Developing Voice in EFL Learners’ Argumentative Writing through Dialogical Thinking: A Promising Combination

Research Article

Hoda Divsar*1
Khorshid Amirsoleimani2

Received: 2019-10-26 | Revised (4): 2020-07-02 | Accepted: 2020-07-05

Abstract

Dialogical thinking paves the way for EFL learners to express their thoughts in discussion, be able to convince the intended audience effectively, and provide reasons for the way they think, which consequently leads to the manifestation of individual voice. This study examined the effect of teaching dialogical thinking on the development of voice in the writing skill of a group of intermediate female EFL learners. To this end, twenty-two EFL learners were selected randomly and were assigned to two groups, namely experimental and control groups. Oxford Placement Test was administered to ensure the participants’ homogeneity in terms of their language proficiency level. The experimental group received

1 Assistant Professor of TEFL, Department of Teaching English as a Foreign Language (TEFL), Payame Noor University (PNU), Tehran, Iran, (Corresponding Author); Hodadivsar@gmail.com
2 MA in TEFL, Department of Teaching English as a Foreign Language (TEFL), Payame Noor University (PNU), Rasht, Iran; a-khorshid@hotmail.com

DOI: 10.22051/lghor.2020.28900.1210
treatment on the dialogical thinking based on Alexander’s (2008) dialogical teaching model in the form of analytical discussions on eight controversial topics in ten sessions. On the other hand, the control group was asked to take part in the routine written classroom activities. To measure the expression of voice, Helms-Park and Stapleton’s (2003) Voice Intensity Rating Scale (VIRS) was used. The findings revealed that dialogical pedagogy stimulated logical arguments, sound reasoning, and sensible evaluations, and consequently, led to the manifestation of the individual voice via linguistic and rhetorical devices.

**Keywords:** dialogical thinking, EFL, instruction, voice, writing.

**Introduction**

To improve how to think, students must be encouraged to express their unique horizons, values, and world views in a dialogic environment (Lee & Gray, 2019; Marchenkova, 2005; Matsuda, 2015). The augmentation of voice empowers not only individuals’ confidence, but also reinforces the manifestation of their individual perspectives and viewpoints (Olinger, 2011). Ivanic and Camps (2001) defined self-voice as the manifestation of the writer’s personal views, authorial presence, and authoritativeness. In this regard, it refers to the way an author presents his/her views in relationship with the audience and the way s/he postulates meaning into a written text form the writer’s self-voice. In simple terms, an individual voice can be perceived as the process of constantly producing, forming, modifying, and comprehending the internal/external identities that shape us as writers within the enclosure of language and discourse (Hirvela & Belcher, 2001).

Paul and Elder’s (2004) dialogical thinking expanded the scope of critical thinking by taking the social context into account leading to the social constructivism which was based on the distributed cognition. According to social constructivism, learning is constructed through critical dialogues among the individuals (Boulter, 2012; Paul & Elder, 2004). Challenging the established cultural norms, the traditional values, the conventional customaries, and the acknowledged claims that lead to the outgrowth of the opposing viewpoints are the consequences of the dialogical thinking. As Tanaka (2014) puts it, in dialogical thinking instead of deducing and inferring the accurate point of view, learners take part in the rigorous evaluation of numerous plausible points of view that are shaped according to diverse belief systems.

Therefore, teaching will not merely involve the transmission of subject knowledge, as in monologic and teacher-centered classrooms, but will develop the individuals’ capacity “to engage in the dialogues through which knowledge is constantly being constructed, deconstructed, and reconstructed” (Wegerif, 2007, p. 60). Alexander and Wolfe (2008) suggested that dialogical pedagogies initiated inroads into the established ways of classroom communication through which the learners were considered as mere supporters of the teachers’ ideas and their own voices were hardly heard. Therefore, rather than authoritative monologic answers, we should see a consent mutual understanding
of the issues raised and discussed by the individuals. This paper addresses the instruction of dialogic thinking as well as voice and the importance of their development in writing skill.

**Literature Review**

Dialogic thinking, by and large, plays a significant role in the personal and professional lives of individuals (Gemmell, 2008; Helms-Park & Stapleton, 2003). Apart from the central role of the knowledge of linguistic codes and the necessity of a rich schematic knowledge of the topic, learners have to attend to a plethora of relevant elements such as dialogic thinking, critical thinking, contextual factors, the audience, the purpose, self-identification, and the self-voice. Presently, most academic writing curriculums draw others’ attention to the significance of dialogic thinking and self-voice which are essential for academic accomplishment and lifelong learning achievement (Alagozlu, 2007; Alkhoudary, 2015). It means that instead of accepting others’ ideas blindly, individuals are encouraged to critically examine the validity of the ideas presented in the texts and evaluate the ideas of other people. In that case, they can develop their authorial presence and the autonomy of their thoughts.

Numerous researches have been conducted into the quality of dialogic thinking and dialogic discourse (Alexander, 2017; Hall, 2018; Hemati & Valadi, 2017). Alagozlu (2007) explored the ways of developing writing skills by promoting critical thinking skills and authorial voice to tackle the writing barriers and to deal with the demands of the multicultural world. Ramanathan and Atkinson (1999) stated that in order to think critically, one must have an individual voice to articulate his/her viewpoints on such issues, sharpen his/her own stance against that of others who disagree, and empower his/her personal views with whatever resources of evidence and support are feasible. This implied that individuals are supposed to examine the condition critically, persuasively back their decisions, estimate and appraise the counter-arguments, and reasonably weigh different pieces of evidence that may support their positions. As Ramanathan and Atkinson (1999) pointed out, such attitudes, however, have normally been found to be problematic for second language writers.

