

The Effect of Individual and Collaborative Assessment Training on Iranian Male and Female EFL Learners' Writing

Zahra Riahi

Email: riahi_z@yahoo.com
IAU Research and Science Center,
Khouzestan Branch, Iran;

Abstract

Individual and collaborative assessments have engaged all researchers' attention over the last century. This study explored the probable improvement in Iranian EFL learners' writing with respect to individual and collaborative self-assessment training as well as examining the effect of gender on assessment training. The participants in this study were 80 (40 males, 40 females) English learners in a language institute. This sample was divided into four equal groups. Two groups worked on assessment individually, and two groups performed assessment collaboratively. Data were collected from a self-assessment questionnaire and writing scores based on IELTS task 1 writing band score descriptors and Magoosh Essay Rubric for the GRE and GMAT. Results revealed the positive effect of collaborative assessment training on writing. Additionally, the Two-Way Analysis of Variance (Two-Way ANOVA) indicated that the female participants in the collaborative group outperformed males in the same group. The study provides implications for teachers and institutional policymakers in how to improve English learners' writing proficiency.

Keywords: Assessment Training, Collaboratively, Individually, Gender, Writing Skill

1. Introduction

Most learners decide to learn English with the aim of being able to speak, but few students consider the importance of learning and practicing English writing. If one wants to be fluent in English, one should aim at learning all language skills (Yang, 2013). It seems that second language (L2) writing has established its status in current understandings of language education. L2 writing proficiency plays an important role in any individual's educational and professional progress (Costa, 2019; Simin & Tavangar, 2009).

As a productive skill, writing has received much attention due to the effect it can exercise on international communication. Therefore, writing skill, like other skills, can be rated and assessed. Assessment has been described as an essential pedagogical activity process (Bijsterbosch et al., 2019; Cheng, 2017). According to Ellis and Yuan (2004), teachers utilize writing as an important skill to evaluate learners' success, failure, achievement, and progress. Students are not usually able to evaluate their writing task accurately (Bjork, 1999; Fahimi & Rahimi, 2015).

Two major types of assessment are used in evaluation, including teacher-assessment and self-assessment. In traditional courses of writing, the teacher provides some information needed to write about the topic for learners and then asks learners to write a text that includes certain criteria and finally gives feedback. Learners improve their writing based on feedback. On the other hand, in self-assessment, the learner participates in the feedback process (Comert & Kutlu, 2018).

Self-assessment is one of the most significant elements in strategy as well as autonomous learning, which enable learners to evaluate their progress. If there is not any room for learner-directed learning, learners become passive and demotivated because no progress can be experienced. Self-assessment provides a situation for active and focused students to assess their progress according to communication (Harris, 1997). Students can evaluate and judge their performance based on their criteria, learning goals, and expectations (Henner-Stanchina & Holec, 1985).

As Chamot and O'Malley (1994) pointed out that: learners perform different types of thinking skills in higher order and strategies of learning by using self-assessment as well as self-rating strategy provides feedback to them and can be the direction for their future learning. In this way, there is little individual goal orientation that leaves students any room for self-assessment. Blue (1988) suggested that self-assessment has many constraints as the procedure of accurately assessing learner proficiency and called learners' ability in assessing their progress.

Assessment individually and collaboratively has engaged various researchers' attention over the last century (Ballantyne et al., 2002; Brown, 2004; Herbert, 2006; Kurt, 2014; Mangelsdorf, 1992; Michael & Valdez, 1996; Topping, 1998). The motivation for introducing and combining these types of assessments is that students can show more responsibility in learning a language. It is believed that self-assessment may assist students to be responsible for their learning (Brown & Hudson, 1998; Meihami & Esfandiari, 2020). Another characteristic of this assessment is locating and pinpointing the strengths and weaknesses of students. In this way, they can think about their weaknesses and attempt to provide a way for compensating their shortcomings (Davison & Leung, 2009).

The other significant point of assessment is related to the balance that is made by the cooperation between students. The next effective point is associated to the constructive effect of assessment on students' language learning (Hasani & Moghadam, 2012). In large classes that teachers do not have enough time to monitor all members, assessment individually and collaboratively can play a crucial role.

Moreover, several studies have revealed gender as one of the significant variables that can influence students' achievement and interest (Bernat & Liayed, 2007; Zeynali et al., 2012) as well as their writing skills (Beard & Burrell, 2010; Troia et al., 2013). Regarding assessment training, males and female students might have differences in writing performance (Beard & Burrell, 2010; Olinghouse, 2008; Troia et al., 2013). However, little research has examined both assessment training

(individual and collaborative) types together in male and female learners. Thus, this research explores the effect of individual and collaborative assessment on Iranian EFL learners' writing skill.

This study addressed the following research questions:

1. Does assessment training have any significant effect on Iranian EFL learners' writing skill in individual and collaborative groups?
2. Does gender have any significant effect on Iranian EFL learners' writing skill in individual and collaborative groups?

2. Literature Review

2.1. Writing assessment

Nevertheless, a bulk of research has shown that there is much variation in the scores given to essays (Alias, Masek, & Salleh, 2015; Brown & Abeywickrama, 2010; Brown & Hudson, 1998). Learners know about various scoring by raters which can increase anxiety in their writing process. Passing L2 writing exams is an essential part of graduation; therefore, the consistency and reliability of scoring are necessary for students (Arter & McTighe, 2001; Perlman, 2003). It should be taken into consideration that assessing and scoring L2 writing requires a meaningful and time-consuming effort (Honsa Jr, 2013).

