Impact of Peer-Assessment and Self-Assessment on Paragraph Writing Ability of EFL Learners

*Amaneh Babaie,**Parisa Farrokh

*English Language Department, Rasht Branch, Islamic Azad University, Rasht, Iran

**English Translation Department, Lahijan Branch, Islamic Azad University, Lahijan, Iran

Abstract

This study was designed to investigate the effects of peer-assessment and self-assessment on the writing performance of Iranian intermediate EFL learners. To this end, 45 female learners, who were studying at Kish English Institute in Rasht were selected based on their performance on Oxford Quick Placemen Test (OQPT). The purpose of OQPT is to homogenize the participants on the basis of proficiency level. The participants were assigned to one control and two experimental groups while each group consisted of 15 participants. A pretest of writing was administered to all the participants. Next, a control group and two experimental groups all received treatment for 8 sessions in which the control group received instruction about paragraph writing while the two experimental groups were trained to do self-assessment or peer-assessment respectively in addition to typical writing instruction. Finally, all the three groups were exposed to the posttest of writing. Based on the statistical results, both the self and peer assessment affected EFL learners' writing ability significantly. However, the extent of improvement that occurred was not the same for both experimental groups, that is, peer-assessment was more effective than self-assessment.

Keywords: Assessment; Peer-assessment; Self-assessment; Writing Ability

1. Introduction

The process of language teaching and learning is no longer static, since the advent of new methods and horizons has caused further complications in the development of instruction and education. Presently, educators are reproved if they are losing the chance of assessing their students' performance using different methods. An increasing amount of suggestions are offered to classroom instructors and teachers at universities and schools, to include the learners in the assessment process.

The complexity of writing and its teaching originates from the way of writing itself. It can be a way for individuals to express themselves, and at the same time, it can serve to reflect the individual creator's participation in a social group. Furthermore, it is thought as ability, while this aptitude is itself a procedure subordinate upon a scope of different aptitudes and, in addition, a procedure that is vivid and formed by the writer's changing objectives for writing.

Another complexity of instructional writing originates from the classrooms as instructive settings. Educators negotiate between the classes as a social gathering and individual understudies in that gathering. This becomes a difficult task when the class size is twenty or thirty, and learners are from different cultural or social backgrounds. Besides, educators regularly negotiate between their yearnings to show writing as a deliberate procedure, and to instruct the changed "abilities" thought about as essential to that procedure, in which, aptitudes are controlled by their understudies.

To deal with this mind boggling teaching act, educators of all levels must get to be comfortable with and cautious onlookers of writers and writing, looking for the kind of data about understudies that helps them as instructors react to the inquiries—the difficulties—inalienable in their understudies' endeavors. In different parts of this work, the author surveys the sorts of interrelated research information about writing that may advise educators' perceptions of their understudies and their choices about how best to bolster their understudies' endeavors.

Today, assessment in education is tending towards a radical development from customary testing to assessing learning. In other words, many researchers have attempted to probe into the efficiency of implementing new methods of assessing language learning of different learners, since traditional approaches to writing assessment are not complete due to some reasons. First, they are not wise enough to assess learners´ writing ability based on only one draft, which is written under timed conditions and about an unfamiliar topic. Second, a single piece of writing cannot be a good indicator of the learners´ overall writing ability. In spite of these disadvantages, some teachers still use traditional approaches in assessing writing skill in EFL contexts like Iran (Heidari,2011; Iraji, Enayat, & Momeni,2016; Khodashenas, & Rakhshi, 2017).

In fact, the teachers act like a reader and an editor, first they read the paper and then edit it for grammatical and mechanical mistakes (Vangah, Jafarpur, & Mohammadi, 2016).However, self- and peer-assessment can prepare pupils for effectiveness and improvement in their lives. They are also connected to the purpose of life-long learning and integrated into different subjects and fields (Chen, 2008).Consequently, the reason for this study is to research whether there is any relationship between self-assessment and peer-assessment and writing

performance of Iranian intermediate EFL students. Using the findings in this study can improve the condition and status of language teaching, especially in the context of Iran. The findings may encourage the teachers, who still believe in their own traditional techniques in teaching and writing to shift their attitudes, and follow more practical techniques.

2. Review of Literature

The importance of writing skill is not realized or utilized by teachers and learners, despite that it plays a vital role for a man's scholarly and informative advancement, while it demonstrates one's reasoning improvement. Also, a suitable instructional system can upgrade complex deduction advancement. Furthermore, language, particularly the skill of writing, is a communicative instrument used to present one's idea and culture.

Despite the fact that writing is the skill which a learner needs to learn, the process of teaching this skill starts after the learner masters three other skills, listening, speaking, and reading, possibly because writing is the most muddled skill. The learners must be outfitted with the other abilities before figuring out how to write on the grounds that the other skills are normally reflected through the skill of writing. In this way, the skill of writing needs genuine consideration (Bowen, Madson, & Hilferty, 1985). Learners nowadays can barely convey ideas through writing in the present society due to different issues and complexities in this skill including, but not limited to, vocabulary, expressions, syntactic developments, writing components, and thought organization.

Therefore, improving writing skill has frequently been troublesome and difficult to learners. They are required to devote all the time and exertion in learning and rehearsing, to be able to compose their thoughts by using language components. On the other hand, the multifaceted nature of the instructing and learning writing ability as well as assessing it is additionally a fairly troublesome procedure, including numerous strides. In general, this tedious duty has exclusively been regarded and put on the shoulders' of writing instructors.