Voice refers to authorial identity and presence (Hirvela & Belcher, 2001; Hyland 2002; Lee & Gray, 2019) that enables the individuals to laudably articulate their ideas that may be against the socioculturally acknowledged norms and violates the relative avoidance of personal disagreement and social discord (Ramanathan & Atkinson, 1999). In writing, it records the sounds of the people on the page (Elbow, 1981) to express the intended meaning with regard to the topic, the audience and the purpose of communication. It is a personal and singular characteristic of a particular writer which represents his/her different perspectives (Dean, 2006). Sedova (2017) proposed that dialogic teaching encompassed stages of regression brought about by a disagreement among the
fundamentals of the dialogic arguments and stages of development which, on the other hand, became effective when the dismantled issues were brought into a coordinated harmony.

According to Stapleton (2002), voice can be perceived as recognized discursive features related to individualism originated in written texts in some cultures. He further added that "learners for learning to write in English should develop an individualized identity, or to impart their writing with voice, whereas suggesting that doing so is an alien concept in some L2 cultures" (p. 40). In fact, in the light of the socio-constructivist perspective, self-representation and authorial presence are crucial in perceiving written text as social communication between L2 readers and writers, through which meaning is assigned (Guinda & Hyland, 2012; Hyland, 2010). Matsuda (2001) emphasized the social aspect of voice and defined it as the collective impact of the proper employment of discursive and non-discursive elements by the individuals either deliberately or unintentionally from the yet existing social repertoires. Accordingly, Tardy (2012) considered the authorial voice as a multifaceted issue that includes individual dimensions such as uniqueness and authoritativeness, social features such as the resources of one's representation and authorial attendance, and dialogic aspects like the interoperation between the people and the social aspects, including writer-reader mutual communication. Research employing social-constructivist definition of 'voice' (Hirvela & Belcher, 2001; Matsuda, 2001; Matsuda & Tardy, 2007; Prior, 2001; Tardy & Matsuda, 2009) argued that 'voice' is not utterly confined to individualism and highlighted that the individual and the social voice are reciprocally constitutive and unavoidable with an imperative function in advanced academic literacy.

The identification of the authorial voice in the written discourse is not an easy task as it is not just a simple and mere citation of the ideas of other individuals, but a relatively complicated set of linguistic strategies (Alagozlu, 2007). In an attempt to capture the features associated with voice, Helms-Park and Stapleton (2003) developed Voice Intensity Rating Scale that encompasses the textual features that are influential in the elaboration of voice in one's writing. Based on this scale, the notion of voice consists of four components: assertiveness, self-identification, repetition of central point, and authorial presence (Helms-Park & Stapleton, 2003). The first two components of the scale are based on sentence-level linguistic features while the last two components evaluate the strength of individualized voice at the level of paragraph and beyond. Assertiveness is established through the manipulation of linguistic devices such as intensifiers and hedges and self-identification is demonstrated via the employment of the first-person pronouns (Helms-Park & Stapleton, 2003; Ivanic & Camps, 2001; Ramanathan & Atkinson, 1999). Hedges are used to reduce or soften the illocutionary force of the utterances and to moderate the claims. Intensifiers, on the other hand, are classes of words, mostly adverbs, used to adjust gradable adjectives, adverbs, verbs, or past participles. Self-identification refers to the application of the first and second person, vocative case, active voice, and explicit voice markers (Scollon et al., 1998) to signify personal opinion. Advocating the use of self-identification features, Harwood (2005) indicat-
ed some advantages of using personal pronouns in academic writing in order to facilitate the development of personal views pronounced in the arguments. Reiteration of the central point deals with how frequently and explicitly the major argument is repeated and rearticulated and the authorial presence assesses the overall manifestation of the author's voice that accounts for one's distinct inner self (Ramanathan & Atkinson, 1999). The manifestation of voice signifies the writers' willingness, confidence, and "self-trust to make claims, conviction, and gumption to support those claims" (Elbow, 1994, p. 10).

An empirical study was undertaken by Helms-Park and Stapleton (2003) with the aim of finding the relationship between the elements involved in L1 voice and the quality of academic writing in L2. The participants were required to give reasons in favor of or against Canada's immigration policy. Voice Intensity Rating Scale was used to measure the voice and Jacobs et al.'s (1981) ESL Composition Profile was employed by the overall writing quality. The results revealed no significant relationship between the quality in general and overall voice intensity or between overall quality and any of the four constituents of voice.

The intensity of authorial voice with regard to the overall quality of the argumentative writing was investigated by Zare-ee et al. (2014). As one of the components of the authorial voice, assertiveness was found to have a positive relationship with the overall quality of the academic writing quality. Having analyzed the strategies for expressing the voice, they came up with nine strategies for voice expression. At the sentence-level, the high-voice participants employed intensifiers most frequently to express assertiveness, while the low-voice participants tried to arrange other lexico-grammatical tools. At the text-level, however, both the high-voice and low-voice participants were more concerned with the effect of the topic on their voice expression.

In addition to the various writing difficulties, most EFL learners are grappling with expressing their own voice and identity in their writings. Matsuda (2001), in the context of L2 writing, observed that a problem that Japanese students faced in expressing voice in English written discourse was due to their lack of familiarity with voice-expression strategies employed in English. As Doukmak (2014) documented, most learners lack the ability to demonstrate their personal voices in writing due to the supremacy of the teacher's voice in classroom. Teacher should play the role of a facilitator in the students' quest to improve their participations, to develop their knowledge, to display their voice and identity, and finally to transfer the power from herself/himself to the students (Molinari & Mameli, 2013).

