development. In this regard, Barrot (2018) declared that, as a rather new concept in applied linguistics, CAF has already attracted the attention of some scholars in SLA (e.g., Skehan, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). Meanwhile, Norris and Ortega (2009) proposed an organic approach to exploring CAF in L2 learning, highlighting the dynamic nature of CAF. Later, Lu (2011) investigated different measures of syntactic complexity and offered ESL teachers remarkable insights into the implementation of these measures as indices of writing proficiency for university students. The three dimensions of CAF are briefly reviewed below.

As the first known dimension of L2 proficiency, complexity mainly consists of syntactic complexity (Norris & Ortega, 2009) and lexical complexity (McNamara, Crossley, & McCarthy, 2010). As commented by Skehan (2009), complexity manifests the level of advanced language, including the use of subordination. There is no anonymous agreement on the exact metrics and indices for complexity measurement. For instance, Wolfe-Quintero et al. (1998) proposed that clauses per T-unit assess complexity, that is, the number of dependent clauses per total clauses and the number of dependent clauses per T-unit. Meanwhile, McNamara, Crossley and McCarthy (2010) proposed word length (WL) and word frequency (WF) as the two important measures of lexical complexity, and Malvern et al. (2004) used the type-token ratio (TTR) as an index of lexical complexity that refers to the number of different words divided by the total number of words in a text.

As the second known dimension of L2 proficiency, accuracy is associated with the ability to be free from errors while producing language (Skehan, 2009). Again, no fixed metrics have been proposed by scholars as indices for measuring accuracy. In their seminal work, Wolfe-Quintero et al. (1998) proposed that accuracy can be evaluated via calculating the number of EFT, error-free T-units per T-unit (EFT/T), and errors per T-unit (E/T). In this regard, Nakatani (2010) suggested that the most effective way to measure accuracy is by calculating the number of global errors (i.e., those that lead to misunderstanding) and local errors (i.e., minor errors that do not cause communication breakdown).

As the third known dimension of L2 proficiency, fluency is referred to as "the capacity to produce speech at a normal rate and without interruption" (Skehan, 2009, p. 510). Wolfe-Quintero et al. (1998) declared that fluency can

best be measured in writing performance by means of T-unit length, error-free T-unit (EFT) length, and clause length. Lu (2011) proposed that longer production which is related to fluency, can represent a higher level of proficiency. Meanwhile, Skehan (2003) emphasized the distinction between breakdown fluency, speech fluency, and repair fluency, measured via silence-related, time-related, and self-correction metrics, respectively.

Much of the greater part of the literature acknowledges the effectiveness of task-based output activities on CAF measures with respect to written performance (e.g., Barrot, 2018; Norris & Ortega, 2009). For example, Norris and Ortega (2009) found that the frequent use of coordination represents a lower level of proficiency while frequent use of subordination represents a higher level of proficiency. In the same vein, Wigglesworth and Storch (2009) examined the impact of pair and individual writing on CAF dimensions, and they concluded that collaborative writing significantly displayed an increase in accuracy but not in complexity and fluency. Furthermore, Ellis and Yuan (2004) investigated the impact of planning on CAF and concluded that pre-task planning improves learners' fluency and complexity. More recently, el Majidi et al. (2020) examined the role of debate, as an L2 pedagogical tool in writing development, in the context of secondary school and they confirmed that debate-based instruction increased aspects of writing proficiency in terms of CAF.

Taken together, following the guidelines suggested by Swain's (1993) output hypothesis based on which output enables learners to produce language more deeply, the current study examines the impact of debating vs. dictogloss on L2 writing development among intermediate EFL learners. The study hypothesizes that task-based collaborative output activities could enhance

writing proficiency in terms of CAF. Therefore, the following research questions are addressed in this study:

1) Do L2 learners find dictogloss and debating conducive to writing development in terms of CAF?

2) How do the students react to the influence of output-based speaking activities on their writing development?

Method

Participants

The study, initially, consisted of 37 English-major junior BA students selected based on convenience sampling from Islamic Azad University of Quchan, located in Khorasan Province of Iran. They participated in a supplementary course of Panel Discussion and Writing Development during the academic year. They had already passed their 8-credit courses of Conversation, Grammar, and Reading Comprehension and a 2-credit course of Advanced Writing. They were divided into two intervention groups based on convenient sampling. Moreover, the researcher administered the Oxford Quick Placement Test (OPT) to the participants for homogeneity purposes. Those volunteers who were classified in intermediate level were included in the present study. Six of the students were not classified in intermediate level (two students in class A and four students in class B) so that their writing performance on the pre-test and post-test was not included in the data analysis. Thus, the final sample, included in the experiment, consisted of 31 students (class A =15 & class B =16) (females: n=17, 54.80%; males: n=14, 45.20%; Mean age=21.48, SD=.96). Each group had two hours of English per week working on productive language skills entailing speaking and writing skills.

Instruments

The first instrument utilized to measure L2 learners' language proficiency was OPT as a language proficiency test that contains 60 multiple choice vocabulary and grammar items. The scoring criteria categorize the participants into four levels of English language proficiency: elementary (1-14), pre-intermediate (15-29), intermediate (30-44), and upper-intermediate (45-50). Those volunteers who were classified in intermediate level were included in the present study.

The second instrument was semi-structured interview questions utilized to discover the learners' perspectives of the role of output-based speaking activities in their writing development (Appendix A). The questions centered on their learning experience, familiarity, and reflection on the type of task activities presented during the course. The contents of the questions were checked by three experts in applied linguistics and three experts in assessment and testing who had been teaching English language courses at university levels. Having received the feedback from the experts, the researcher revised the questions to take care of the content validity.