Lumely (2002) conducted a research on criteria of assessment in a large-scale writing test and also explored that how raters signify these criteria. This study aimed to investigate the procedure that marking decisions were made by raters and they used an analytic rating scale that was designed for multiple-choice tests. In this study, four reliable, practiced, and trained assessors contributed and marked two sets of 24 texts. After that, four raters provided some think-aloud protocols in which they described the process of rating the second set. The researcher took advantage of assigning a code to explain the data of thinking aloud for analyzing the subsequence of ratings, and the explanation the assessors suggested about the categories of scoring in the analytic rating scale.

According to Andrade, Du, and Wang, (2008), teachers keep their criteria for assessment without informing what constitutes their assessment of L2 writings, generating the inconsistent assessment of student performance. For example, one teacher may prioritize the linguistic structures in the assessment, while another may be more interested in idea development. This always results in disputes between teachers and students when the results are publicized. The results are impressionistic judgment of writing proficiency that depends more upon teacher-raters and text qualities. This kind of assessment has been argued not to be successful in making a true distinction among learners' writing performance (Huang, 2012).

Recently, Mede and Atay (2017) studied Turkish English teachers' assessment literacy at the preliminary syllabus suggested by Turkish private and state universities. Most of the Turkish EFL teachers were untrained, and they needed more advanced teaching in preparing classroom tests, applying tests that were ready-made, giving feedback on assessment, and utilizing self-assessment or peer-assessment. They found that most of the participants required basic or advanced assessment training.

2.2. Self-assessment

Self-assessment has developed as a specific research in pedagogical purposes of L2 learning since 1976. Students are able to evaluate their performance based on assessment criteria by self-reflective and self-assessment activities (Blanche & Merino, 1989). According to Noels, et al. (2000), self-assessment could be used as a method for enhancing motivation and improving learners' autonomy, and in this environment, they did not feel more anxious.

Sajedi (2014) proposed the impact of self-assessment on EFL learners' writing performance. The teacher asked learners to rate their compositions based on sub-skills such as organization, content and structure, and grammar. The findings in this study recommended that learners in the experimental group significantly performed composition better than learners in the control group. Moreover, Kostons, van Gog, and Paas (2012) asserted that for improving self-regulated learning, the participation of learners in self-assessment of their performance on a learning task could be more effective. The findings in this study recommend that self-assessment and task-selection skills have the significant effect on self-regulated learning, and teaching these skills can improve students' knowledge which can achieve from

self-regulated learning. In addition, they were often agreed with the impact of self-assessment on their writing (Banlı, 2014; Bayat, 2010; Bing, 2016; Cömert & Kutlu, 2018; Lam, 2010; Oscarson, 2009).

2.3. Collaborative assessment

According to Brown (2004), self-assessment and peer-assessment improve students' autonomy and motivation. Moreover, there have been a relatively large amount of experimental studies focusing on the impact of peer assessment and teamwork on writing performance (Meihami & Esfandiari, 2020; Nicol, 2010; Sadler & Good, 2006).

Peer assessment can lead to student-centered learning and prepare learners to edit their work (Lejk & Wyvill, 1996). It could have critical pedagogical values for students (Patri, 2002) because they assess their peers' learning by participating in the process of evaluation (Cheng & Warren, 2005).

Collaborative writing in EFL contexts can be described as a team working to perform a common goal while the members communicate and negotiate with each other (Alhadabi & Karpinski, 2020; Abrams, 2019; Hsiu-Chen Hsu, 2019; Orikassa, 2012;). According to McDonough, De Vleeschauwer, and Crawford (2018), EFL college learners from Thailand composed accurate and precise paragraphs when they wrote collaboratively compared to learners who performed individual writing.

2.4. Rubrics

One of the most important writing evaluation tools is a rubric. Rubrics can explain teachers' expectations clearly and help learners in self-evaluation and identify their strengths and weaknesses (Panadero & Jonsson, 2013). The term "instructional rubrics" refers to the rubrics supporting the writings of learners and developing their works with respect to standards of writing and assessment (Andrade, 2001). Instructional rubrics have some advantages for learners, including (a) rubrics can define and describe the quality of writing as authentic writing, (b) rubrics can spot the weaknesses in learners' writing, and (c) they can help learners in avoiding errors as well as rubrics can help learners in evaluating the progress of their works (Arter & McTighe, 2001).

An acceptable and general educational rubric definition claims that a rubric is a scoring tool for the qualitative scoring of valid written texts (Karkehabadi, 2013; Muhammad, Lebar, & Mokshein, 2018). These rubrics consist of some criteria rating the essential criteria of assessment proficiency and various standards of attaining criteria. The rubric defines for both teachers and learners what should be considered significant and what should be looked for in the assessment performance (Arter & McTighe, 2001; Cabrera, et al. 2017, Perlman, 2003).

For example, Shi (2001) made a comparison between native and non-native raters as well as between experienced and novice ones. The purpose of the study was to explore which group outperforms in using rubrics in assessing L2 essay writing. The findings indicated that each rater used their criteria in scoring L2 writing, but the experienced raters scored more closely up to the standard rubrics.

Cohen, Lotan, Scarloss, Schultz, and Abram (2002) conducted research on the quality of writing of content area with the beneficial effect of rubrics. In this study, 163 learners took part, and instructional rubrics were employed in five 6th grade classes. These groups were assessed based on their content knowledge and showed a high improvement on formative assessments and essays. In addition, Schooen (2012) conducted a comparative study between trained raters and the raters who never received any training in scoring. Peculiarly, the results suggested that those who did not receive any training scored the writing more precisely rather than the trained raters. Results also revealed the impact of previous achievement and treatment on all writing scores. In addition, the best impact of treatment on scores for participants' criteria on rubric was observed. Findings indicated that employing a model in generating rubric criteria for all assignments and self-assessment improves learners' writing (Andrade et al., 2008).