Although dialogic thinking and voice have been widely discussed in second language writing, on the pedagogical level, little attention has been given to the development of voice through dialogic thinking in writing instruction. To fill such a gap, this paper attempted to explore the effect of teaching dialogical thinking on the development of voice in intermediate EFL learners' writing.
Method

Participants

This study was carried out at Kish Language Institute, Tehran. Twenty-two intermediate female EFL learners with the age range of 15 to 22 years were randomly selected and then were assigned as the members of the experimental and control groups. To ensure the students’ homogeneity with respect to their proficiency, Oxford Placement Test was used. All of the participants were Persian native speakers.

Instruments and Materials

The following tests and tasks were used to select the participants and to collect the data:

Oxford Placement Test (OPT). Oxford Placement Test version 1.1 (2001) published by Oxford University Press and University of Cambridge Local Examinations Syndicate includes 60 items measuring the participants’ general knowledge of grammar, vocabulary, reading, and writing. This test was used to homogenize the participants in terms of their proficiency level. The Cronbach’s Alpha reliability coefficient was computed to be .91 indicating an excellent reliability estimate.

Writing Pre-Test. An IELTS writing test was employed as the pretest in the present study. In the writing task, the students wrote an argumentative discursive essay on a controversial topic provided to them to present general information, offer a solution, and justify the evidence. The topic was “why nowadays the families are not so close as they used to be in the past and how they could be brought closer together”. They were given 30 minutes to write at least 250 words on the given topic and support their arguments with further explanation. The participants’ writings were assessed based on IELTS detailed performance descriptors criteria including task response, lexical resources, coherence and cohesion, grammatical range and accuracy. The validity of scoring was ensured through inter-rater reliability. The discursive writing conventions such as the order of presenting the information, the style used, and the way the paragraphs started and finished were also taken into consideration. Irrelevant sentences, off-topic responses, and disconnected text were also penalized in scoring.

Writing Tasks. Considering the participants’ level of proficiency, eight controversial topics and consequently passages were selected. Asking the participants about which topics were among their interests and priorities, the researchers collected a pool of controversial topics from which some were randomly selected. The topics, as well as the passages, were selected with the consultation of two experienced EFL university instructors who had the experience of teaching materials preparation and curriculum development courses. The passages were selected in line with Flesch readability standard based on which the difficulty level of the texts was checked to match the students’ proficiency level. The topics were also selected meticulously to be controversial enough to raise different
opinions and viewpoints. The controversial prompt would kindle and stimulate the individuals to voice their own views on the suggested topic. The following items were the selected topics:

1. The actual school system is not good.
2. The positive and negative influences of TV programs (Some argue that the negative effects of TV programs on youths are more than the positive ones).
3. The educational values of computers (Some people say computers facilitate education but do not necessarily enrich it).
4. Success in life means money.
5. Husbands and wives should have the same educational level.
6. Friendship is the most important relationship in life.
7. Doing a job that you like is more important than earning more money.
8. Team-work or individual work? Being a team worker is more important than being a brilliantly creative person.

**Voice Intensity Rating Scale.** To measure the expression of voice in the writing samples, Helm-Park and Stapleton’s (2003) Voice Intensity Rating Scale (VIRS), the analytic rating scale for voice expression in writing, was used as an instrument. The scale is based on a careful analysis of voice features from the literature. To ensure the reliability of this instrument in Iranian EFL context, it was also piloted on 14 participants \( n_{\text{cont.}} = 7; n_{\text{Exp.}} = 7 \) having characteristics similar to those of the main participants. Cronbach Alpha was run to ensure the internal consistency of this scale and it was found to be .83. Based on this scale, the voice elements are classified in two levels and four scales. Table 1 shows the Voice Intensity Rating Scale components at both sentence and text levels.

<table>
<thead>
<tr>
<th>Sentence level Scales</th>
<th>Criteria</th>
<th>Text level Scales</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Assertiveness</td>
<td>a. hedges</td>
<td>3. Reiteration of central point</td>
<td>a. restating the central point frequently and directly</td>
</tr>
<tr>
<td>2. Self-identification</td>
<td>b. intensifiers</td>
<td>b. the explicit rearticulation of the main argument</td>
<td></td>
</tr>
<tr>
<td>3. Reiteration of central point</td>
<td>c. boosters</td>
<td>a. the overall manifestation of the author’s voice</td>
<td></td>
</tr>
<tr>
<td>4. Authorial presence and autonomy of thought</td>
<td>a. use of first person pronoun</td>
<td>b. the intangible quality of identity in writing</td>
<td></td>
</tr>
<tr>
<td>5. Authorial presence and autonomy of thought</td>
<td>b. active voice</td>
<td>c. author’s explicit views</td>
<td></td>
</tr>
<tr>
<td>6. Authorial presence and autonomy of thought</td>
<td></td>
<td>d. disclosure of personal attitudes</td>
<td></td>
</tr>
</tbody>
</table>
These two first clusters segregate linguistic features of voice mainly at the sentence level. The last two groups measure the strength of personalized voice beyond sentence. The scale includes four voice components of assertiveness that is quantified as the frequency of hedges and intensifiers in a passage, self-identification that is shown through the use of first-person pronouns and grammatical voice, reiteration of central point which is measured by the rearticulation of the central idea and authorial presence and autonomy of thought which is measured by the degree to which the author clearly expresses his or her own views.