Assessment is the way toward collecting and talking about data from different and assorted sources to build up a profound comprehension of what understudies know, comprehend, and can do with their insight as a consequence of their instructive encounters. The procedure finishes when evaluation results are utilized to enhance consequent learning.

Alternative assessment can be extensively characterized as any evaluation strategy that is an option to a conventional standardized test. It applies to any kind of assessment that varies from the numerous test types, such as timed, one-shot approaches and multiple-choice that describes most classroom and standardized evaluation. It is known as a strategy to be recognized additionally

as direct, authentic, and performance assessments. This method incorporates rating scales, checklists, rubrics, diaries, logs, journals, video and audio tapes, conferences, portfolio, peer-assessment and self-assessment.

Lopes (2015), in "Alternative Assessment of Writing in Learning English as a Foreign Language: Analytical scoring and self-assessment," explained that while the instructive advantages of optional assessment are progressively perceived and elective assessment strategies brought into distinctive instruction, in Cape Verde, numerous EFL instructors did not realize its significance. The primary reason of the study is thus to incorporate and survey effective assessment tools to evaluate the writings of EFL pupils of Cape Verde secondary schools. The subject of exact and reasonable assessment of writing most likely constitutes the real quandary in both L1 and L2 writing fields, and EFL field is not a special case. The paper, likewise, expected to make Cape Verdean EFL educators mindful of the diverse sorts of optional assessment that exist to survey understudies, concentrating on analytical and self-assessment device, as vital method for picking up a dynamic image of understudies' scholarly and linguistic advancement.

Self-assessment is presently assuming critical part in language learning and instructing. The technique includes understudies in judging their own learning, especially their accomplishment and learning results. Blue (1994) distinguishes a few advantages of self-assessment, for example, empowering more prominent exertion, boosting fearlessness, and encouraging consciousness of refinements between competence and performance, and also mindfulness of learning qualities and shortcomings. Self-assessment aptitudes help understudies step by step, to build up a basic demeanor toward learning for the duration of their lives and after, to accomplish impeccable self-governance.

In this relation, Boud (1989) maintains that self-and peer-assessment rehearsals have been perceived as meeting an assortment of unessential needs in different settings of advanced education. Firstly, he watches that self-assessment practices may have been preferences for evaluating vast quantities of understudies, in this way permitting understudies to somewhat diminish their educators of the time and exertion of testing. The contention takes after that, if understudies could go up against some assessment undertakings, then scholarly staff would have more opportunity to plan and lead instructively advantageous learning and assessment exercises for them. This research aims to examine the effect of self-and peer-assessment on the writing performance of intermediate EFL students.

Numerous researchers have prescribed that before the genuine evaluation, learners ought to be prepared on the most proficient method to utilize self-assessment and peer-assessment. Based on the ideas of Oscarson (1989),

preparing can build the unwavering quality of learners' self-assessment. Additionally, Birjandi and Siyyari (2010) noticed that, as the students practice and prepare more, they assess accurately. In a study conducted by Jafarpur and Yamini (1995), the incapability of preparing and its effect on self-assessment accuracy was reported. They expressed that preparing helped learners to the degree that they could only judge their peers. Conrad and Goldstein (2009), on the one hand, and Min (2012) and Nakanoshi (2015) on the other, contended that if learners work on peer-assessment, the expertise will assume an essential part in building up their own writing ability. Peer-assessment, unlike the assessment by teacher, gives the learners chances to think and reason in arranging with diverse thoughts prompting the improvement of learners' attention to audience. In a follow-up study done by Khonbi and Sadeghi (2012), the effect of self-assessment, peer-assessment and teacher-assessment techniques on Iranian EFL learners' general English proficiency was investigated. The results of study indicated differences in the effect of the three techniques in favor of peer-assessment. Meanwhile, Meihami and Varmaghani (2013) probed whether self-assessment impacts Iranian EFL learners' writing skill. The results of their study showed that self-assessment significantly affected the writing ability of the students.

In spite of the fact that instructors review or check every bit of work, there is a vast gap in the learning cycle (Farhady, 2003). In this appreciation, self-assessment and peer-assessment can give learners the chance to have a different experience of assessment, to consider their shortcoming and qualities, to figure out how to advance in learning, and lastly, to include effectively in the assessment process prompting another experience.

From an alternate point of view, it is trusted that a suitable bridge between learners' learning and their favored learning styles can enormously impact and enhance learners' learning potential and performance, particularly for EFL learners. In this way, when learning styles might have a part in language learning, does it have similar effect on language assessment and, all the more particularly, on self-and peer-assessment? Albeit such an inquiry is by all accounts of most extreme significance, little work has yet been done in such manner.

The purpose of this study is to examine the effect of self-assessment and peer-assessment on the writing performance of intermediate EFL students. The research question in this study is as follows:

RQ: Does peer-assessment or self-assessment have any effect on Iranian intermediate EFL learners' writing performance?

In order to study the research question in hand, the following null hypothesis was framed.

RH0: Peer-assessment or self-assessment does not have any effect on Iranian intermediate EFL learners' writing performance.