To measure the syntactic complexity, the researcher opted for three indices suggested by Norris and Ortega (2009) entailing: 1) global complexity (number of words per T-unit, MLT), 2) complexity by subordination (mean number of clauses per T-unit, C/T), and 3) clausal/phrasal complexity (mean length of clauses, MLC). To measure lexical proficiency, the researcher used the TTR (Malvern et al., 2004) as well as WL and WF (McNamara, Crossley, & McCarthy, 2010). To measure accuracy, the researcher first segmented the written texts into clauses. Following Miller's (2008) guidelines, nonfinite clauses were also classified as subordinate clauses because "they express propositions and, like finite clauses, consist of a verb plus complements and

adjuncts" (p. 85). Altogether, Following el Majidi et al. (2020) guidelines, the following indices were calculated: 1) error-free clauses (EFCs), 2) lexical errors per 100 words, 3) syntactic errors per 100 words, 4) morphological errors per 100 words, and 5) prepositional errors per 100 words. Finally, to measure fluency, the number of words that a student could produce in 15 minutes was counted (Plakans, Gebril, & Bilki, 2019).

Data Collection Procedure

The current study began with a quantitative approach using experimental design to investigate the effect of debating and dictogloss on writing proficiency, supplemented by a qualitative interview method to conduct the current research. The obvious advantage of employing a mixed-methods approach is that by adopting both quantitative and qualitative research methods, the researcher could benefit from the strength of both (Riazi, 2017). To be more exact, the study followed a quantitatively dominant sequential mixed-methods research including a qualitative phase (Johnson & Christensen, 2012) to improve the experiment further by conducting a semi-structured interview 'that can greatly enhance the study's internal validity" (Dörnyei, 2007, p. 173). As such, the study was carried out in two major steps as follows:

During the first step of the study, before the treatment phase, the students were homogenized in terms of overall language proficiency by means of OPT, and following this, they were assigned into two experimental groups. Prior to the treatment phase, the researcher administered the pre-test which was a conversational free-opinion topic "The merits and demerits of living in a small city" which was of interest to students since they all shared the same experience. The students were allotted 15 minutes to write a single paragraph

on the issue. The treatment phase lasted for eleven sessions beginning from January 2019 to May 2019 with classes holding once a week for 90 minutes for each group during which the participating teacher who was the author allocated approximately the first 45 minutes to free-opinion speaking tasks and the second 45 minutes of his time in each class to writing tasks. The themes of the classes included topics generally about psychology, culture, media, education, art, and sport. Considering the experiment, class A, consisting of 15 students, was taught through dictogloss instruction, and class B, consisting of 16 students, was taught through debating instruction. Both instructions dealt with the speaking and writing activities, meeting the requirements of the course. The classroom physical environment was spacious and convenient, with students seated in circle seating arrangements.

The teacher supplied two tasks, including a simple and clear instruction for each group. The first task required the learners to participate in a speaking activity in which they discussed the topic. In the second task, learners were engaged in a writing activity which was discourse and text-oriented. They were required to fulfill this task by themselves. After they accomplished the tasks, the teacher went over the answers and offered corrective feedback.

Following the procedure suggested by Wajnryb (1990), the group exposed to dictogloss instruction received the collaborative output tasks in four stages, including preparation, dictation, reconstruction, analysis, and correction. During the first stage, the teacher started the class with a warm-up, reviewing the necessary vocabulary and collocations to cope with the task. Moreover, they were also divided into groups of four individuals or maybe three individuals because of the odd number of students in a session. During the second stage, they listened to a related text read by the teacher at normal speed, and they were asked not to take any notes. However, they were required

to note down the keywords while listening to the text for the second time so that they could reconstruct the text. During the third stage, they worked in small groups, and the teacher encouraged them to reconstruct the text. Finally, in the last stage, they analyzed, compared, and corrected their texts. The teacher was also around to provide them with feedback if needed.

The group exposed to debate instruction was required to discuss the topic that was already chosen for the session which was of interest to the students (e.g., fashion, Instagram), and they were given a chance to be for or against in the debate mostly in groups of four individuals. Following the procedure developed by el Majidi et al. (2018), each debate consisted of three levels. The students were informed about the triple levels during the first session. Initially, the students were asked to interact with information, arguments, and texts on the learner-content level. Then, on the learner-instructor level, the teacher stimulated the students to debate on the topic and he offered feedback on their performance. Finally, on the learner-learner level, the students debated the topic in their groups. Following the treatment phase, the researcher administered the post-test of writing, including a conversational free-opinion topic "The advantages and disadvantages of pursuing academic studies," which was of interest to the students since they all shared the same concern. The students were allotted 15 minutes to write a single paragraph on the issue. The students' writings on the pre-test and post-test were typed and filed separately for further analysis.

During the second step of the study, the researcher gathered data from six students (three students from each group) who were exposed to output-based speaking activities, employing a semi-structured interview method. The data were obtained over a series of three weeks in August 2019. Each of the interviews was a bit different in length so as for the interviewer to make sure

that the interviewees' responses provide no more new information. The learners were free to answer the questions in English or Persian; however, their responses were transcribed into English and then analyzed by classifying, that is, by tracing commonalities across them.

Data Analysis

Descriptive and inferential statistics were used to analyze the data, as for the study's first objective regarding the effect of dictogloss and debating on writing development in terms of CAF. Syntactic and lexical complexity was measured by means of an automatic L2 syntactic complexity analyzer (Lu, 2010). In this regard, to determine the word length and weighted percentages of the data, the researcher also made use of the Nvivo Software Version 10, queries option. Accuracy was measured manually, through which the researcher identified mistakes for each text and categorized them into four categories entailing lexical, syntactic, morphological, and propositional errors. In addition, the number of EFC was calculated manually. Fluency was measured with reference to the number of words that each student could produce in 15 minutes. Moreover, to assess inter-rater reliability for the hand-coded measures with respect to accuracy indices, the researcher, using SPSS software version 22, ran the Kappa value, and the results confirmed that the inter-rater agreement for accuracy indices were acceptable since the values were .81, .84, .76, .81, and .72 for EFCs, lexical errors, syntactic errors, morphological errors, and propositional errors, respectively, following the guidelines suggested by Cohen (1960). Thereafter, a paired-samples t-test was run to determine if post-test gain was significant after the treatment for each group. Then, ANCOVA was run to compare the significant difference between the two groups while controlling the pre-test scores as the covariance.