**Data Collection and Analysis**

The study was conducted in ten sessions in which sample essays supplemented by writing activities were employed to make the participants familiar with the basic structure of essay writing through dialogical thinking (See Appendix A for the detailed lesson plan). A writing pre-test was given to both groups prior to the study to examine their use of voice in writing. As the treatment, in experimental group, dialogical thinking based on Alexander's (2008) dialogic teaching was offered and students participated in consensus tasks that generally engaged them in coming to an agreement on a certain issue. Students were expected to read an article about the issue that was supposed to be discussed during that session. The issue featured opposed viewpoints in the community, in the education, and in the society. Each week students completed in-class writing tasks to show their learning of the course materials. For these tasks, they were asked to clearly state their positions on a controversial issue and to follow paragraph and essay organization principles taught in the course. Assignments for other weeks were completed through multiple in-class drafting, out of class writing, or cooperative writing. With respect to dialogical thinking instruction, students read various articles selected by the researchers on different issues including the actual school system, the positive and negative influences of TV programs, the educational values of computers, 'success in life means money', educational level of husbands and wives, friendship, doing a job to earn more money, and 'team-work or individual work'.

Students were required to participate in the discussions and complete assignments such as analyzing each topic and listing 10 reasons to reject or accept it. They were required to discuss the opinions raised on the selected topics. The topics featured clear and powerful controversial viewpoints, which the participants examined and assessed through whole-class and group discussions. The process of perceiving and apprehending controversial standpoints started by uncovering the conventional belief system that shapes assumptions, which in turn shape opinions. In addition, the researchers delivered lectures on the instruction and facilitated the practice in preparing criteria lists and recognizing dissimilarities in paradigms. After the discussions, the participants wrote essays on comparing and contrasting the divergent viewpoints from the readings, and prioritizing fairness and academic responsibility in their own writing.
Each session, students were asked to enlist and write their viewpoints around the agreements.

As part of this study, the participants were exposed to examples of boosters, hedges, intensifiers, lexical bundles, and reiteration of the central points, and were supposed to identify the instances in the passages covered in the class. For every passage up to the final one, the feedback on the target devices on voice was given to the participants. As a part of practice, the participants were given a handout with neutral statements and they were supposed to rewrite them using boosters, intensifiers, and hedges to allow them to see how to intensify or tone down one’s ideas.

Altogether, the learners completed the three writing tasks within the experimental period as part of their class requirements. After the ninth sessions, for the final essay, they wrote about two topics. One of them was selected from among the topics discussed during the treatment and another one was selected by the course instructor.

To measure the expression of voice, Helm-Park and Stapleton’s (2003) Voice Intensity Rating Scale (VIRS) was adopted. The lesson plan was prepared for the experimental group based on Alexander’s (2008) dialogic teaching that included five principles. The principles are presented below:

- Collective: Learners and teachers together address learning tasks in a group or as a class;
- Reciprocal: Listening to each other, learners and teachers share ideas and consider other possible different views;
- Supportive: Learners freely express their thoughts, without fear of being embarrassed due to possible wrong responses, and they assist each other to arrive at common understandings;
- Cumulative: Learners and teachers expand their own and each other’s knowledge and understandings;
- Purposeful: Teachers design and boost dialogic teaching with specific educational objectives in view (Alexander, 2006).

Upon the completion of the course, the final argumentative essays were analyzed through quantifying the number of instances of lexical bundles, hedges, boosters, rhetorical devices and other voice strategies that shed light on the use of each of the devices in their essays. In the following sample excerpt, merely the use of voice strategies was underlined and the other writing components such as the accuracy of punctuation and the grammatical and lexical deviations were not pointed out.
It is important to point out that by the time this argumentative essay was written, the participants were already familiarized with the voice strategies and had employed boosters, hedges, intensifiers, and the other voice strategies in their previous written assignments. In addition to the elements of cohesion, coherence, unity and other lexical and syntactic considerations, the writing essays were analyzed in terms of the use of voice strategies.

As highlighted in the excerpt, as far as assertiveness was concerned, five hedges and thirteen boosters were employed. Regarding self-identification, five first-person pronouns and twelve cases of active voice were utilized. Concern-
ing *reiteration of central point*, at least eight cases were detected. With respect to the *authorial presence and autonomy of thought*, ten cases were spotted in the form of overall presence, intangible quality of identity, and the author’s explicit views.

To ensure the reliability of scoring, following an analytical scoring, inter-rater reliability was computed based on the IELTS performance descriptors criteria. To this end, the raters who were professional IELTS instructors were prepared for the rating task through two 30-minute training sessions in which Helm-Park and Stapleton’s (2003) Voice Intensity Rating Scale (VIRS) and the IELTS detailed performance descriptors criteria were explained to them. The raters were then provided with the copies of the VIRS and IELTS band descriptors and the randomly selected writing papers of the participants. Finally, they were asked to rate the selected writings. As explained in more detail in the following section, the inter-rater reliability was ensured.

To analyze the data, both descriptive and inferential statistics were run. To this end, means and standard deviations were computed first. In addition, Cronbach Alpha was run to ensure the OPT reliability. Furthermore, to confirm the normality of the distribution of the obtained data, One-Sample Kolmogorov-Smirnov Test was run. Moreover, to explore whether teaching dialogical thinking has any significant effects on the development of voice in intermediate EFL learners’ writing, independent samples t-test and paired t-tests were run.

**Results**

*Inter-rater Reliability Analysis for the Two Raters*

Inter-rater reliability was used to assess the consistency between the ratings provided by the two raters and the degree of the agreement between them. It was computed based on the data obtained from the pilot study participants (*n* = 14). It is to be noted that in the pilot phase, there were seven participants in each group, namely control and experimental groups. The consistency of the two raters’ judgments was also tested using intra-class correlation coefficient (ICC) analysis that yielded a comparatively high level of inter-rater reliability for the writing test scores in the two administrations in pre and post-tests. The item statistics for the scores given by the two raters are presented in table 2.