According to Bradford et al., 2016; Thompson, 2013; and Zhang, 2012, using assessment rubrics can enhance learners' writing skills. Sundeen (2014) conducted a research on instructional rubrics for writing quality. Participants were 89 high school students in the 10th grade. To collect the data, the researcher used the 6-Point Writer's Rubric developed by Education Northwest for assessing essays and compositions. The six criteria in this set of the rubric are ideas, voice, organization, sentence fluency, conventions of writing, and word choice. All participants along with three English teachers showed the effect of explicit rubric teaching in improving writing.

There is another study performed by Andrade and Bouley (2003). These researchers examined the rubric referenced self-assessment in writing. They focused on examining the effect of self-assessment on students' written essays. These participants were in the 7th and 8th grades. In this study, they wrote two essays on historical fiction and a response to literature. Researchers taught learners instructional rubrics based on criteria and gradations of the best quality for these essays. Findings indicated the positive effect of self-assessment on the writing skill.

2.5. Gender

In a study conducted by Read, Francis, and Robson (2005), they reported on results concerning essay assessment and gender in history essays. This study focused on the assessment of and feedback given to two sample essays by 50 historians based in England and Wales universities. Findings revealed a significant variation attributed to the classification of essays and to negative and positive points and interpretations put about quality and to back up the concept that the 'quality' of writings and essays for assessment is eventually constructed by everyone that reads essays and cannot be objectively determined. In this study, researchers concluded that gender constructions could be evident in presentation more than practice.

According to the above-mentioned empirical studies, few researchers have addressed the effect of individual and collaborative assessment training on learners' writing. Therefore, this literature gap in this regard is filled by doing this research. Furthermore, this study tries to investigate the probable improvement in male and female EFL learners' writing with respect to both types of assessments (collaborative and individual).

3. Methodology

3.1. Participants

The participants in the study were 90 English learners in an Iranian English language institute, Iran. Their homogeneity was examined by using a Preliminary English Test (PET) and 80 learners were selected. Their level was upper-intermediate. There were 40 males and 40 females in the sample, aged between 18 and 22 years. To evaluate the effect of assessment training on learners' writing, the researcher divided this sample into four groups of equal size. Two groups worked on assessment individually, and other groups worked collaboratively. The experiment lasted twelve 60-minute sessions.

3.2. Instruments

To homogenize the learners, the Preliminary English Test (PET) was administered. The sample of this test used in the present study contains three parts: listening (25 items), writing (7 items), and reading (35 items). The reliability of 67 items of the PET test was estimated by the researcher, which was 0.82.

This study employed an assessment questionnaire in order to investigate the probable effect of self-assessment on improving learners' writing skills. This questionnaire was a reliable and suitable part of English as a Second Language Program (ESLP) 182 questionnaire. The total reliability of the self-assessment questionnaire is 0.95. Since the reliability of $0 < \rho < 1$ is considered reliable, this questionnaire can be reliable. This questionnaire has questions that include various aspects of students' writing self-assessment; punctuation, grammar, as well as mechanics skills; general writing strategies; before writing; during writing; after writing, punctuation, grammar and mechanics strategies; general learning strategies. The participants were asked to evaluate their writing skills for each item on a Likert-type scale between 1 and 5, ranging from never or almost never true to always or almost always true.

The next instrument in collecting and analyzing data was writing and analyzing scores based on IELTS task 1 writing band score descriptors (Appendix A) and Magoosh Essay Rubric for the GRE and GMAT (Appendix B) in learning writing and assessing it. IELTS task 1 writing band score descriptors assess four criterion areas: coherence and cohesion, task achievement, lexical resource and grammatical range as well as accuracy. Magoosh Essay Rubric for the GRE and GMAT evaluates the following criteria: quality of ideas, organization, writing style, grammar and usage, as well as the summary. The teacher defined essay rubrics for helping learners to evaluate their writings and practicing them to assess their writings.

3.3. Procedures

At the beginning of the experiment, the Preliminary English Test (PET) was performed to make certain learners' homogeneity, and 80 EFL upper intermediate male and female participants were selected from 90 learners. Then, the participants were randomly assigned to four groups: one female group and one male group performed assessment training individually, and the other ones (one female group and one male group) received their treatment collaboratively.

At the second stage, the ESLP questionnaire was administered to the four groups. Although the learners did not have any information about the assessment, the teacher asked them to fill the questionnaire. Then, they were requested to do a piece of IELTS sample writing and assess it themselves because learners did not have any information or idea about the assessment process in their writings at this stage. This was done to be the basis of the following performance of students in the writing skill. Then, the teacher scored all students.

At the third stage, the teacher taught all the groups the criteria of writings rubrics based on IELTS task 1 writing band score descriptors and Magoosh Essay Rubric for the GRE and GMAT. Four criterion areas of IELTS task 1 writing such as coherence and cohesion, task achievement, lexical resource and grammatical range and accuracy as well as five evaluation writing criteria of Magoosh Essay Rubric like the quality of ideas, organization, writing style, grammar and usage, as well as summary were taught, and scores of each evaluation of criterion were analyzed by the teacher. In this way, the writing performance of the participants based on evaluation criteria for achieving each of the scores from zero to nine in IELTS task 1 writing and zero to six in Magoosh Essay Rubric were described. At the end of each session, participants in all groups were asked to write some essays of the IELTS, GRE, and GMAT writing samples, and assess them based on writing rubrics criteria. Two groups (one male and one female) performed the task individually, and other groups (one male and one female) did it collaboratively.

At the last stage, the ESLP questionnaire was administered to students again in order to examine their views on their abilities in writing. This stage aimed to find the probable change in the students' ideas about their capabilities.

3.4. Data Analysis

Data collected from the ESLP questionnaire was fed into the SPSS and analyzed quantitatively. That is, descriptive statistical procedures were utilized to find out the status of learners' views on their abilities on writing skills before and after the treatment sessions. Then, two-Way Analysis of Variance (Two-Way ANOVA) which is an inferential test, was conducted to detect and compare two different assessment training (individual and collaborative) on EFL learners' writing skill across gender.