Following this, as for the second objective of the study regarding the students' reactions to the influence of output-based speaking activities on their writing development, the researcher made use of "theme-based categorization" (Dörnyei, 2007, p. 245) to label the responses emerging from the open-ended interview questions. The inter-rater agreement and inter-rater reliability for coded transcripts were also taken care of. It is worth mentioning that the inter-coder agreement needs the two coders to reach an agreement through discussion (Garrison et al., 2006), and the inter-coder reliability requires that the two coders choose the same code for the same unit of text (Krippendorff, 2004).

Results

Dictogloss vs. Debating and Writing Proficiency

The first objective of this study was to examine the impact of dictogloss and debating on the written performance of intermediate EFL learners. To this end, the researcher compared the sets of scores obtained from the pre-test and post-test for each learner in both conditions (including the two experimental groups) and on each measure (see Appendix B for an example of the analysis of writing performance of a participant).

Table 1.

Means and Standard Deviations of Writing Proficiency in terms of CAF

	dictogloss	debate			
measures	**index**	**pre-test**	**post-test**	**pre-test**	**post-test**
Syntactic	MLT	11.12 (.65)	11.32 (.74)	10.60 (.68)	10.67 (.68)
complexity	MLC	8.18 (.43)	8.35 (.39)	8.41 (.44)	8.72 (.41)
	C/T	1.13 (.18)	1.05 (.17)	1.26 (.33)	1.21 (.32)

Lexical complexity	TTR	.32 (.04)	.35 (.05)	.36 (.05)	.44 (.06)
	Word frequency	1.27 (.33)	1.20 (.24)	1.48 (.40)	1.38 (.33)
	Word length	4.98 (.23)	5.64 (.33)	5.31 (.44)	5.38 (.51)
Accuracy	EFCs	.34 (.05)	.38 (.06)	.46 (.06)	.52 (.10)
	Lexical errors	1.63 (.39)	1.46 (.37)	1.41 (.30)	1.14 (.29)
	syntactic errors	1.66 (.59)	1.32 (.52)	1.78 (.47)	1.67 (.44)
	morphological errors	2.45 (.37)	2.38 (.37)	2.32 (.39)	2.28 (.41)
	propositional errors	1.10 (.16)	1.04 (.18)	1.22 (.31)	1.16 (.29)
Fluency	number of words	96.06(6.78)	100.33(5.5)	98.81(5.31)	104.62(3.83)

Note. EFCs = error free clauses; MLT= number of words per T unit; MLC=mean length of clauses; C/T=mean number of clauses per T unit; TTP=type-token ration

The results obtained from the descriptive statistics, including the means and standard deviation, showed that the students improved over the intervention period in a number of CAF measures (see Table 1). It seems that the students in both groups showed improvement following the instructions in terms of aspects of writing proficiency, and the debate group outperformed the dictogloss group across the majority of indices of CAF at the post-test.

Table 2.

Results of Paired-samples t-test for the Dictogloss Group

measures	**index**	**df**	**t**	**sig.**
Syntactic complexity	MLT	14	-1.32	.20
	MLC	14	-2.51	.02
	C/T	14	1.56	.14
Lexical complexity	TTR	14	-2.16	.04
	Word frequency	14	1.23	.23
	Word length	14	-10.67	.00
Accuracy	EFCs	14	-2.20	.04
	Lexical errors	14	4.75	.00
	syntactic errors	14	7.32	.00
	morphological errors	14	1.98	.06
	propositional errors	14	1.47	.16
Fluency	number of words	14	-4.70	.00

As shown in Table 2, the results of the paired-samples t-test confirmed that the students in the dictogloss group showed significant improvement over the instruction. To be more specific, as for syntactic complexity, the participants showed significant improvement in terms of clause length, $t_{(29)}$=-2.51, p=.02. However, the other two measures, including MLT and C/T fell short of significance. Concerning lexical complexity, it was found that the TTR was significantly different, and the students could use more different words in their post-test writings. Moreover, the results revealed that the mean word length, $t_{(29)}$=-10.67, p=.00, was significantly different too. As for the word frequency, although the students improved from pre-test to post-test (see Table 1), this improvement fell short of significance. Taking accuracy into account, the researcher found that the students displayed improvement across three accuracy measures in terms of EFCs, $t_{(29)}$=-2.20, p=.04, lexical errors, $t_{(29)}$=4.75, p=.00, and syntactic errors, $t_{(29)}$=7.32, p=.00. However, no significant difference was found in morphological errors and propositional errors. With regard to fluency, the results indicated that the students could produce longer texts in their post-test, $t_{(29)}$=-4.70, p=.00.

Table 3.