**Table 2.**

*Item Statistics for the Scores Given by the Two Raters*

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest scores rater A</td>
<td>58.0000</td>
<td>6.48074</td>
<td>7</td>
</tr>
<tr>
<td>Pretest scores rater B</td>
<td>59.1429</td>
<td>6.66905</td>
<td>7</td>
</tr>
<tr>
<td>Posttest scores rater A</td>
<td>62.5714</td>
<td>5.19157</td>
<td>7</td>
</tr>
<tr>
<td>Posttest scores rater B</td>
<td>62.4286</td>
<td>4.11733</td>
<td>7</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest scores rater A</td>
<td>63.5514</td>
<td>7.41299</td>
<td>7</td>
</tr>
<tr>
<td>Pretest scores rater B</td>
<td>63.5714</td>
<td>7.45782</td>
<td>7</td>
</tr>
<tr>
<td>Posttest scores rater A</td>
<td>86.5714</td>
<td>5.44234</td>
<td>7</td>
</tr>
<tr>
<td>Posttest scores rater B</td>
<td>86.8571</td>
<td>5.78586</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 2 displayed the item statistics for the scores assigned by the two raters, showing the mean (M) and standard deviation (SD) of the data from each rater for pre and post-tests. Overall, it appeared that rater (B) measured slightly higher writing scores than rater (A) both in pre and post-tests.

After the means and standard deviation for the scores were given by the two raters for both pre and post-tests, 'average measures' were computed individually for the writing pre and post-tests. Table 3 shows the intra-class correlation coefficients (ICC) for the scores.

Table 3. The Intra-class Correlation Coefficients for the Scores by Two Raters for Pre- and Post-test Scores

<table>
<thead>
<tr>
<th>Groups</th>
<th>Correlation</th>
<th>95% Confidence Interval</th>
<th>P Test with True Value 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Cont.</td>
<td>Single Measures</td>
<td>.906</td>
<td>.553</td>
</tr>
<tr>
<td></td>
<td>Average Measures</td>
<td>.951</td>
<td>.712</td>
</tr>
<tr>
<td>Exp.</td>
<td>Single Measures</td>
<td>.952</td>
<td>.749</td>
</tr>
<tr>
<td></td>
<td>Average Measures</td>
<td>.975</td>
<td>.856</td>
</tr>
<tr>
<td>Cont.</td>
<td>Single Measures</td>
<td>.928</td>
<td>.645</td>
</tr>
<tr>
<td>Exp.</td>
<td>Average Measures</td>
<td>.963</td>
<td>.784</td>
</tr>
<tr>
<td></td>
<td>Single Measures</td>
<td>.949</td>
<td>.734</td>
</tr>
<tr>
<td></td>
<td>Average Measures</td>
<td>.974</td>
<td>.847</td>
</tr>
</tbody>
</table>

The estimated inter-rater reliability estimates between the two ratings for the pre-test scores of the control group ($r = .951$), with 95% CI (.712, .992), and for the experimental group ($r = .975$), with 95% CI (.856, .996) were computed and both of them were quite strong. Furthermore, the estimated reliability between the two raters for the post-test scores of the control group ($r = .963$), with 95% CI (.784, .994), and for the experimental group ($r = .974$), with 95% CI (.847, .995) were also calculated which were both quite wide. Therefore, inter-rater reliability of the writing measurement for the writing pre and post-test between the two raters was ensured.

Results of the Pre-test Scores of the Voice Rate in Writing

After assigning the participants into control and experimental groups, they were given a writing test to unveil the possible pre-existing differences between the two groups with respect to the voice rate in writing before offering the treatment to the experimental group. Tables 4 displays the results of an independent samples t-test which was run to analyze the participants’ pretest writing scores.
Table 2 displayed the item statistics for the scores assigned by the two raters, showing the mean (M) and standard deviation (SD) of the data from each rater for pre and post-tests. Overall, it appeared that rater (B) measured slightly higher writing scores than rater (A) both in pre and post-tests. After the means and standard deviation for the scores were given by the two raters for both pre and post-tests, 'average measures' were computed individually for the writing pre and post-tests. Table 3 shows the intra-class correlation coefficients (ICC) for the scores.

Table 3. The Intra-class Correlation Coefficients for the Scores by Two Raters for Pre- and Post-test Scores

The estimated inter-rater reliability estimates between the two ratings for the pre-test scores of the control group \( r = .951 \), with 95% CI (.712, .992), and for the experimental group \( r = .975 \), with 95% CI (.856, .996) were computed and both of them were quite strong. Furthermore, the estimated reliability between the two raters for the post-test scores of the control group \( r = .963 \), with 95% CI (.784, .994), and for the experimental group \( r = .974 \), with 95% CI (.847, .995) were also calculated which were both quite wide. Therefore, inter-rater reliability of the writing measurement for the writing pre and post-test between the two raters was ensured.

Results of the Pre-test Scores of the Voice Rate in Writing

After assigning the participants into control and experimental groups, they were given a writing test to unveil the possible pre-existing differences between the two groups with respect to the voice rate in writing before offering the treatment to the experimental group. Tables 4 displays the results of an independent samples t-test which was run to analyze the participants’ pre-test writing scores.