3. Methodology

3.1 Research design

Experimental design was used in the study. The participants of this research were 45 female students at intermediate EFL level, composing three classes at Kish Language Institute. All the participants were female students, who were mainly high school students. Their ages were ranged between 15-18 years old. To respect the quality of the study for further supervision by the instructor and recording the results accurately, the students were divided into three different groups of 15 students. In each class, the students were exposed to same content and teaching method with the same instructor. The first experimental group received self-assessment, the second group peer-assessment and control group received no other treatment than normal writing lessons. Finally, their developments during the process were analyzed and proper comparisons were presented as seen in all groups.

3.2 Instruments

3.2.1 Oxford quick placement test

To be sure of the homogeneity in two experimental groups, proficiency test was administered to establish participants' homogeneity. OQPT was administered to make sure that the participants were homogenous in terms of their language proficiency. This enables teachers to have good understanding about the level of students. The test contains 50 multiple choice questions, which assess student's knowledge of key grammar and vocabulary, a reading text with 10 graded comprehension questions, and a writing task for assessing student's ability to produce the language.

3.2.2 Writing test

In order to compare the effect of treatment on students' writing performance, one writing task was adapted from "Paragraph Development" (1990) by Arnaudet and Barrett, as the pretest. Before the treatment, students were given atopic ("Compare two people you know who are very similar") to write about it during 45 minutes. Students were supposed to present their points of view with convincing evidence, focus on topic and avoid irrelevancies, and use English accurately and appropriately. After attending 8 sessions, a post-test of writing ability was given to all groups. Since the aim of this study was to show the degree of progress from the pre-test to the post-test in all groups of the study, the given topic for the post-test was the same as the topic of the pre-test.

3.2.3 Analytic scoring rubric

The modified version of Wang and Liao's (2008) as a writing scoring rubric was utilized in this study, which consisted of five subscales: focus, elaboration, organization, convention and vocabulary, each with five levels.

3.3 Materials and procedure

In the current study, writing knowledge was aimed. First, the researcher selected three classes at intermediate level to do the research. All the subjects were learning English at Kish Institute in Rasht. Second, OQPT was administered to check the homogeneity of the participants. Next, a pretest of writing subject, "Compare two people you know who are very similar" from "Paragraph Development" (1990) by Arnaudet and Barrett, was administered to measure students' writing knowledge before treatment. Then, the treatment stage started. The treatment consisted 8 sessions of 2 hours each to teach paragraph writing to all students in three groups. The two experimental groups was additionally treated with self- and peer-assessment respectively for 30 minutes duration in every session. Hence, all the groups were equal in terms of the material, teacher and amount of instruction, except the respective type of assessment for the experimental groups in their writing performance, during the study. The instructional course went on for two hours in every session for all groups. Keeping in mind that the end goal is to acquaint the members with self- and peer-assessment, the learners ought to concentrate on the rating scales while evaluating their own and peers' performance on writing. Therefore, the instructor explained the modified Wang and Liao's (2008) rating scale utilized for scoring the participants´ paragraphs in the first session of the treatment. It consisted of five subscales: focus, elaboration, organization, convention and vocabulary, each with five levels, 1 (poor) to 5 (perfect), for each scale, from low quality writing (5) to perfect one (25) for each single writing. Subsequent to expounding on the items, the researcher clarified the explanation by giving a few illustrations. To obviously set up the assessment criteria, every understudy got a sample paragraph, which was assessed before by the researcher in light of the same scale. Then, the participants in experimental groups were acquainted with self- and peer-assessment. In other words, in one experimental group, every learner's writing was assessed by the learner herself and in another experimental group, by the peers of the learner.

In the control group, the instructor used the same writing instruction and material. In contrast to experimental groups, the students in the control group did not receive any treatment regarding self and peer assessment but a traditional method to teach writing. After eight sessions, the posttest was administered to measure the students' progress.

3.4 Data collection and analysis

In the analysis stage of this research, the results achieved from the writing tests were summarized, and the procedures of descriptive statistics (including frequencies, means, standard deviations, etc.) along with inferential statistics namely One–way ANOVA were run. Before running the main statistical analysis of the present study, normality that is the main assumption of parametric tests was established for all of the distributions. Moreover, the reliability of the instruments employed in the study was estimated through a pilot study on (N=10) EFL students, who were representatives of the main group in terms of their general English language proficiency. The results of the study and relevant interpretation are presented in the following sections.

4. Results and Findings

4.1 Reliability analysis for the writing test

Reliability of the writing test was estimated by test-retest method, by giving the same test two times to the pilot study group. Afterwards, the correlation coefficient between the two sets of scores was computed. The test-retest reliability analysis relies on an assumption that no important change should be observed in the participants' English writing proficiency from one test to another using the same writing topic whereas the two tests are separated by a certain time interval. It should be noted that the time interval of two weeks between the two test administrations appears logical. The reliability estimate is given in the following table.

Table 1 Correlations between the first and second administration of the writing test

Correlations		
		Second administration
First administration	Pearson Correlation	.946**
	Sig. (2-tailed)	.000
	N	10

**Correlation is significant at the 0.01 level (2-tailed).

To comment on the results of the Pearson test, Cohen's (1988) guidelines were used. According to this guideline, values between 0.0 and 0.1 show that there is *Little* or *no* relationship between X and Y, values between 0.1 and 0.5 show that X and Y are *weakly* related, and the correlation indices between 0.5 to 0.9 show that the relationship is *strong*. Furthermore, the correlation index of (1) shows *perfect* relationship between the two variables (Cohen, 1988). Consequently, the findings revealed that the relationship between the two

administrations of the same writing test was relatively high (r=.946). As a result, the reliability of the writing test was established through the results of the test-retest reliability analyses.