4. Results

4.1. Investigating First Research Question

The first research question aimed to explore whether individual and collaborative assessment has any significant effect on EFL learners' writing skills.

Table 1

Tests of Between-Subjects Effects

Dependent Variable: Post-Test

		Type III			Mean		
Source	Sum of Squares	df	Square		F	g.	Si
Model	Corrected	2281.27	36	63.369	1.3	.1	
		3 ^a			88	51	
	Intercept	362087.687	1	362087.687	793	.0	
					0.754	00	
	Groups	242.992	1	242.99	5.3	.0	
				2	22	26	

Pre-Test	981.775	20	49.089	1.075	.407
Groups *	1080.616	15	72.041	1.578	.121
Error	1963.214	43	45.656		
Total	571921.000	80			
Corrected	4244.487	79			
Total					

a. R Squared = .537 (Adjusted R Squared = .150)

Table 1 shows the sig value or probability value is .121, safely above the cut-off. The sig value for the interaction is greater than .05, and the interaction is not statistically significant, indicating that the assumption is not violated.

The actual difference in the mean scores between individual and collaborative groups is quite large. As Table 2 indicates, the mean score for the individual group (M=82.12, SD=7.1) was significantly different from that of the collaborative group (M=86.35, SD=7.01):

Table 2

Descriptive Statistics of Individual and Collaborative Writing Skill

Dependent Variable: Post-Test

	Groups	Mean	Std. Deviation	N
Individual and Collaborative	Individual	82.1250	7.10070	40
	Collaborative	86.3500	7.01847	40
	Total	84.2375	7.32992	80

Table 3 shows that there is a significant difference in the writing skill through individual group in comparison with the collaborative group:

Table 3

Tests of Between-Subjects Effects of the Individual and Collaborative Writing Skill

Dependent Variable: Post-Test

	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Model	Corrected	522.989 ^a	2	261.495	5.410	.006	.123
	Intercept	13405.073	1	13405.073	277.359	.000	.783
	Pre-Test	165.977	1	165.977	3.434	.068	.043

Groups	317.069	1	317.069	6.560	.012	.079
Error	3721.498	77	48.331			
Total	571921.000	80				
Corrected	4244.487	79				
Total						

a. R Squared = .123 (Adjusted R Squared = .100)

The sig value is .01, which is greater than .05; therefore, the collaborative group outperformed the individual group on writing skill test scores.

4.2. Investigating Second Research Question

The second research question examined the impact of using two different assessment training (individual and collaborative) on EFL learners' writing skill across gender through conducting Two-Way Analysis of Variance (Two-Way ANOVA).

Table 4 summarizes the results of the post-tests, including the means and standard deviations of the two groups for the female and male EFL learners. As Table 4 indicates, the mean score for the collaborative female group ($M=89.60$, $SD=5.86$) was significantly different from that of the collaborative male group ($M=83.10$, $SD=6.67$):

Table 4

Descriptive Statistics of Individual and Collaborative Groups across Gender

Dependent Variable: Post-Test

Groups		Gender	Mean	Std. Deviation	N
Individual	Female	F	83.4500	6.95455	20
	Male	M	80.8000	7.17158	20
	Total	T	82.1250	7.10070	40
Collaborative	Female	F	89.6000	5.86156	20
	Male	M	83.1000	6.67202	20
	Total	T	86.3500	7.01847	40
Total	Female	F	86.5250	7.07102	40
	Male	M	81.9500	6.93542	40
	Total	T	84.2375	7.32992	80

Table 5 shows the results from the Two-Way ANOVA and whether there is a statistically significant difference between the group means. It is shown that the significance value is 0.00, which

is below 0.05 and, therefore, there is a statistically significant difference in the mean of the writing skill post-test scorers between individual and collaborative groups across gender:

Table 5

Tests of Between-Subjects Effects for the Individual and Collaborative Groups across Gender

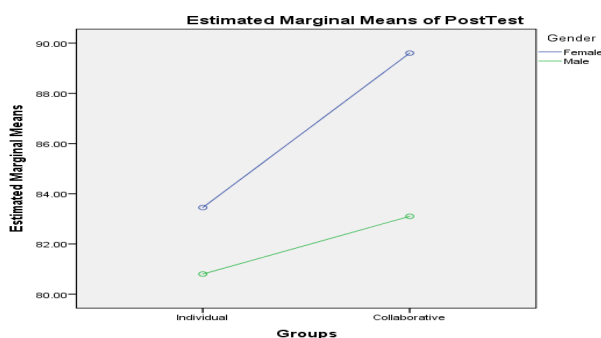
Dependent Variable: Post-Test

	Source	Type III Sum of Squares	df	Mean Square	F	g.	Si Eta Squared	Partial Eta Squared
Model	Corrected	849.737 _a	3	283.246	6.341	.001		.200
	Intercept	567676.513	1	567676.513	12708.864	.000		.994
	Groups	357.013	1	357.013	7.993	.006		.095
	Gender	418.613	1	418.613	9.372	.003		.110
Gender	Groups *	74.113	1	74.113	1.659	.202		.021
	Error	3394.750	76	44.668				
	Total	571921.000	80					
Total	Corrected	4244.487	79					

An interaction between groups and gender could not be demonstrated, $F(1, 76) = 1.65$, $p = 0.2$." Further, the [partial eta squared](#) is only 0.021 for the interaction effect. This can be insignificant. Since there is no interaction effect, we should regard to the significant impacts, both of which have $p = 0.00$. The Partial eta squared is 0.95 for groups and 0.11 for gender. That is, the relative impact of groups and gender is strong.