Table 4. Group Statistics for the Pre-test scores of the Voice Rate in Writing

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>58.5714</td>
<td>6.41891</td>
<td>2.42612</td>
</tr>
<tr>
<td>Experimental</td>
<td>12</td>
<td>63.5714</td>
<td>7.34523</td>
<td>2.77624</td>
</tr>
</tbody>
</table>

For the participants’ writing pre-test in terms of the measured voice rate, the mean scores for both groups were computed \( M_{\text{control}} = 58.87; M_{\text{experimental}} = 63.57 \). Furthermore, the standard deviation for the control group was slightly smaller than that of the experimental group \( \text{SD}_{\text{control group}} = 6.41; \text{SD}_{\text{experimental group}} = 7.34 \). Table 5 shows the independent samples t-test for the pre-test writing scores.

Table 5. Independent Samples T-test for the Pre-test Scores of the Writing

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>Pretest scores</td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.39</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-1.35</td>
</tr>
</tbody>
</table>

The independent samples t-test showed the results of the Levene’s test for the equality of the variances for the distributions. As the Sig. value for the Levene’s test was found to be higher than the alpha level, i.e., .05, the first row, i.e., ‘equal variances assumed’ was considered to explain the results of the voice rate in learners’ writing. This indicated that the assumption of equal variances was not violated for the two tests and both groups were approximately equal in terms of variance as far as voice was considered in their writing.

Results of the Post-test Scores of the Voice Rate in Writing

The independent-samples t-test was run to compare the two groups’ voice rate in writing pretest. Normality, as the main assumption of t-test, was checked before we ran the main statistical analyses. To this end, the Skewness and Kurtosis values were computed and the trimmed means were obtained. The Skewness and Kurtosis values were all within the range of ±2, indicating the normality of the distributions. After establishing the normality assumption, the t-test
was run to answer if there were statistically significant differences between the two groups in terms of the mean scores of the voice rate. The independent-samples t-test was run to see whether there was a statistically significant difference in the mean scores of the groups under study. Table 6 shows the group statistics for the two groups on posttest of writing.

Table 6.
Group Statistics for the Two Groups on Post-test of Writing (voice rate)

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>62.5000</td>
<td>4.60072</td>
<td>1.73891</td>
</tr>
<tr>
<td>Experimental</td>
<td>12</td>
<td>86.7143</td>
<td>5.54419</td>
<td>2.09551</td>
</tr>
</tbody>
</table>

In the group statistics, as table 6 shows, the means and the standard deviations for each of the groups are given. For the posttest of writing that was administered after the treatment to measure the development of voice, the mean scores of the voice rate for the control and experimental groups were \( M_{\text{Control}} = 62.50 \) and \( M_{\text{Experimental}} = 86.71 \), respectively. Furthermore, the extent of the deviation of the scores for the control group was smaller than that of the experimental group (SD_{\text{Control Group}} = 4.60; SD_{\text{Experimental Group}} = 5.54). Table 7 shows the group statistics for the two groups on writing posttest.

Table 7.
Independent Samples Test for the Two Groups on Post-test of the Writing (voice rate)

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>Posttest scores</td>
<td>Equal variances assumed</td>
<td>.61</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
</tr>
</tbody>
</table>

The independent samples t-test presented the results of the Levene’s test for equality of variances. This tested whether the variation of scores for the groups under the study was the same. The output of this test determined the correct t-value that SPSS provided for use. Since the Sig. value for the Levene’s test was larger than the alpha level (.05), the first row of the table, referred to “Equal variances assumed,” was used. In Table 9, the significance level for Levene’s test was (.44). This was larger than the cut-off point (.05). This meant that the assumption of equal variances was not violated for the posttest scores, too.

As the value in the Sig. (2-tailed) column was lower than .05, there was a statistically significant difference in the posttest mean scores of voice rate for
each of the two groups. In this study, the Sig. (2-tailed) value was (.00). Since this value was lower than the required cut-off point (.05), a significant difference in the posttest means of the voice rate in writing for the control and experimental groups was found. The mean difference between the two groups is also shown in this table (mean difference = 24.21), along with the 95% Confidence Interval (CI) of the difference showing both the lower and upper values (see Table 7).

Calculating the Effect Size for the Independent-Samples T-Test (Posttest Scores)

Effect size statistics shows the magnitude of the existing differences between groups. Eta squared was used to compute the effect size. Eta squared shows the proportion of the variance in the dependent variable under study explained by the independent (type of instruction) variable. Eta squared value for t-test was computed using the information given in the output.

Replacing with the appropriate values from table 7: Eta squared = 79.03 / 79.03 + (7+7-2) = (.8681). The guidelines (proposed by Cohen 1988) for interpreting this value are .1 = small effect, .3 = medium effect, .5 = large effect. It was found that the effect size of .8681 shows a large effect. Expressed in percentage (eta square value was multiplied by 100), 86.81% of the variance in the dependent variable under study explained by the independent (type of instruction) variable. Eta squared value for t-test was computed using the information given in the output.

To sum up, there was a significant difference in the mean scores for the control (M control = 62.50, SD control = 4.60) and experimental group (M experimental= 86.71, SD experimental= 5.54; t (14) = 8.89, p = .00). The magnitude of the differences in the means (mean difference = 24.21, 95% CI: -30.14 to -18.28) was medium (Eta squared = .8681). Thus, dialogical thinking has a statistically significant effect on intermediate EFL learners’ development of voice in writing.

The Results of Paired Samples T-Test

To investigate the extent of the participants’ development of voice in writing within the groups, paired samples t-tests were also run. These tests showed the participants’ progress in writing pre-test and post-test of the voice as shown in Table 8.