4.2 Inter-rater reliability analysis

Inter-rater reliability (IRR) was used to assess the consistency between the ratings provided by the two raters, and the degree of agreement between the two raters, who made the independent ratings for the writing tests. In fact, two different raters, who were experienced foreign language teachers scored the written production of the participants. The consistency of the two raters' judgments was tested using inter-class correlation coefficient (ICC) analysis that showed a comparatively high level of inter-rater reliability for the writing test scores given in two administrations in pre and post-tests.

Table 2 Item statistics of scores given by the two raters

Groups		Mean	Std. deviation	N
Control	Rater A pretest scores	12.7333	1.38701	15
	Rater B pretest scores	12.8667	1.59762	15
	Rater A posttest scores	12.8667	1.30201	15
	Rater B posttest scores	13.6667	1.23443	15
Experimental A (Peer-assessment)	Rater A pretest scores	12.9333	1.43759	15
	Rater B pretest scores	12.0667	1.66762	15
	Rater A posttest scores	16.5333	1.76743	15
	Rater B posttest scores	16.2000	1.42428	15
Experimental B (Self-assessment)	Rater A pretest scores	13.2000	1.85934	15
	Rater B pretest scores	12.8667	1.55226	15
	Rater A posttest scores	16.0000	2.07020	15
	Rater B posttest scores	15.9333	1.83095	15

Table 2 displays the mean and standard deviation (SD) of the scores 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 of the control group. However, the scores assigned by Rater A for pre and posttest scores of both experimental A and B groups were higher than those assigned by Rater B. Additionally, the scores given by rater B were less variable than scores given by rater A, except for pre-test scores of the control group and pre-test scores of experimental A (see means & standard deviations in Table 2).

After computing the means and standard deviation for the scores given by the two raters for both pre and post-tests, Inter-class correlation coefficients (ICC) were computed individually for the pre and post-tests of writing.

Table 3 Inter-class correlation coefficients (ICC)of the scores given by two raters for pre and post-test scores

	Groups		Inter-class correlation	95% Confidence Interval		Sig.
				Lower Bound	Upper Bound	
Pre-test	Control	Average Measures	.85	.57	.95	.000
	Experimental A (Peer-assessment)	Average Measures	.83	.49	.94	.001
	Experimental B (Self-assessment)	Average Measures	.85	.55	.95	.001
Post-test	Control	Average Measures	.94	.84	.98	.000
	Experimental A (Peer-assessment)	Average Measure	.88	.64	.96	.000
	Experimental B (Self-assessment)	Average Measures	.91	.74	.97	.000

The estimated inter-rater reliability between the two raters for the pretest scores of the control group was (r=.85), with 95% CI (.57, .95), that for the experimental group A was (r=.83), with 95% CI (.49, .94), and for the experimental group B was (r= .85), with 95% CI (.55, .95) which were acceptable. Furthermore, the estimated reliability between the two raters for the post-test scores of the control group was (r=.94), with 95% CI (.84, .98), that for the experimental group A was (r=.88), with 95% CI (.64, .96) and for the experimental group B came to (r= .91), with 95% CI (.74, .97), which were acceptable. Therefore, the reliability of the writing measurement for the pre and post-test of writing between the two raters was supported.

4.3 Examination of normality of distribution

The parametric test of Pearson product-moment correlation was employed to estimate the test-retest reliability analysis for the writing test. Moreover, One-way ANOVA was run to provide answer to the research question. The assumption of parametric tests, including Pearson correlation and One-way ANOVA i.e. normality was examined before running the main statistical analyses. Skewness analysis and trimmed means were used to examine the normality of the scores. The table 4 presents the association between these distributions, and shows the normal distribution of these variables.

Table 4 Statistics for the first and second administration of the writing test

Groups			Rater A pretest	Rater B pretest	Pretest scores	Rater A posttest	Rater B posttest	Posttest scores
Control	Mean		12.73	12.86	12.80	12.86	13.66	13.26
	95% Confidence Interval for Mean	Lower Bound	11.96	11.98	12.02	12.14	12.98	12.58
		Upper Bound	13.50	13.75	13.57	13.58	14.35	13.95
	5% Trimmed Mean		12.70	12.85	12.75	12.85	13.62	13.24
	Skewness		-.005	.370	.363	-.167	.214	-.030
	Kurtosis		-1.505	-.025	-.877	-1.204	-.782	-.844
Experimental A (Peer-assessment)	Mean		12.93	12.06	12.50	16.53	16.20	16.36
	95% Confidence Interval for Mean	Lower Bound	12.13	11.14	11.70	15.55	15.41	15.52
		Upper Bound	13.72	12.99	13.29	17.51	16.98	17.20
	5% Trimmed Mean		12.87	11.96	12.41	16.53	16.16	16.35
	Skewness		.632	1.159	1.161	.116	.452	.258
	Kurtosis		-.108	1.297	.873	-1.296	-.632	-1.43
Experimental B (Self-assessment)	Mean		13.20	12.86	13.03	16.00	15.93	15.96
	95% Confidence Interval for Mean	Lower Bound	12.17	12.00	12.14	14.85	14.91	14.92
		Upper Bound	14.22	13.72	13.91	17.14	16.94	17.00
	5% Trimmed Mean		13.22	12.85	13.06	16.05	15.92	15.99
	Skewness		-.028	.122	-.085	-.279	.435	.022
	Kurtosis		-.983	-1.54	-1.67	-.467	-.656	-.528

In Table 4, the descriptive statistics including the means, 95% confidence interval for the means, the trimmed means, and values of Skewness and Kurtosis for the variables are presented. To compute the 5% trimmed mean, the top and bottom 5 percent of the cases were removed, afterwards, a new mean value was calculated, that was the trimmed mean. Then, the original mean scores and the new trimmed means were compared to determine the differences for all the tests, and it was found that the extreme scores did not have a strong effect on the means. The trimmed means were all within the range of 95% confidence interval for the means. In other words, since the trimmed means and the original mean values were not much different for the writing tests of both groups, the values were not too different from the remaining distribution. Consequently, these cases were retained in the data file.