Results of Paired-samples t-test for the Debate Group

measures	index	df	t	p
Syntactic complexity	MLT	15	-1.47	.16
	MLC	15	-4.19	.00
	C/T	15	1.42	.17
Lexical complexity	TTR	15	-8.50	.00
	Word frequency	15	1.73	.10
	Word length	15	-1.20	.24
Accuracy	EFCs	15	-2.55	.02
	Lexical errors	15	6.31	.00
	syntactic errors	15	2.37	.03

measures	index	df	t	p
	morphological errors	15	1.78	.09
	propositional errors	15	1.61	.12
Fluency	number of words	15	-4.40	.00

As indicated in Table 3, the results of the paired-samples t-test confirmed that the students in the debate group displayed significant improvement over the instruction. More specifically, concerning syntactic complexity, they showed significant improvement in terms of clause length, $t_{(29)}$=-4.19, p=.00. Nevertheless, the other two measures of MLT and C/T did not show a significant difference. As for lexical complexity, the researcher found that the TTR was significantly different and the degree of vocabulary variation in the students' writings increased over the instruction. Furthermore, the results revealed that the students improved with respect to word frequency and word length from pre-test to post-test (see Table 1); nevertheless, their progress fell short of significance. As for accuracy, it was found that the students displayed improvement across three accuracy measures in terms of EFCs, $t_{(29)}$=-2.55, p=.02, lexical errors, $t_{(29)}$=6.31, p=.00, and syntactic errors, $t_{(29)}$=2.37, p=.00. However, the results showed that there was no significant difference with respect to morphological errors and propositional errors. Regarding fluency, the researcher found that the students were able to produce longer texts in their post-test, $t_{(29)}$=-4.40, p=.00.

Table 4.

Results of ANCOVA for Comparison between the two groups

measures	index	df	F	Sig	partial eta sq.
Syntactic complexity	MLT	1	1.72	.20	.05
	MLC	1	4.74	.03	.14

measures	index	df	F	Sig	partial eta sq.
	C/T	1	.70	.40	.02
Lexical complexity	TTR	1	15.28	.00	.35
	Word frequency	1	.33	.56	.01
	Word length	1	39.10	.00	.58
Accuracy	EFCs	1	6.83	.02	.34
	Lexical errors	1	4.64	.04	.14
	syntactic errors	1	13.93	.00	.33
	morphological errors	1	.28	.59	.01
	propositional errors	1	.33	.57	.01
Fluency	number of words	1	4.59	.04	.14

As shown in Table 4, the results of ANCOVA showed a significant difference between the two groups in terms of MLC, TTR, word length, EFCs, lexical errors, syntactic errors, and fluency. To elucidate, with regard to the means in Table 1, the debate group improved more than the dictogloss group in terms of MLC, $F_{(1, 28)}$=4.74, p=.03, partial eta squared=.14, TTR, $F_{(1, 28)}$=15.28, p=.00, partial eta squared=.35, EFCs, $F_{(1, 28)}$=6.83, p=.02, partial eta squared=.34, lexical errors, $F_{(1, 28)}$=4.64, p=.04, partial eta squared=.14, and fluency, $F_{(1, 28)}$=4.59, p=.04, partial eta squared=.14. The results of eta squared showed a large effect size for all five measures (Cohen, 1988). However, the dictogloss group improved more than the debate group in terms of word length, $F_{(1, 28)}$=39.10, p=.00, partial eta squared=.58, and syntactic errors, $F_{(1, 28)}$=13.93, p=.00, partial eta squared=.33. The results of eta squared indicated a large effect size for the two measures. No significant difference was found in terms of MLT, C/T, word frequency, morphological errors, and propositional errors.

Results of Interviews with the Students

As for the study's second objective regarding the students' reactions to the role of collaborative output-based instructions in writing development, the researcher constructed semi-structured open-ended questions, employing theme-based categorization, to analyze data (Dörnyei, 2007). Some of the statements made by the students, along with the researcher's brief explanations, are reported below:

As for the learners in the dictogloss group, they mentioned that learning new vocabulary and collocations through dictogloss could make the task of acquiring unknown words easy and the amount of pressure on their minds becomes low because the teacher provides them with new unknown vocabulary and their collocations on the board before delving into the task. The first participant said:

> [I was not already familiar with this type of instruction through which the teacher-oriented my attention to the new vocabulary. The problem was that when we compared our writings, especially during the first sessions, my peers did not take it seriously to analyze the language, but step by step, they took part in the discussion before starting the writing tasks].

The researcher found out that the four language skills could be integrated through dictogloss so that learners were being engaged in accomplishing the task, and working in small groups was more interesting and less stressful for the students. The second participant mentioned:

> [I had always considered the tasks as competitive tasks, and it was all good experience that I was taking part in a cooperative task, enabling me to write longer sentences].

It seems that working on the language through receptive skills and then practicing the language through productive skills were altogether motivating for the students. The students also meant that the learning culture was shifted to an anti-reductionist and holistic language paradigm since they could integrate the language skills. Through comparing and analyzing the texts in small groups, the students found an opportunity to focus on syntactic errors. However, they believed that while doing the writing tasks, they had mounting concerns about accuracy and correct constructions of grammar. Of course, one of the students said that because they were exposed to adequate input and output activities before doing the writing task, they were willing to write creatively and fluently. The third participant pointed out:

> [I enjoyed writing after discussing the topic to reconstruct it. I saw myself involved in doing the task without feeling anxious. The only major problem that I encountered was that I was not sure about the correct proposition of the words].

The controversial point was that the students had no idea of the type/token ratio in their writings, and the number of repeated words was rather high, perhaps due to the fact that they were not acquainted with such distinctions. Moreover, working on the three aspects of language proficiency simultaneously needed much more time to draw their attention to all aspects of writing proficiency.

The students in the debate group stated that working on meaning through the three stages of debating could improve their vocabulary repertoire since they could use the words more effectively and express their voices freely. The first participant said:

[I was really satisfied with this type of activity since I had no stress to speak in English with peers, but I focused more on fluency rather than accuracy. I also found that I was using different clauses more than before in my writings].

Actually, learners in this group stated that working on debate-based activities was quite challenging, and speaking in English in small groups reduced their stress and anxiety. Moreover, they learned to create argumentative writings and used different words that, in turn, resulted in longer sentences. It seems that debating helped them develop their ideas and did not concern much about writing mechanics. The second participant mentioned:

[What was more important for me was to write a meaningful and creative text, and while writing, I was more engaged with the idea development, and I tried to write complex sentences and avoid using repeated words. While writing, I tended to write critically about the positive and negative points related to the topic. I was able to write more with less interruption].