Table 8.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair 1</td>
<td>Pretest</td>
<td>58.5714</td>
<td>10</td>
<td>6.41891</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>62.5000</td>
<td>10</td>
<td>4.60072</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair 2</td>
<td>Pretest</td>
<td>63.5714</td>
<td>12</td>
<td>7.34523</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>86.7143</td>
<td>12</td>
<td>5.54419</td>
</tr>
</tbody>
</table>
The mean score of the control group for the voice rate in writing improved from pre-test ($M = 58.57$) to post-test ($M = 62.50$). For the experimental group, the mean score noticeably improved from pre-test ($M = 63.57$) to post-test ($M = 86.71$).

In order to see if these differences between pre and posttest scores of the voice rate in writing were statistically significant, paired samples t-tests were run on the pre and posttest writing scores for the two groups. The results are presented in Table 9.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>95% Confidence Interval of the Difference</th>
<th>T</th>
<th>Df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Pretest scores - Posttest scores</td>
<td>-3.92</td>
<td>7.33</td>
<td>-10.71 - 2.85</td>
<td>-1.41</td>
<td>6</td>
<td>.206</td>
</tr>
<tr>
<td>Experimental</td>
<td>Pretest scores - Posttest scores</td>
<td>-23.14</td>
<td>10.67</td>
<td>-33.01 - 13.26</td>
<td>-5.73</td>
<td>6</td>
<td>.001</td>
</tr>
</tbody>
</table>

As it was shown in Tables 8 and 9, both groups showed development of voice in their writing. This improvement could be argued to be statistically significant for the experimental group only ($p \leq .05$). The mean difference between pre and posttests for the control group was 3.92 for the development of voice in learners’ writing. However, the mean difference between pre and posttests for the experimental group amounted to 23.14. This suggested that the experimental group outperformed the control group in writing posttest. As a result, it could be concluded that dialogical thinking had a statistically significant effect on intermediate EFL learners’ development of voice in writing. The results of paired samples t-tests revealed that both groups progressed in the writing post-test. However, this progress was statistically significant simply for the experimental group ($p \leq .05$).

**Discussion**

The results revealed that the manipulation of voice strategies assisted the participants to accomplish various objectives in their writing and to portray themselves in their essays. Voice strategies such as the reiteration and the restatement of the central point improved the quality of coherence and unity in their writings (see Figure 1). With regard to the boosters and hedges, they enabled the participants not only to state their opinions but also to adjust the level of
directness and emphasis. The results also revealed that these strategies improved the participants’ overall writing skill and helped them unveil their emotions and attitudes more clearly. Through the disclosure of personal attitude, their autonomy of thought as an individual and their personal voice were imprinted. Based on the findings, the learners started making rhetorical moves that qualified them to be more cautious when suggesting or challenging the values and beliefs, and to treat the topics with more relative authority and confidence. Hyland (2002) also accentuated that writing calls not only for the exploration of the norms not only at the sentence levels but also at the discoursal and authorial levels. The EFL learners’ awareness of voice may establish rich ground for them to better understand their teachers’ feedback and to handle unity, coherence, cohesion, sentence, and structural skills, coupled with boosters, intensifiers, hedges, and other voice components to cultivate a stronger sense of authoritativeness and authorial presence in their writing.

In light of the findings of the present study, it was found that voice expression strategies, as elements of good writing (Matsuda, 2001), needed to be detected, discussed, evaluated, modeled, and taught to the EFL learners to help them have their own authorial voice and claim the ownership of their scripts. Lack of due attention to the development of voice, the components of assertiveness, boosters, hedges, self-identification, authorial presence, and stance-taking strategies in composition classes may leave the learners on their own to guess what is taken to be good writing. The findings are supported by Escobar and Fernandez (2017) who confirmed that EFL composition courses should offer an opportunity to the learners to learn not only the basic norms but also to build a discoursal and authorial voice as EFL writers.

The findings of this study are indicative of the credibility of the manipulation of dialogic teaching and its effect on the writing pedagogy and the development of voice in EFL learners’ writing. The findings are consistent with those of Fahim and Mirzaii (2014) who acknowledged the acceptability of the employment of dialogic critical thinking tasks and their weighty influence on writing pedagogy. They continued that dialogic critical thinking tasks, inherently, engage individuals in dialog; consequently, they would have the potentiality for developing voice in them considerably.

Along with the findings of the present study, Tanaka (2014) stated that dialogical thinking guides students to consider and understand the controversies through the analysis and appraisal of opposing belief systems behind controversial perspectives. The attainment of voice through dialogical thinking helps the learners to posit their ideas, views, and attitudes based on logical argumentations (Alagozlu, 2007; Hyland, 2010; Ibrahim & Nambiar, 2012). This stance suggests that critical thinking and self-voice mutually improve the quality of the L2 writing. As Barnawi (2011) confirmed, dialogic thinking and voice are “indispensable ingredients in EFL college writing/composition instruction, which will allow students to express themselves clearly, and put their own viewpoints into their writing” (p.191).
The findings suggested that dialogical thinking enables the learners to engage in argumentative discussions to meet the mutual understanding and thereby shape and strengthen their voice. As Matusov (2007) argued, dialogism can lead to the creation of voices by providing opportunities for the learners to engage collaboratively with each other and provide reasons for the way they think. The results dovetail with Chappell (2014, p. 98) who pointed out that engaging students in dialogues and writing activities help them manage reasonably what they think. Yaqubi and Rashidi (2019) also acknowledged that dialogic classes provide opportunities for all students to have voice and contribution and propose their own comments, viewpoints, and perspectives.

Merely getting a voice into one's writing is not enough (Arend & Sunnen, 2016; Escobar & Fernández, 2017) and it should be accompanied by sound and sense reasoning and thinking. Dialogical thinking is needed to encourage and stimulate reasonable arguments, sound judgments, and sensible evaluations. Regarding dialogic thinking, the results were in line with those of Frijters et al. (2008) who contended that dialogic teaching results in an extra positive influence on the critical thinking skills of the students regarding "generative fluency of reasoning and quality of value orientation" (p. 66). Dialogic pedagogy expedites learners' exploration of meaning, evidence, and application of reasoned arguments (Jamali, 2015) that consequently initiates the development of higher level of cognitive complexity (Reznitskaya, 2012).