Additionally, the results of Skewness analyses for both first and second administration of the test are shown in Tables 4. It was obtained by dividing the

statistics of Skewness by the standard error, and it showed that the assumption of normality was observed in the distribution of the scores and the distribution were all normal and symmetric. In other words, the Skewness and Kurtosis values reported in the table were all within the range of ± 2, supporting that the distributions were normal.

4.4 OQPT results for sampling purpose

To choose homogenous subjects as the sample with respect to their general language proficiency, OQPT was administered to sixty (N= 60) EFL students. The participants took three-part test that included items related to structure, vocabulary, and reading comprehension with a maximum possible score of sixty (60) points. A cut-point of one standard deviation above and below the mean was set, and 45 EFL learners whose proficiency scores were within this range ($\pm$ 1 SD from the mean) were selected as the main participants of the present study. The descriptive statistics for OQPT is available in Table 5.

Table 5 Statistics for OQPT

N	Valid	60
	Missing	0
Mean		32
Median		31
Mode		29
Std. Deviation		4
Variance		17
Skewness		1
Std. Error of Skewness		.30
Kurtosis		1.18
Std. Error of Kurtosis		.60
Range		17
Minimum		27
Maximum		44
Sum		1932.00

Table 5 delineated the findings of group statistics for OQPT scores that was given to pick out homogeneous subjects, with regard to their general English language proficiency. Measures of central tendency including mean, median, and mode together with measures of dispersion such as range, variance, and standard deviation as well as measures of distribution (i.e., Skewness and Kurtosis) were computed for OQPT. Thus, the cut-point of ($32\pm$ 4) was set and (n=45) EFL learners whose proficiency scores were within this range (28 to 36), were selected as the main participants of the present study.

4.5 Examining the Research Question

4.5.1 Descriptive statistics for the pre-test scores of writing test

To answer the research question about whether or not the self and peer assessment have significant effect on EFL learners' writing ability, one-way ANOVA was used to analyze the scores of writing pre-test for both control and experimental groups. This is to examine if there is initial difference between the two groups in terms of writing ability. Before applying one-way ANOVA, homogeneity of variances was examined by means of Levene's test and the results are presented in Table 6.

Table 6 Levene's test to examine the homogeneity of variance among pre-test scores

	pretest scores		
Levene's statistics	df1	df2	Sig.
1.164	2	42	.322

The finding of Levene's test has shown that the group variances are not statistically different among the scores of writing pretest since the p-value or $P_{\text{pre-test}}$ equals to .322 and is greater than .05. The result of Levene's statistics thus supported the hypothesis that there is no significant difference in terms of group variances of the writing pretest scores.
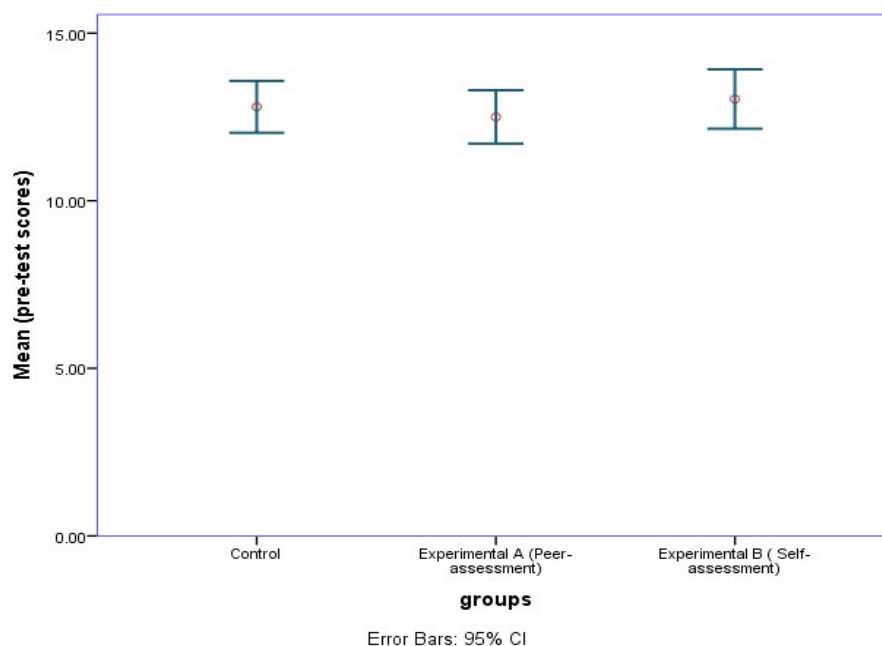

Error Bars: 95% CI

Figure 1 Error bars for examining the homogeneity of variances of pretest

Figure 1 depicts that the average performance was nearly the same for the three groups at the beginning of the study. Furthermore, the extent of diversity in performance was almost identical. ANOVA assumes fairness of variance across the groups, and that assumption was established for these distributions. The following table represents descriptive statistics for the three groups that was run to compare the means of the three groups, in the pre-test with respect to their writing ability.