The statements by the interviewees revealed that debate-based instruction could act as scaffolding, providing materials for them to write down critically since they felt that they could go on to argue successfully about the topic with the readers. The researcher also found out that providing the learners with positive energy and emotional feedback would galvanize their minds and they would try their best to create excellent writings. The third participant noted:

[During the class time, the teacher asked us to note down the arguments and the mistakes from classmates, and these techniques helped me a lot].

Actually, the writing tasks were no more tedious for the students who got into them enthusiastically, and writing did not mirror dictation to them. Their writings no longer looked like incoherent text. They could use the words in new sentences since their intervention phase was not based on repetition and memorization but based on note-taking, rephrasing, and creativity. Through debate-based activities, they demonstrated the ability to develop more complex sentences with fewer lexical and syntactic errors since they have already gained the necessary skills to write, focusing on CAF during the treatment.

Finally, taking the inter-coder reliability and agreement into account, having coded the data, the researcher provided a second person who was one of his colleagues, holding a Ph.D. in TEFL and teaching at the same university as a faculty member with the data to code. After that, the second researcher coded the responses by eliciting the commonalities and formulated rather similar findings with minor differences. In doing so, initially, following the guidelines suggested by Campell et al. (2013), the researcher divided the number of coding agreements by the number of agreements and disagreements combined, and they achieved acceptable inter-rater reliability. To be more specific, as of the dictogloss group, there were 21 common themes that at least one of the researchers invoked a code, and of these, there were 14 cases that both of the two coders had invoked the code. Regarding the debate group, there were 25 common themes that at least one of the researchers invoked a code, and of these, there were 19 cases that both of the two coders had invoked the code. Therefore, the inter-coder reliability for the dictogloss group was 66 percent (14/21=.66), and for the debate group was 76 percent (19/25=.76). However, after negotiating discrepancies, they reached 78 percent inter-coder reliability for dictogloss (11/14=.78) and 84 percent inter-coder reliability for

debating (16/19=.84). Therefore, following the coding reliability and agreement, the number of common themes that emerged from the students' responses was reduced to 11 codes for dictogloss and 16 codes for debating.

The coding results for the dictogloss group included: 1) facilitative, 2) focused attention, 3) productive, 4) low stress, 5) fewer errors, 6) integration of language skills, 7) collaboration, 8) higher self-efficacy, 9) motivating, 10) paraphrasing and 11) accurate complex sentences. The coding results for the debate group included: 1) productive, 2) active involvement, 3) fluent writing, 4) challenging, 5) lexical repertoire, 6) free opinion, 7) enjoyment, 8) accurate fluency, 9) idea development, 10) problem-solving, 11) arguing, 12) emotional feedback, 13) using longer complex sentences, 14) writing without distress, 15) content knowledge, and 16) motivating.

Discussion

The primary goal of this study was to evaluate the effectiveness of dictogloss and debating interventions on L2 learners' writing proficiency. The results obtained from the study confirmed that both dictogloss and debating were conducive to writing development in terms of CAF, showing an increase in different indices of syntactic and lexical complexity and accuracy as well as in fluency following the treatment. The findings also showed that the debate group improved more than the dictogloss group in terms of MLC, TTR, EFCs, lexical errors, and fluency, while the dictogloss group improved more than the debate group in terms of word length and syntactic errors. Finally, the results of the interviews with the students regarding the role of task-based collaborative output activities on their written performance yielded several commonalities, indicating that the use of task-based collaborative output

activities as teaching tools is actually fruitful in enhancing L2 writing proficiency.

Taking the first objective of the study into account, the obtained results revealed that EFL learners experienced collaborative output activities conducive to their writing development in terms of CAF, which is in line with the study carried out by el Majidi et al. (2020), who investigated the effect of debate-based instruction, as an L2 collaborative teaching tool, on writing development in the context of secondary school and they concluded that debate-based instruction increased aspects of writing proficiency in terms of CAF. Similarly, the results of the study are consistent with those of Barrot (2018) who found that features of sociocognitive-transformative approach such as collaboration and metatalk significantly showed an increase in fluency and complexity in L2 learners' written performance.

The findings of the study displayed improvement in written performance in nearly all aspects of writing proficiency in terms of CAF which are both in agreement and disagreement with those of Wigglesworth and Storch (2009) who concluded that pair writing positively affect learners' CAF and collaborative writing showed an increase in the accuracy of essays but not in fluency and complexity. In the same vein, Hartshorn et al. (2010) worked on the effect of dynamic written corrective feedback on CAF and they found that although writing fluency and complexity were largely unaffected by this instructional pedagogy, significant improvement was observed for writing accuracy. The analyses of the students' writings revealed that the number of complex sentences via coordination and subordination increased in their writing production. Likewise, Norris and Ortega (2009) reported the frequent use of subordination as indicative of writing development.

Indeed, the significant increases in nearly all measures of complexity over the intervention indicate that the treatment for both groups was fruitful in assisting students to better produce language which is not consistent with the claim of Skehan (1996) that accuracy decreases as students write more complex clauses and vice versa. It seems that collaborative activities can trigger students' writing competence to elevate their levels of attention, involvement, and self-efficacy in doing productive tasks. Moreover, working on collaborative activities enhances critical thinking and enables students to express their opinions, which has already been concluded by Green and Klug (1990), who found that debate format is an effective way to modify learners' opinions and to teach critical thinking and writing skills. Similarly, Farid et al. (2017) found that the collaboration aspect of dictogloss enhances the students' writing ability. In a similar vein, the debate-based instruction seems to be related to 'L2 willingness to communicate' (McIntyre et al., 1998) because debating can increase students' readiness to enter the discourse, and since sensory emotioncy has a significantly positive relationship with students' willingness to speak and willingness to write (Makiabadi et al., 2019), 'emotioncy-based language instruction' (Jajami & Pishghadam, 2019) could act as scaffolding for students to become actively involved in collaborative output-based activities and enhance their L2 writing proficiency. The results of the study revealed that students in the dictogloss group displayed significant improvement in syntactic errors in comparison to the debate group, which is consistent with the relevant study by Schmitt (2012), who found that disctogloss, as a collaborative writing task, encourages learners to reflect on form in their written performance.