Traditionally, the writing courses yielded to the compliance of the EFL learners with a set of lexical, syntactic, and linguistic resources and the role of the EFL teachers was to teach the writing rules (Johns, 1997). However, the EFL learners' mastery of the basic norms of academic writing does not suffice to develop a strong discoursal and authorial voice. Writing is not the demonstration of ideas in a detached, neutral, and impersonal manner, but rather the manifestation of voice. EFL teachers, therefore, should tailor the writing courses that conventionally focus on unity, cohesion, and coherence so as to address voice sufficiently and enhance the learners' writing ability by engaging them in scaffolded dialogues.

**Conclusion**

In conclusion, this study attempted to develop voice into college EFL writing classrooms through dialogical thinking. The finding revealed that utilizing the dialogical thinking during class instruction affect the development of voice in the intermediate EFL learners' writing. In fact, teaching writing just through the basics and rules fails to prepare learners for the scholarly work that they need (Dizon, 2016).

In the light of the findings of the present study, a number of pedagogical implications for teachers, instructors, material developers, and policy makers are put forward. First of all, through dialogic pedagogy, teachers can empower L2 learners to have a voice in the classroom and consequently in the society (Rahimi et al., 2012). Under this approach, the scholarly texts are not consid-
ereed as "perfect, formally organized language patterns and discourses" (Johns, 1997, p.7), rather they are viewed as medium for self-disclosure and discovery and the expansion of an exclusive individual voice. Accordingly, EFL curriculum developers and textbook writers are highly recommended to include these types of tasks as one of the components of instructional materials to enable the EFL learners to voice their opinions and engage them in dialogical thinking tasks. There seems to be an opportunity to improve students’ writing on challenging topics by instructing and practicing the principles of dialogical thinking. The results imply that EFL practitioners not only need to modify their views about writing, texts and voice, but also have to modify their approach to writing and the methodologies they use. Last but not least, it should be mentioned that teaching dialogical thinking and promoting self-voice in writing requires not only the knowledge of how to apply them practically in the class, but also patience, persistence, and meticulous supervision on behalf of the teachers to support the students and get them to engage in the whole process.

References

Doukmak, R. (2014). Are you sure you don’t have any questions? Dialogic teaching as a way to promote students’ questions. ELTED, 16, 27-33.


Harwood, N. (2005). ‘We do not seem to have a theory ... the theory I present here attempts to fill this gap’: Inclusive and exclusive pronouns in academic writing. Applied Linguistics, 26(3), 343–375.


Appendix
The Lesson Plan Based on Alexander’s (2008) Dialogic Teaching

Objectives:
At the end of the term, the students would have been able to:
- Identify the structure of the genre of the argumentative essay.
- Identify the features of the generic structure and the stages of the essay.
- Identify the linguistic features (i.e., hedges, intensifiers, bundles, boosters, and stance taking strategies, point-making) in the moves of the essay.
- Recognize the use of types of active and passive versions.
- Find the various use of stating, developing, and supporting the argument

Session 1:
- An OPT was administered to select the participants who were at intermediate level.
- Students were familiarized with the course requirements and the nature of the writing activities.
- The learners were informed that they had to complete some writing tasks as part of their class requirements and hand the final draft to the teacher for feedback and scoring. However, they were already informed that they were supposed to hand in their best draft because it contributed to their final score in the course.
- The preliminary topics of writing were taught (topic sentence, support sentences, etc.)

Sessions 2 to 9:
- From week 2 to week 9 of the course meetings, the learners spent 35 minutes of the total class time (90 minutes) each week to complete the assignments through multiple in-class drafting, out of class writing, or cooperative writing.
- The teacher delivered the first passage to the students. Teachers and students addressed learning tasks, i.e., topic sentence, hedges, and intensifiers together as group discussions. (Collective principle)
- The participants were asked to summarize the passage in one page and in two pages they were to argue and list in favor of or against any issues raised in the selected passage individually.
- Teachers and students listened to each other, shared ideas, and considered alternative viewpoints. They were asked to discuss the issue in small groups and present their reasons. They were asked to clearly state their positions, justify their viewpoints, challenge others’ views and finally arrive at a consensus. (Reciprocal principle)
- Students articulated their ideas freely without fear of embarrassment over wrong answers. They helped each other to reach common understandings. They were asked to participate in discussion and completed assignments such as analyzing the topic and listing 10 reasons for rejecting or accepting the issue and its effect on society. (Supportive principle)
- Students were supposed to discuss and evaluate opposing viewpoints of the controversial issue in a kind of a consensus task leading to an agreement on the certain issue. In fact, the teacher and students built on their own and each other’s ideas and chain them into coherent lines of thinking and enquiry. (Cumulative principle)
- The teacher delivered the necessary instructions on the topic sentence, hedges, and intensifiers. They practiced cooperatively building the criteria lists and identifying differences in the paradigms. (Purposeful principle)
- The students were asked to explore hedges and boosters across the genre-moves of the essay.
- Next session, they had to handle a written paper based on their new understandings and viewpoints.
- They were asked to follow essay organization principles taught in the course.
- The topic for the next session was introduced.

Session 10:
- The final writing tasks was administered in which the participants were asked to write about two topics. One of them was chosen from among the topics they discussed during the treatment and another one was selected by the teacher.