Table 7 Descriptive results for the pretest scores of writing

	N	Mean	SD	Std. Error	95% Confidence Interval for Mean		Pre-test minimum score	Pre-test maximum score
					Lower Bound	Upper Bound		
Control	15	12.80	1.39	.36	12.02	13.57	11.00	15.50
Experimental A (Peer-assessment)	15	12.50	1.43	.37	11.70	13.29	11.00	15.50
Experimental B (Self-assessment)	15	13.03	1.59	.41	12.14	13.91	10.50	15.00
Total	45	12.77	1.46	.21	12.33	13.21	10.50	15.50

The results of the one-way ANOVA indicated that there was no significant difference in pre-test scores among the Control group ($\bar{X}$ = 12.80, SD = 1.39), Experimental group A ($\bar{X}$ = 12.50, SD = 1.43) and Experimental group B ($\bar{X}$ = 13.03, SD = 1.59), (F = (2, 42) = .489, p = .617 > .05) (see Table 8). Thus, the groups were equal at the beginning of the study, with respect to their writing ability.

Table 8 One-way ANOVA for the pretest scores of writing

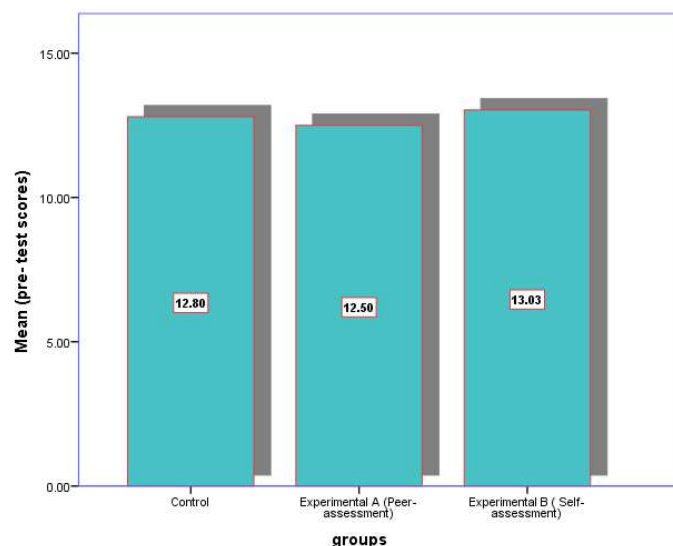|	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	2.144	2	1.072	.489	.617
Within Groups	92.133	42	2.194		
Total	94.278	44			

Figure2Comparison of the three groups on pretest of writing

4.5.2 Descriptive statistics for the post-test scores of writing test

To examine the possible effects of the two types of assessment (peer-assessment and self- assessment) on Iranian intermediate EFL learners' writing ability, One-way ANOVA was used. Levene's test was run to the results of the post-tests and the model assumption was checked (see Table 9).

Table 9 Levene's statistics for the test of homogeneity of variances of posttest scores

	posttest scores		
Levene's statistics	df1	df2	Sig.
1.228	2	42	.303

Table 9 presents the equality of variances across the posttest scores. The significance value for the dependent variable (posttest scores of writing) is greater than 0.05, so it can be concluded that the equal variances assumption is met for this variable ($F_{posttest}(2, 42)=1.228$, sig(.303) ≥ .05).

After confirming the homogeneity of variances assumption, descriptive statistics was run to the results of post-test of writing test. Table 10 lists the means and standard deviations, as well as other values, for each group on post-test.

**Table 10 Descriptive statistics for writing scores (post-test)**

	N	Mean	SD	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Posttest minimum score	Posttest maximum score
Control	15	13.26	1.23	.31	12.58	13.95	11.50	15.50
Experimental A (Peer-assessment)	15	16.36	1.51	.39	15.52	17.20	14.50	18.50
Experimental B (Self-assessment)	15	15.96	1.87	.48	14.92	17.00	12.50	19.00
Total	45	15.20	2.06	.30	14.57	15.82	11.50	19.00

A comparison of the means across the three groups showed that the experimental group A ($\bar{X}$= 16.36) performed generally better than the experimental group B ($\bar{X}$ = 15.96), as well as the control group C ($\bar{X}$ = 13.26).

Table 11 One-way ANOVA for the three groups on posttest of writing

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	85.300	2	42.650	17.408	.000
Within Groups	102.900	42	2.450		
Total	188.200	44			