More specifically, as for lexical complexity, it seems that the collaborative output activities enable students to use lexical patterns in their

writings such as (*gain knowledge, monotonous life…)* which is consistent with the previous research into the role that task-based collaborative output activities play in formulaic expressions and writing development (Ellis & Yuan, 2004; el Majidi, 2020). In addition, Hyland (2019) found that reading-to-write pedagogy improves lexical complexity in written performance. Furthermore, both groups gained improvement in TTR measure of lexical complexity as an indication of the use of different words in a text which is consistent with the previous study undertaken by Crossley, Salsbury and McNamara (2014) who concluded that the use of less frequent words is related to lexical proficiency in writing. As of syntactic complexity, the obtained findings revealed that MLC, as an indication of complexity at the clausal level, reached significance in both groups over instruction, and also the debate group outperformed the dictogloss group in terms of MLC and previous research documented this index as an important measure to determine syntactic complexity development (e.g., Norris & Ortega, 2009; Lu, 2010). The results of the study also indicated a decrease in students' written performance in terms of C/T, or subordination, from pre-test to post-test in both the dictogloss group (pre-test: M=1.13; post-test: M=1.05) and the debate group (pre-test: M=1.26; post-test: M=1.2) and this decrease is associated with an increase in L2 writing proficiency (Norris & Ortega, 2009; el Majidi, 2020).

To be more specific, concerning accuracy measures, the researcher found that both groups significantly improved from their pre-test to post-test in four out of five indices of accuracy entailing EFCs, lexical errors, syntactic errors, and morphological errors, and the debate group represented more improvements than the dictogloss group in two indices of accuracy entailing EFCs and lexical errors which is consistent with the study carried out by el Majidi et al. (2020) who found that the debate group reached statistical

significance in two indices of EFCs and lexical errors over the intervention. However, in this study, the dictogloss group improved more than the debate group in one accuracy index, including syntactic error. It seems that debate-based instruction helps students to empower their vocabulary repertoire more than the dictogloss because of the nature of the discussion that needs more words to negotiate an argument; however, dictogloss instruction helps students to enhance their syntactic knowledge more than the debate-based instruction, perhaps due to the fact that dictogloss provides learners with the opportunity to reconstruct a text together and their joint product after comparing and analyzing their texts will be better than individual reconstruction, and this set of finding is aligned with the relevant study by Kuiken and Veder (2002) who concluded that dictogloss increases syntactic quality and deepens learners' awareness of linguistic rules.

Finally, as the students mentioned in their interviews, they could use the words they have already learned in the intervention which is aligned with the findings obtained by Jianling (2018) who observed transfer patterns in similar tasks activities which has been referred to as 'generic relationship between the task' (Hyland, 2007, p. 149). The commonalities emerged from the students' responses showed that collaborative tasks are motivating, enjoyable, and helpful for the students to reduce their stress. Previous research also documented that a collaborative environment motivates students and improves their writing fluency (Albadi, 2018). Indeed, in line with the social constructivist perspective (Vygotsky, 1987), collaborative activities can act as scaffolding enabling students to develop their potentiality in writing skills. In this respect, Swain and Lapkin (2001) found that collaborative tasks could enable students to tackle linguistic problems that are beyond their individual abilities.

Conclusion

The findings of this study make a strong case for addressing the issue of task-based collaborative output activities and embracing the future perspective of EFL learners' writing proficiency in terms of CAF. Actually, the attention devoted to the students' development of language production in their L2 settings is of great value since students extend their writing proficiency while engaging in discourse-based and socially-created tasks (Modarresi & Alavi, 2014). Following chaos/complexity science (Larsen Freeman, 1997), CAF possesses a dynamic, complex, and emergent nature, what has been referred to as "an organic nature of CAF" (Norris & Ortega, 2009, p. 556). Consequently, many interacting factors are at play to elevate CAF indices among which are the types of output and interaction. Although the precise measure of CAF indices is a demanding task, awareness of these factors helps diagnose the students' weaknesses and strengths in terms of CAF. This research has yielded a deeper insight into the effective role that the integration of language skills can perform in the development of writing proficiency, which is considered a required skill for attaining academic success and activating higher-order thinking processes that would enable students to think creatively analytically. Indeed, the present study reinforces the conclusion that collaborative output tasks are highly challenging and arguing, requiring higher mental loads. The active participation and focused attention nested in the collaborative environments can foster learners' awareness of the different aspects of language proficiency.

There are potentially helpful implications for L2 learners and teachers in relation to the provision of collaborative activities for writing courses. EFL students are expected to engage in challenging output tasks and socially oriented activities to improve their writing proficiency. They should learn to

interact with their classmates while integrating language skills and employ the output activities that work best for them to produce creative, accurate, and fluent written performance in L2. Learners are recommended to reflect on their writing performance in terms of CAF, trying to write longer sentences with less frequent works and self-assess their writing development during the term. L2 teachers are suggested to provide the students with output activities that are derived from novel findings in the field of SLA. Teachers should present samples of students' writing works and measure them in terms of CAF indices for students on the board to create a mental image of actual writing proficiency.