The mean differences between individual and collaborative groups were significantly different, and it is assumed that they were not at the same level of writing skill. It is shown in the figure below graphically:

Graph 1 *Relationship between Individual and Collaborative Post-Test Scores across Gender*



The preceding graph (Figure 1) visually illustrates the individual and collaborative group scores across gender. Further, it shows that EFL learners in the collaborative group outstripped the ones in the individual. It also shows that EFL learners in the female group outperformed the ones in the male group.

5. Discussion

As discussed earlier, this study tried to investigate whether assessment training had any significant effect on Iranian male and female EFL learners' writing skill in individual and collaborative groups or not.

The results confirmed that passing the treatment sessions about assessment training could help learners write a good writing and soon specify mechanics of writing in essays. Learners learned which components in essays are important. This finding is in line with those trying to find out the items required for writing scoring in that their findings revealed that idea and content, vocabulary, organization, grammar, and the mechanics of writing as the key components in L2 writing assessment (Rafoth & Rubin, 1984, Davidson & Lloyd, 2005, Betsis et al. 2012).

Results of the present study showed that there was a significant difference between individual and collaborative assessment of essays in the post-treatment sessions. It means that, writings of the collaborative groups outperformed the individual groups. These findings are in line with those of Chang, et al. (2012), in which they worked on portfolio assessment and confirmed that mean scores of peer-assessment were high and scores of self-assessment and teacher-assessment were low. The same results were found by Abolfazli and Sadeghi (2013) who concluded that when peer-assessment is compared to self-assessment due to their effects on learners' scores, the peer-assessment group performed better than the self-assessment group. In another study, Meihami and Esfandiari (2020) confirmed the positive impact of collaborative methods in teaching language, especially learners' performance in writing.

In addition, in the collaborative groups, the females outperformed males in the same group. These findings of the current study are consistent with those of Jafari and Ansari (2012), who found female EFL learners who performed writing accuracy collaboratively outperformed males in the same group.

6. Conclusion and implications

The findings of the present study indicated that there is a significant difference between individual and collaborative assessment of essays in the post-treatment sessions. In addition, results confirmed that passing the treatment sessions about self-assessment collaboratively could help learners write a good writing and soon specify mechanics of writing in essays. Therefore, the students were often satisfied with the impact of self-assessment on their writing skills. Moreover, the study found that female learners in the collaborative group outperformed the male group who performed assessment writing collaboratively.

The findings of the present study imply that teachers can incorporate assessment-training strategies into their classes. They can try to teach learners to evaluate their writing by using essay rubrics. Assessment training strategies with applying appropriate writing rubrics can also be helpful for EFL syllabus designers and material developers because they play a significant role in learners' writing performance by paying attention to main writing scores criteria such as coherence and cohesion, task achievement, lexical resource, grammatical range, quality of ideas, organization, writing style, as well as the summary.

This study has some limitations. Firstly, the age of the participants was limited to the range of 18 to 22 years of age. Therefore, the findings of the study would not be applicable to young learners or those in far different age ranges. Secondly, the present study was limited to the EFL learners at the upper-intermediate level. It is likely that the results would differ with learners at the beginning level of language proficiency. Furthermore, regarding the limited time, it was not possible to conduct the study with a large sample to obtain more generalizable outcomes. In this way, further studies should be done with a large number of EFL/ESL learners to decrease the probable limitations in the generalization of the results. More studies are necessary to employ more reliable instruments in screening assessment to find more reliable and valid findings.

References

- Abolfazli, Z. & Sadeghi, K. 2013. The effect of assessment type (self vs. peer) on Iranian university EFL students' course achievement. *Social and Behavioral Sciences*. 70(1), 1552-1564. ISSN: 1877-0428.
- Abrams, Zs. I. (2019). Collaborative writing and text quality in Google Docs. *Language Learning & Technology*, 23(2), 22–42. 10125/44681
- Alhadabi, A., & Karpinski, A. C. (2020). Grit, self-efficacy, achievement orientation goals, and academic performance in University students. *International Journal of Adolescence and Youth*, 25(1), 519–535. <https://doi.org/10.1080/02673843.2019.1679202>
- Alias, M., Masek, A., & Salleh, H. H. M. (2015). Self, peer, and teacher assessments in problem-based learning: Are they in agreements? *Procedia-Social and Behavioral Sciences*, 204, 309-317.
- Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education (Tempe, Ariz.)*, 4(4), 1–28.
- Andrade, H., & Boulay, B. (2003). Role of Rubric-Referenced Self-Assessment in Learning to Write. *The Journal of Educational Research*, 97(1), 21–30. <https://doi.org/10.1080/00220670309596625>
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3–13.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Corwin Press Inc.
- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, 27, 427–441.
- Banlı, S. (2014). *The role of self-assessment practices in the improvement of freshman students' writing performance and awareness* (Unpublished master thesis). Çağ University Institute of Social Sciences, Mersin.
- Bayat, Ö. (2010). İngilizce yazılı anlatım derslerinde uygulanan akran ve öz değerlendirme etkinliklerine yönelik öğrenci görüşleri. *Dil Dergisi*, 150, 70–81.
- Beard, R., & Burrell, A. (2010). Writing attainment in 9- to 11-year-olds: Some differences between girls and boys in two genres. *Language and Education*, 24(6), 495–515.
- Bernat, E., & Lioted, R. (2007). Exploring gender effect on EFL learners' beliefs about language learning. *Australian Journal of Educational & Developmental Psychology*, 7, 79–91.
- Betsis, A., Haughton, L., & Mamas, L. (2012). Succeed in the new Cambridge proficiency (CPE) - student's book with 8 practice tests.
- Bijsterbosch, E., Béneker, T., Kuiper, W., & van der Schee, J. (2019). Teacher Professional growth on assessment literacy: A case study of prevocational geography education in the Netherlands. *Teacher Educator*, 54(4), 420–445. <https://doi.org/10.1080/08878730.2019.1606373>
- Bing, X. (2016). A study of the effects of student self-assessment on the EFL. *Journalism and Mass Communication*, 6(2), 91–107.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). MIT Press.
- Blanche, P., Merino, B. 1989. Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*. (39)3, 313- 338. ISSN: 1467-9922.
- Blue, G. M. (1988). Self-assessment: The limits of learner independence. *Individualization and autonomy in language learning. ELT Documents*, 131, 100–118.