The results revealed that the three groups were different at the end of the study ($F = (2, 42) = 17.408$, $p = .000 < .05$) (See Table 11). Thus, the groups were statistically different at the end of the study with respect to their writing ability. The following figure illustrates the mean plot for the results of the post-test of writing.
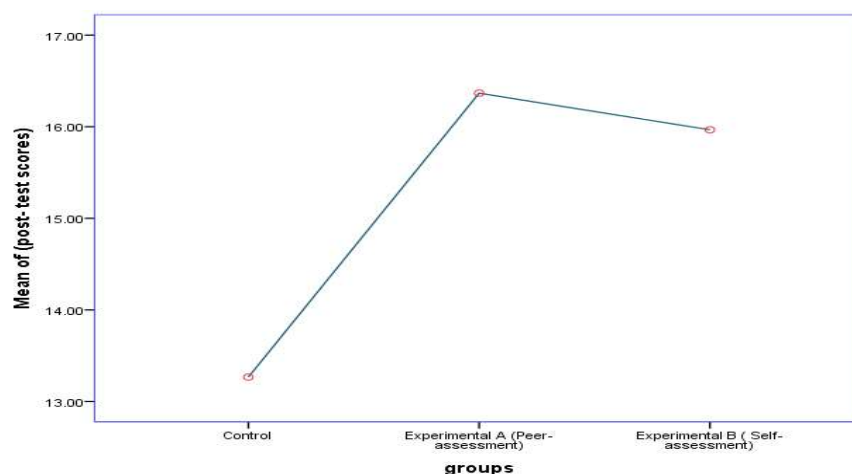
Figure 3 Mean plot for the results of the post-test of writing

This time, the results revealed that different types of assessment affected the writing ability of the three groups, differently ($p = 0.00 < 0.05$). In fact, learners' performance in experimental groups far outweighed that of the control group, in the post-test. The relatively high value of $F = 17.408$ showed that there was, indeed, a significant effect of the two types of assessment namely, peer-assessment and self-assessment on Iranian intermediate EFL learner's writing ability. After the revelation that the groups differed in some way, post-hoc test disclosed more about the structure of the differences. In other words, to find out the location of the differences among the three groups, a Scheffe test was conducted. The table 12 makes multiple comparisons among the three groups based on the results of Scheffe test.

Table 12 Post-hoc Scheffe test for post-test of writing

(I) groups	(J) groups	Mean Difference (I-J)	Std. Error	Sig.	Scheffe 95% Confidence Interval Lower Bound	Upper Bound
Control	Experimental A (Peer-assessment)	-3.100*	.571	.000	-4.5504	-1.6496
	Experimental B (Self-assessment)	-2.700*	.571	.000	-4.15	-1.24
Experimental A (Peer-assessment)	Control	3.100*	.571	.000	1.64	4.55
	Experimental B (Self-assessment)	.400	.571	.784	-1.05	1.85
Experimental B (Self-assessment)	Control	2.700*	.571	.000	1.24	4.15
	Experimental A (Peer-assessment)	-.400	.571	.784	-1.85	1.05

*. The mean difference is significant at the 0.05 level.

When the reference is made to the mean difference of the three groups on posttest of writing test, it can be seen that the positive influence of peer-assessment has made the largest difference between the mean scores of the experimental group (A) and the control group (C) (mean difference= 3.10). On the other hand, the lowest mean difference was seen between group (A) and group (B) (mean difference= .40). Therefore, the hypothesis that different types of assessment (peer versus self-assessment) does not have any effect on Iranian intermediate EFL learners' writing ability is rejected.
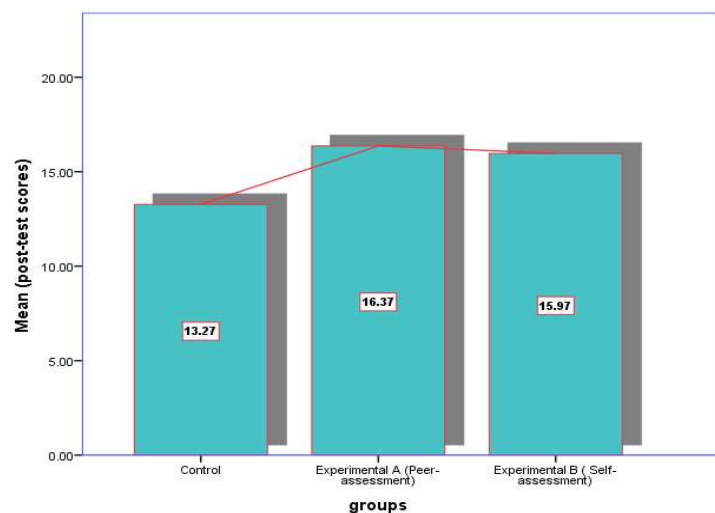
Figure 4 Comparison of the three groups on posttest of writing

5.Discussion

The results of the study affirmed that providing opportunity for the students to use self- and peer-assessment is helpful and influential in improving their writing ability. The results showed significant change in the experimental groups. The findings of the present study are in line with the results of numerous studies concerning the valuable use of self- and peer-assessment, to improve the paragraph writing abilities among EFL learners. Birjandi and Siyyari (2010) explored the impact of doing self-assessment and peer-assessment on the section of writing performance and their accuracy on Iranian students of English major.

The results are also consistent with the findings of previous studies in this field, such as those of Conrad and Goldstein (2009), Min (2012), Nakanoshi (2015) and Khonbi and Sadeghi (2012). It is that peer-assessment, as it may, ended up being more compelling in enhancing the writing performance of the learners than self-assessment. In other words, the results of the present study corroborate the results of the previous findings in this field i.e. the impact of self- and peer-assessment on EFL learners' writing abilities.