Although this study offers some fascinating insights, it has a number of limitations too. First, care should be taken in terms of the external generalizability of the findings since the sample is not representative of all English-major students at the university level. Second, due to the limited number of students available, the study included no control group. Actually, whereas the use of a control group is generally recommended, in some circumstances, the inclusion of a control group might not be possible for practical reasons (Mackey & Gass, 2016; Plonsky, 2017). Third, this experiment was conducted for 11 sessions in a nearly four-month period, and as commented by Storch (2009), development in L2 skills cannot be taken place over a 2- to 4-month period, so that further longitudinal studies with longer duration are needed to explore the extent to which task-based collaborative output activities can enhance writing proficiency in terms of CAF. Finally, the present study opted to design similar tasks with the same difficulty level; however, further research can examine the relationship between task complexity and writing complexity.

Finally, we are in the rather initial steps of experimentally investigating the role of collaborative output tasks in aspects of writing proficiency in the Iranian context, and now the door is open for carrying out further research concerning the impact of output-based instruction on writing proficiency in order to create a complete picture of CAF indices in relation to learning foreign languages.

Declaration of Conflicting Interests

The author declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Albadi, A. A. M. (2016). Twitter mobile application: A source of communicative and authentic learning. *Journal of Teaching English for Specific and Academic Purposes*, *4*, 49-60.

Alexopoulou, T., Michel, M., Murakami, A., & Meurersc, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning, 67*, 180-208.

Allen, H. W. (2018). Redefining writing in the foreign language curriculum: Toward a design approach. *Foreign Language Annals, 51*, 513-532.

Barrot, J. S. (2018). Using the sociocognitive-transformative approach in writing classrooms: Effects on L2 learners' writing performance. *Reading & Writing Quarterly, 34*(2), 187-201.

Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching, 7* (3), 377-396.

Campbell, J. L., Quincy, Ch., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semi-structured interviews: Problems of unitization and inter-coder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294-320.

Cho, H. (2017). Synchronous web-based collaborative writing: Factors mediating interaction among second-language writers. *Journal of Second Language Writing*, *36*, 37-51.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20* (1), 37–46.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings. *Applied Linguistics*, *36*, 570-590.

Crowhurst, M. (1991). Research review: Patterns of development in writing persuasive/ argumentative discourse. *Research in the Teaching of English*, *25*(3), 314-338.

Dehghan, M., & Mohammad Amiri, M. (2017). Collaborative output tasks and their effects on learning English comparative adjectives. *Journal of Language Teaching Skills*, *36*(1), 1-26.

Dobao, A. F. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40-58.

Doody, O., & Condon, M. (2012). Increasing student involvement and learning through using debate as an assessment. *Nurse Education in Practice*, *12* (4), 232-237.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Dörnyei, Z. (2019). What makes an L2 task engaging? In W. Z. Ahmadian MJ (Ed.), *Researching L2 task performance and pedagogy: In honor of Peter Skehan* (pp. 53-66). Amsterdam: John Benjamins.

el Majidi, A., de Graaff, R., & Janssen, D. (2018). Students' perceived effect of in-class debates in second language learning. *The European Journal of Applied Linguistics and TEFL*, *7*(1), 35-57.

el Majidi, A. E., de Graaff, R. , & Janssen, D. (2020). Debate as L2 pedagogy: The effects of debating on writing development in secondary education. *The Modern Language Journal*, *104* (4), 804-821.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59-84.

Farid, A. A. P. A., Setyarini, S., & Moecharam, N. Y. (2017). The implementation of dictogloss storytelling in improving 8th-grade students' writing skill. *Journal of English and Education*, *5*(1), 85-91.

Ferris, D. R. & J. Hedgcock (2004). *Teaching ESL composition: Purpose, process, and practice*. New York, NY: Routledge.

Fortune, A. (2005). Learners' use of meta-language in collaborative form-focused L2 output tasks. *Language Awareness, 14,* 21-38.

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*(1), 45-57.

Gallego, M. (2014). Second language learners' reflections on the effectiveness of dictogloss: A multi-sectional, multilevel analysis. *Studies in Second Language Learning and Teaching, 4* (1), 33-50.

Ganji, M., & Ketabi, S. (2015). Cloze and editing tasks: Are they effective in teaching lexical collocations? *Journal of Applied Linguistics and Language Research, 2*(5), 48-61.

García Mayo, M. P. (2002). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics, 12,* 156-175.

Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *Internet and Higher Education, 9,* 1-8.

Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple-text source-based writing. *System, 99,* https://doi.org/10.1016/j.system.2021.102524.

Green, C. S., & Klug, H. G. (1990). Teaching critical thinking and writing through debates: An experimental evaluation. *Teaching Sociology, 18,* 462-471.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010).

Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44,* 84-109.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*, 148-164.

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge: Cambridge University Press.

Jajarmi, H., & Pishghadam, R. (2019). Emotioncy-based language instruction: A key to enhancing EFL learners' vocabulary retention. *Applied Research on English Language, 8*(2), 207-226.

Jianling, L. (2018). The impact of face-to-face oral discussion and online text-chat on L2 Chinese writing. *Journal of Second Language Writing*, *41*, 27-40.

Johnson, M. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing, 37*, 13-38.

Johnson, R. B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Los Angeles: Sage.

Kanakri, A. (2016). Second language writing instruction: Teaching writing to English language learners. *Beyond Words, 4* (2), 88-101.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Kroll, B. (2003). *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press.

Kuiken, F. & Vedder, I. (2002). The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research, 37*(3-4), 343-358.

Larsen-Freeman, D. (1997). Chaos complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.

Lee, L. (2001). Online interaction: Negotiation of meaning and Strategies used among learners of Spanish. *ReCALL, 13*(2), 232-244.

Leki, I., Cumming, A., & Silva, T. (2010). *A synthesis of research on second language writing in English*. NewYork: Routledge

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*, 474-496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62.

Lustigova, L. (2011). Speak your mind: Simplified debates as a learning tool at the university level. *Journal of Efficiency and Responsibility in Education and Science*, *4*, 18–30.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in an L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*(4), 545-562.

Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design*. New York, NY: Routledge.