Bradford, K. L., Newland, A. C., Rule, A. C., & Montgomery, S. E. (2016). Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. *Early Childhood Education Journal*, 44(5), 463–472.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2 ed.). NY: Pearson education.

Brown, H. D. (2004). *Language assessment: principles and classroom practices*. Pearson/Longman.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.

Cabrera, S., Rosario, J., Castillo, P., & Jimenez, J. (2017). Implementing rubrics to assess writing skills in an Adults Advanced EFL (English as a Foreign Language) Class at ICDA (Instituto Cultural Dominicano Americano). *International Journal of Innovation and Applied Studies*, 20(2), 681–710.

Chamot, A. U., & O'malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Addison-Wesley Publishing Company.

Chang, C. C. TSENG, K. H., LOU, S. J. 2012. A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students. In: *Computers & Education*. 58(1), 303-320. ISSN: 0360-1315

Cheng. (2017). Assessment in the language classroom. Palgrave. <https://link.springer.com/book/10.1057/978-1-137-46484-2>

Cohen, E., Lotan, R., Scarloss, B., Schultz, S., & Abram, P. (2002). Can groups learn? *Teachers College Record*, 104(6), 1045–1068. <http://www.tcrecord.org>

Cömert, M., & Kutlu, Ö. (2018). The effect of self-assessment on achievement in writing in English. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research*, 8(1), 107-118. <http://ebad-jesr.com/>

Costa, D. (2019). Indian English: A national model. *Journal of English as an International Language*, 14(2), 16-28.

Davidson, P., & Lloyd, D. (2005). Guidelines for developing a reading test. In D. Lloyd, P., Davidson, & C. Coombe (Eds.), *the fundamentals of language assessment: A practical guide for teachers in the Gulf*. TESOL Arabia Publications

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415.

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

Fahimi, Z., & Rahimi, A. (2015). On the impact of self-assessment practice on writing skill. *Procedia: Social and Behavioral Sciences*, 192, 730–736.

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference*, 17.0 update (10a ed.). Pearson.

Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal*, 51(1), 12–20.

Hasani, M. T., & Moghadam, C. R. (2012). The effect of self-assessment on Iranian EFL learners' writing skills. *The Iranian EFL Journal*, 32(6), 371–388.

Henner-Stanchina, C., & Holec, H. (1985). *Evaluation in an autonomous learning scheme*. Longman.

Herbert, N. (2006). Peer assessment: Is it fair? *UniServe Science Assessment Symposium Proceedings*. Retrieved from http://sydney.edu.au/science/uniserve_science/pubs/procs/2006/herbert.pdf

Hsu, H.-C. (2019). Wiki mediated collaboration and its association with L2 writing development: An exploratory study. *Computer Assisted Language Learning*, 32(8), 945–967. <https://doi.org/10.1080/09588221.2018.1542407>

Honsa, S., Jr. (2013). Self-assessment in EFL writing: A study of intermediate EFL students at a Thai university. *Voices in Asia Journal*, 1(1), 34–57.

Huang, S. C. (2012). Like a bell responding to a striker: Instruction contingent on assessment. *English Teaching*, 11(4). <http://edlinked.soe.waikato.ac.nz/research/files/etpc/files/2012v11n4art7.pdf>

Jafari, N., & Ansari, N. A. (2012). The effect of collaboration on Iranian EFL learners' writing accuracy. *International Education Studies*, 5(2), 125–131. <https://doi.org/10.5539/ies.v5n2p125>

Karkehabadi, S. (2013). Enhance Student Performance Why Use a Rubric? North Virginia Community College.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22, 121–132. 10.1016/j.learninstruc.2011.08.04

Kurt, M. (2014). Collaborative assessment: Fostering ownership in assessment. Retrieved from: <https://www.researchgate.net/publication/320596429>

Lam, R. (2010). The role of self-assessment in students' writing portfolios: A classroom investigation. *TESL Reporter*, 43(2), 16–34.

Lejk, M., & Wyvill, M. (1996). A survey of methods of deriving individual grades from group assessments. In: *Assessment & Evaluation in Higher Education*. 21(3), 267-280. ISSN: 0260-2938.

Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the teachers? *Language Testing*, 19(3), 246–276.

Mangelsdorf, K. 1992. Peer reviews in the ESL composition classrooms: What do students think? *ELT Journal*. 46 (3), 274-284. ISSN: 0951-0893

McDonough, K., De Vleeschauwer, J., & Crawford, W. (2018). Comparing the quality of collaborative writing, collaborative prewriting, and individual texts in a Thai EFL context. *System*, 74, 109–120. <https://doi.org/10.1016/j.system.2018.02.010>

Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *tömer Ankara üniversitesi*. Retrieved from <http://dergiler.ankara.edu.tr/dergiler/27/2188/22675.pdf>

Meihami, H., & Esfandiari, R. (2020). Comparative effects of self-assessment, peer-assessment, and teacher assessment on EFL learners' writing performance. *XLinguae*, 13(4), ISSN 1337-8384, ISSN 2453-711X, DOI: 10.18355/XL.2020.13.04.7

Michael, J., & Valdez, L. (1996). *Authentic assessment for English language learners: practical approaches for teachers*. Pearson P T R.