Moreover, this study is congruent with the one, conducted by Saito and Fujita (2004). They found that peer-evaluation of writing quality is more like educator's evaluation than self-evaluation. To legitimize this discovery, they advanced mental components, for example, understudies' self-esteem, confidence, social estimation of humility and propensities for overestimating self-capacity as mindful for this finding. On the off chance, their avocation is thought to be correct, and it appears that the current study was sufficiently effective in controlling these mediating variables. It should be noticed that in the current research, the understudies realized that no high-stake choice was to be made, taking into account their peer- or self-evaluations, despite the fact that the

understudies demonstrated no confidence and readiness in the first place, the members were continuously urged to realize that they could make it, as an instructor or master rater does. Likewise, it ought to be noticed that the members of this study were given a leaflet containing full depictions, outlines, and scripts on what constituted top notch writing.

The self-and peer-assessment group of understudies showed changes in their ordinary rating accuracy, after eight sessions of self-and peer-assessment forms. And the differences between these two gatherings in the degree they had improved in rating exactness that was not noteworthy, both in self-evaluation and peer-evaluation. In other words, the effects of self- and peer-evaluation on the rating precision of the three gatherings were comparable. Also, the individuals' assessing change is parallel with the adjustment in their writing performance. Thus, the conclusion is that the more able the understudies get in their compositions, the better they understand what constitutes better composition, and regularly the more they show precision in their assessments.

The advancement was seen from the pre-test to the post-test in three groups i.e. the treatment and the control. In any case, the members in the treatment group, who received self- or peer-assessment treatment of this study, could do better writings in contrast with the control group, who got no treatment in terms of self or peer assessment. Consequently, the present study has uncovered that using self or peer assessment process in class for education has several advantages as compared with conventional types of teaching writing (i.e. just teaching different parts of writing). In any case, the study has demonstrated that the treatment group, who got teaching and assessment evaluation, exceeded the control group in their writing performance capacity. Accordingly, the requirement for self or peer-assessment seems to be advantageous to prepare EFL learners and EFL classes for writing. Instructors must pay consideration on it in their classes for EFL learners.

6. Conclusion

To conclude, the statistical results of the study confirms that both self and peer-assessment affect the Iranian EFL learners' writing ability, positively. However, the extent of improvement that occurred was not the same for both the experimental groups; that is, peer- assessment would help learners, largely in paragraph writing to perform better than those learners, who received self-assessment. From the statistical inference of the post-hoc Scheffe test, a significant difference can be seen among the three groups. The significant difference is between the control group and experimental group A, as well as the control group and the experimental group B. However, the difference between the two experimental groups was not statistically significant. Hence, language

instructors, particularly those educating the composition ability, are suggested to apply more instructive practices, such as self- and peer-assessment in their instructing; this matter can promise both the learning of the understudies and expanding their inspiration, which is a critical variable in learning.

References

Arnaudet, M. L., & Barrett, M. E. (1990). *Paragraph development: A guide for students of English*. New Jersey: Prentice Hall.

Birjandi, P., & Siyyari, M. (2010). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *Iranian Journal of Applied Linguistics, 13*(1), 23-45.

Blue, G. M. (1994). Self-Assessment of Foreign Language Skills: Does It Work? *CLE Working Papers*, *3*, 18–35.

Boud, D. (1989). The role of self-assessment in student grading. *Assessment in Higher Education, 14*(1), 20-30.

Bowen, J. D., Madson, H, S., & Hilferty, A. (1985). *TESOL techniques and procedure*. London: Newbury House Pub.

Chen, Y. M. (2008). Learning to assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*(2), 235-262.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associate.

Conrad, S., & Goldstein, M. (2009). Self-, peer-, and teacher-assessments in Japaneseuniversity EFL writing classrooms. *Language Testing*, *26*(1), 75-100.

Farhady, H. (2003). Classroom assessment: A plea for change. Paper presented at *the Second Conference on Issues in English Language Teaching, Tehran, Iran.*

Heidari, F. (2011). Alternative assessment procedures in Iranian EFL writing classes: The washback effect and learners' attitudes. *Iranian Journal of Applied Language Studies, 3*(1), 51-76.

Iraji, H.R., Enayat, M. J., & Momeni, M. (2016). The effects of self- and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies, 6*(4), 716-722.

Jafarpur, A., & Yamini, M. (1995). Do self-assessment and peer-rating improve with training? *RELC Journal, 26*(1), 63-85. DOI: 10.1177/003368829502600104.

Khodashenas, M. R., & Rakhshi, F. (2017). The effect of electronic portfolio assessment on the writing performance of Iranian EFL learners. *International Journal of Research in English Education, 2*(3) 67-77.

Khonbi, Z. A., & Sadeghi, K. (2012). The effect of assessment type (self vs. peer vs. teacher) on Iranian university EFL students' course achievement. *Language Testing in Asia, 2(4),* 47-74.

Lopes, L. S. (2015). *Alternative assessment of writing in learning English as a foreign language: Analytical scoring and self-assessment.* Cambridge: Cambridge University Press.

Meihami, H., & Varmaghani, Z. (2013). The implementation of self-assessment in EFL writing classroom: An experimental study. *International Letters of Social and Humanistic Sciences, 9,* 39-48.

Min, H. T. (2012). The effect of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*, 118-141.

Nakanoshi, C. (2015). The effects of different types of feedback on revision. *The Journal of Asia TEFL*, *4*(4), 213-224.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and implications. *Language Testing*, *6*(1), 1-13.

Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, *8*(1), 31-54.

Vangah, F. P., Jafarpur, M., & Mohammadi, M. (2016). Portfolio assessment and process writing: Its effect on EFL students´ L2 writing. *Journal of Applied Linguistics and Language Research, 3*(3), 224-246.