Makiabadia, H., Pishghadam, R., Meidania, N., & Khajavy, G. H. (2019). Examining the role of emotioncy in willingness to communicate: A structural equation modeling approach. *Revista de Psicodidáctica, 24*(7), 120-130.

Malmqvist, A. (2005). How does group discussion in reconstruction tasks affect written language output? *Language Awareness, 14*, 128-141.

Malvern, D.D., Richards, B.J., Chipere, N., & Dura´n, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Hampshire: Palgrave Macmillan.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57-86.

Miller, J. (2008). *An introduction to English syntax* (2nd ed.). Edinburgh, UK: Edinburgh University Press.

Modarresi, Gh., & Alavi, S. M. (2014). Designing and validating a test battery of computerized dynamic assessment of grammar. *TELL, 14* (2), 1-29.

Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing, 32*, 81-87.

Nakatani, Y. (2010). Identifying strategies that facilitate EFL learners' oral communication: A classroom study using multiple data collection procedures. *The Modern Language Journal, 94*(1), 116-136.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555-578.

Persky, H., Daane, M., & Jin, Y. (2003). *The nation's report card: Writing*. Washington, DC: U.S. Department of Education.

Pishghadam, R., Adamson, B., & Shayesteh, S. (2013). Emotion-based language instruction (EBLI) as a new perspective in bilingual education. *Multilingual Education*, *3*(9), 1-16.

Pishghadam, R., Jajarmi, H., & Shayesteh, S. (2016). Conceptualizing sensory relativism in light of emotioncy: A movement beyond linguistic relativism. *International Journal of Society, Culture & Language, 4*(2), 11-21.

Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, *36*, 161–179.

Plonsky, L. (2017). Quantitative research methods. In S. Loewen & M. Sato (Eds.), *Routledge handbook of instructed second language acquisition* (pp. 505-521). New York, NY: Routledge.

Polio, C., & Park, J. H. (2016). Language development in second language writing. In R. M. Manchón, & P. K.

Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287-306). New York: Routledge.

Pourdana, N., Karimi Behbahani, S. M., & Safdari, M. (2011). The impact of task types on aspects of Iranian EFL learners' writing performance: Accuracy, fluency, and complexity. *IPEDR, 20,* 261-266.

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing, 33,* 3-17.

Riazi, A. M. (2017). *Mixed method research in language teaching and learning*. London, England: Equinox.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in A componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 41,* 1-32.

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38–62.

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36*, 1-14.

Skehan, P. (2009). Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532.

Skehan, P. (2013). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.),
Researching pedagogic tasks: Second language learning, teaching and testing (pp.167-185). Oxon, USA: Routledge Taylor and Francis Group.
Smith, K. M. (2012). Dictogloss: A multi-skill task for accuracy in writing through cooperative learning. Retrieved September 15, 2020 from: http://thtjapan.org/publications/2011/069-080_smith.pdf.
Snider, A., & Schnurer, M. (2006). *Many sides: Debate across the curriculum*. New York: International Debate Education Association.
Stewart, T. (2003). Debate for ESOL students. *TESOL Journal*, *12*, 9-15.
Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*, 103-118.
Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. In M. d. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157-177). London: Multilingual Matters.
Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, *50*, 158-164.
Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in*
second language teaching and learning (pp. 471-83). New Jersey: Lawrence Erlbaum Associates.
Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: exploring task effects. In M. Bygate, P. Skehan, & M. Swain, (Ed.), *Researching pedagogic tasks*: *Second language learning, teaching and testing* (p. 98-118). London: Longman.
Talebinezhad, M. R., & Esmaeili, E. (2012). The effects of different task types on EFL learners' acquisition of two grammatical structures (infinitives and gerunds): The case of Iranian high school students. *Theory and Practice in Language Studies*, *2*(8), 1699-1709.
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.
Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445-466.

Wolfe–Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing:Measures of fluency, accuracy, and complexity*. Honolulu, HI:University of Hawai'i at Manoa, Second Language Teaching and Curriculum Center.

Yong, F. L. (2010). Attitudes toward academic writing of foundation students at an Australian-based university in Sarawak. *European Journal of Social Sciences, 13*, 471-477.

Yoon, H. J., & Polio, C. (2016). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly, 34*, 42-57.

Zalbidea, J. (2017). 'One task fits all'? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal, 101*(2), 335-352.

Zohrabi, M., & Hassanpoor, S. (2020). The effect of open vs. closed tasks on improving Iranian EFL learners' oral performance. *Journal of Language Teaching Skills*, *38*(4), 205-238.

Appendix A: Semi-structured Interview Questions

1. Were you already familiar with collaborative output-based activities such as debating or dictogloss?
2. Do you think that speaking activities are useful in improving writing skills? How?
3. Are you familiar with the concepts of complexity, accuracy, and fluency (CAF) in writing development?
4. In your opinion, to what extent do debating or dictogloss enhance your writing proficiency?

Appendix B: A sample analysis

There are merits and demerits of living in small towns. One of the merits of living in small towns is clean air because small towns have little[lex. er.] or no traffic. House prices are low and you get a great house and yard for[pro.er.] your money. It may be easier to meet with[pro.er.] neighbors in a small town. Also, most small communities have lower crime rate[mor.er.] than large cities. One of the biggest demerits of living in a small town is lack of career opportunities. Another demerit of living in small towns is that gossip can travel very rapidly where people know each other.

Measure	Index	value
Syntactic complexity	MLT	12.75
	MLC	9.27
	C/T	1.14
Lexical complexity	TTR	0.63
	Word frequency	1.51
	Word length	5.56
Accuracy	EFCs	0.63
	Lexical errors	0.97
	syntactic errors	0
	morphological errors	0.97
	propositional errors	1.94
Fluency	Number of words	103

Note. EFCs = error free clauses; MLT= number of words per T unit;
MLC=mean length of clauses; C/T=mean number of clauses per T unit;
TTP=type-token ratio.