Muhammad, A., Lebar, O., & Mokshein, S. E. (2018). Rubrics as Assessment, Evaluation and Scoring Tools. *International Journal of Academic Research in Business & Social Sciences*, 8(10), 1417–1431

Nicol, D. 2010. From monologue to dialogue: Improving written feedback in mass higher education. In: *Assessment & Evaluation in Higher Education*. 35(5), 501-517. ISSN: 0260-2938.

Noels, K. A., Pelletier, L. G., Vallerand, R. J. 2000. Why are you learning a second language? Motivational orientations and self-determination Theory: *Language Learning* .50(1), 57-85. ISSN:1467-9922.

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing*, 21(1–2), 3–26. <https://doi.org/10.1007/s11145-007-9062-1>

Orikassa, M. (2012). Effectiveness of implementing world Englishes in English language curricula. *Journal of English as an International Language*, 12(2), 73-88.

- Oscarson, A. D. (2009). *Self-assessment of writing in learning English as a foreign language: A study at the upper secondary school level*. Geson Hylte Tryck.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited. *Educational Research Review*, 9, 129–144.
- Patri, M. 2002. The influence of peer feedback on self- and peer-assessment of oral skills. In: *Language Testing*. 19(2), 109-131. ISSN: 0265-5322.
- Perlman, C. C. (2003). *Performance assessment: Designing appropriate performance tasks and scoring rubrics*. North Carolina, USA.
- Rafoth, B. A., & Rubin, D. L. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication*, 1(4), 446–458.
- Read, B., Francis, B., & Robson, J. (2005). Gender, 'bias', assessment, and feedback: Analyzing the written assessment of undergraduate history essays. *Assessment & Evaluation in Higher Education*, 30(3), 241–260.
- Sadler, D. R., & Good, E. 2006. The impact of self- and peer-grading on student learning. In: *Educational Assessment*. 11(1), 1-31. ISSN: 1062-7197.
- Sajedi, R. (2014). Self-assessment and portfolio production of Iranian EFL Learners. *Procedia: Social and Behavioral Sciences*, 98, 1641–1649.
- Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18(3), 303–325.
- Simin, S., & Tavangar, M. (2009). Metadiscourse knowledge and use in Iranian EFL writing. *Asian EFL Journal*, 11(1), 230-255. Retrieved October 16, 2011 from http://www.asian-efl-journal.com/March_2009.pdf
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74–88.
- Thompson, J. (2013). A writing rubric for Chinese University EFL Students. *Canadian Journal of Applied Linguistics*, 4, 183–194.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.
- Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, 26(1), 17–44.
- Yang, C. (2013). How Chinese beginning writers learn English writing: A survey of writing strategies. *Journal of Educational and Social Research*, 3(1), 9–18.
- Zeynali, S. (2012). Exploring the gender effect on EFL learners' learning strategies: *Theory And Practice in Language Studies*. 2(8), 1614–1620.
- Zhang, L. (2012). Rubric assessment in English writing courses. *Proceedings of the Second Northeast Asia International Symposium Language, Literature, and Translation*, 6, 217-222

Appendix A

IELTS™**WRITING TASK 1: Band Descriptors (public version)**

Band	Task achievement	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	<ul style="list-style-type: none"> fully satisfies all the requirements of the task clearly presents a fully developed response 	<ul style="list-style-type: none"> uses cohesion in such a way that it attracts no attention skillfully manages paragraphing 	<ul style="list-style-type: none"> uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips' 	<ul style="list-style-type: none"> uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	<ul style="list-style-type: none"> covers all requirements of the task sufficiently presents, highlights and illustrates key features/ bullet points clearly and appropriately 	<ul style="list-style-type: none"> sequences information and ideas logically manages all aspects of cohesion well uses paragraphing sufficiently and appropriately 	<ul style="list-style-type: none"> uses a wide range of vocabulary fluently and flexibly to convey precise meanings skillfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation produces rare errors in spelling and/or word formation 	<ul style="list-style-type: none"> uses a wide range of structures the majority of sentences are error-free makes only very occasional errors or inappropriacies
7	<ul style="list-style-type: none"> covers the requirements of the task (A) presents a clear overview of main trends, differences or stages (GT) presents a clear purpose, with the tone consistent and appropriate clearly presents and highlights key features/bullet points but could be more fully extended 	<ul style="list-style-type: none"> logically organises information and ideas; there is clear progression throughout uses a range of cohesive devices appropriately although there may be some under-/over-use 	<ul style="list-style-type: none"> uses a sufficient range of vocabulary to allow some flexibility and precision uses less common lexical items with some awareness of style and collocation may produce occasional errors in word choice, spelling and/or word formation 	<ul style="list-style-type: none"> uses a variety of complex structures produces frequent error-free sentences has good control of grammar and punctuation but may make a few errors
6	<ul style="list-style-type: none"> addresses the requirements of the task (A) presents an overview with information appropriately selected (GT) presents a purpose that is generally clear; there may be inconsistencies in tone presents and adequately highlights key features/ bullet points but details may be irrelevant, inappropriate or inaccurate 	<ul style="list-style-type: none"> arranges information and ideas coherently and there is a clear overall progression uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical may not always use referencing clearly or appropriately 	<ul style="list-style-type: none"> uses an adequate range of vocabulary for the task attempts to use less common vocabulary but with some inaccuracy makes some errors in spelling and/or word formation, but they do not impede communication 	<ul style="list-style-type: none"> uses a mix of simple and complex sentence forms makes some errors in grammar and punctuation but they rarely reduce communication
5	<ul style="list-style-type: none"> generally addresses the task; the format may be inappropriate in places (A) recounts detail mechanically with no clear overview; there may be no data to support the description (GT) may present a purpose for the letter that is unclear at times; the tone may be variable and sometimes inappropriate presents, but inadequately covers, key features/ bullet points; there may be a tendency to focus on details 	<ul style="list-style-type: none"> presents information with some organisation but there may be a lack of overall progression makes inadequate, inaccurate or over-use of cohesive devices may be repetitive because of lack of referencing and substitution 	<ul style="list-style-type: none"> uses a limited range of vocabulary, but this is minimally adequate for the task may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader 	<ul style="list-style-type: none"> uses only a limited range of structures attempts complex sentences but these tend to be less accurate than simple sentences may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	<ul style="list-style-type: none"> attempts to address the task but does not cover all key features/bullet points; the format may be inappropriate (GT) fails to clearly explain the purpose of the letter; the tone may be inappropriate may confuse key features/bullet points with detail; parts may be unclear, irrelevant, repetitive or inaccurate 	<ul style="list-style-type: none"> presents information and ideas but these are not arranged coherently and there is no clear progression in the response uses some basic cohesive devices but these may be inaccurate or repetitive 	<ul style="list-style-type: none"> uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task has limited control of word formation and/or spelling errors may cause strain for the reader 	<ul style="list-style-type: none"> uses only a very limited range of structures with only rare use of subordinate clauses some structures are accurate but errors predominate, and punctuation is often faulty
3	<ul style="list-style-type: none"> fails to address the task, which may have been completely misunderstood presents limited ideas which may be largely irrelevant/repetitive 	<ul style="list-style-type: none"> does not organise ideas logically may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas 	<ul style="list-style-type: none"> uses only a very limited range of words and expressions with very limited control of word formation and/or spelling errors may severely distort the message 	<ul style="list-style-type: none"> attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	<ul style="list-style-type: none"> answer is barely related to the task 	<ul style="list-style-type: none"> has very little control of organisational features 	<ul style="list-style-type: none"> uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling 	<ul style="list-style-type: none"> cannot use sentence forms except in memorised phrases
1	<ul style="list-style-type: none"> answer is completely unrelated to the task 	<ul style="list-style-type: none"> fails to communicate any message 	<ul style="list-style-type: none"> can only use a few isolated words 	<ul style="list-style-type: none"> cannot use sentence forms at all
0	<ul style="list-style-type: none"> does not attend does not attempt the task in any way writes a totally memorised response 			

(A) Academic | (GT) General Training

IELTS is jointly owned by the British Council, IDP: IELTS Australia and Cambridge English Language Assessment. Page 1 of 1

<https://www.ielts.org/-/media/pdfs/writing-band-descriptors-task-1.ashx?la=en>

Appendix B

Magoosh Essay Rubric for the GRE & GMAT					
Score	Quality of Ideas	Organization	Writing Style	Grammar & Usage	Summary
6	Insightfully develops a position on a topic or argument with compelling, persuasive examples and reasons	A clear organizational structure with a logical progression, linking ideas to supporting points from start to finish, from paragraph to paragraph and sentence to sentence	A wide variety of sentence structures and lengths, showing a superior control of word choice with a clear, concise style	Shows a superior control of writing, with impeccable grammar, mechanics, and usage—yet may have a few minor, non-repeated errors	A 6 response is a precise, well-articulated analysis of the complexities of the issue or argument, and demonstrates mastery of the elements of effective writing.
5	Develops a position on a topic or argument with well-chosen examples and reasons	A well-organized structure with a progression of ideas, linking ideas to supporting points from start to finish	Variety in sentence structures, showing control of word choice with a clear style	Shows a control of language, with strong grammar, mechanics, and usage—yet may have a few minor errors	A 5 response is a well-developed analysis of the complexities of the issue and demonstrates a strong control of the elements of effective writing.
4	Develops a position on a topic or argument with relevant examples and reasons	Organized with some progression of ideas, linking ideas and supporting points with some consistency	Some, to little, variety in sentence structure, showing adequate control of word choice with an inconsistently clear style	Shows control of standard English but contains some flaws	A 4 response is a competent analysis of the issue and demonstrates adequate control of the elements of writing.
3	Develops a vague or limited position on a topic or argument with few examples or reasons of questionable importance	Poorly organized with unclear links between ideas and supporting points	A lack sentence variety, showing imprecise use of word choice with an inconsistent, wordy style	Shows some lack of control in grammar, mechanics and usage, containing occasional major flaws with more frequent minor flaws	A 3 response has some competence in its analysis of the issue and in its control of the elements of writing but is clearly flawed.
2	Ideas on a topic or argument are unclear or seriously limited	Disorganized and provides few, if any, relevant links between ideas and examples	Serious and frequent problems with word choice and sentence structure, showing a lack of style	Shows a lack of control in grammar, mechanics, and usage, containing numerous, repeated errors	A 2 response has serious weaknesses in analytical writing.
1	Provides little evidence of the ability to develop a position in response to a topic or argument	No organization or logic, containing irrelevant details and examples with little to no distinction between the two	Severe and persistent errors in word choice, language, and sentence structure, showing no real style	Shows a complete lack of control in grammar, mechanics, and usage, making the essay unreadable and incomprehensible	A 1 response has fundamental deficiencies in analytical writing skills.
0	Completely off topic, blank, or not written in English	Entire lack of structure, blank, or not written in English	Blank or not in English	Blank response, not written in English, or no regard for English grammar, mechanics and usage	A 0 response has a complete lack of analytical writing skills.

<https://2aih25gkk2pi65s8wfa8kzvi-wpengine.netdna-ssl.com/wp-content/uploads/2014/04/Magoosh-Essay-Rubric-for-the-GRE-GMAT.pdf>

Acknowledgements

First and foremost, I am thankful to Almighty God for His blessings in all phases of my life. My sincere thanks go to my colleagues at English Language Institute in Karaj who helped me in data collections procedure during the experimental phase of the study. I must express my deepest thanks to dear Dr. Nazari, who has helped me by his invaluable suggestions during the preparation of this study.