

# Portfolio Assessment, Peer Assessment and Writing Skill Improvement

Ahmad Yaghoubi

Department of English Language Teaching, Faculty of Persian Literature and Foreign Languages, Roudehen Islamic Azad University, Roudehen, Iran

Majid Mobin

Roudehen Islamic Azad University, Roudehen, Iran

**Abstract**—Little research has been conducted in the field of alternative assessment including portfolio and peer assessment in Iran. Moreover, in spite of the assumed advantages of portfolio and peer assessment, the practicality of these techniques needs to be investigated. To this aim, a total of 207 writing papers from 69 participants were collected to investigate the effects of peer and portfolio assessment on the writing components in two experimental groups. The data were collected from students of a general English semester course held at Sharif University of Technology. There were two experimental groups of 37 and 32 students experiencing portfolio and peer assessment respectively. As the treatment, students were supposed to write compositions, and for each composition, they were supposed to invest at least twenty minutes on a topic of interest e.g. sport, science, social issues, etc. And they were then rated by their teacher in group A and their peers in group B based on ten pre-established assessment criteria in each session during the term. After administering the writing post-test, independent t-tests were run to find the differences between the two groups. As the result, it was indicated that portfolio and peer assessment had the same effect on four of the five components of writing namely as content, organization, language use and vocabulary. But only there was a significant difference between portfolio and peer assessment groups on the last component of the study: mechanics in writing. The findings of this study have implications for language teachers, material designers, and educational policy makers.

**Index Terms**—alternative assessments, peer assessment, portfolio assessment

## I. INTRODUCTION

Dochy, Segers, and Sluijsmans (1999) emphasized the idea of using assessment as a learning device in a way that one can switch from a teacher-centered classroom to a learner-centered classroom. In this context, teachers not only monitor learning but also improve it. Orsmond and Merry (1996) also stated that in order to encourage students to be more self-dependent in their own development; we need to change the current tests. To this end, we need to empower students, two instances of which are portfolio and peer assessment. While these assessments help to create a closer relationship among teachers and students, they also foster critical thinking, communication, feedback, responsibility, autonomy, and help students develop useful skills in academic and professional areas (Peng, 2010).

Despite the multiple advantages of new forms of assessments, students often assume that assessments made by teachers are more accurate, which may not always be true. In fact, teacher assessment cannot always be viewed as the most valid method and it can be triangulated with varied sources of assessments (Orsmond and Merry, 1996). On the other hand, there are still controversies about the accuracy of portfolio and peer assessments (Matsuno, 2009; Patri, 2002; Ross, 2006). With regard to the aforementioned merits of portfolio and peer assessment, the pedagogical value of these techniques is hard to be denied; however, the effectiveness of portfolio and peer assessments needs to be empirically tested in an EFL context. In addition, most of the university students in Iran have not experienced portfolio and peer assessment. Thus, this study aimed to investigate the effectiveness of portfolio and peer assessments on writing skill. To this aim, the 5 writing components as content, organization, language use, vocabulary and mechanics were compared between the two groups of students experiencing portfolio and peer assessments.

In recent years, the field of language teaching experienced more attention on learner independence and autonomy (Butler & Lee, 2010). According to the literature, portfolio and peer assessment seem to be a viable option to enhance independent and autonomous learning. Portfolio and peer assessments have many empirically supported merits that include raising students' awareness, extension of assessment range, improvement in learning, promotion of goal-orientation, beneficial post-course effects, and mutual assessment responsibility (Oscarson, 1989). Brown and Hudson (2002) reported the other merits to these alternative assessments as learner involvement in the learning and assessment, demanding short time to conduct in classrooms, increasing motivation, and encouraging learner- autonomy self-awareness and noticing the gap between self- and others' perception. While Saito & Fujita, (2004) added multiple-perspective feedbacks, sensitizing students to the evaluation criteria, encouraging connectivity in the classroom, and

self- reflection to their benefits. Although previous studies have addressed portfolio and peer assessment exclusively, almost a few examined both portfolio and peer assessment on writing skills in an EFL context.

Gaining insight into the degree of effectiveness of portfolio and peer assessment on writing performance can help policy makers, teachers and curriculum designers in gaining trust in these techniques. This study has focused on the effectiveness of these two newly-applied techniques on students writing skill components such as content, organization, language use, vocabulary, and mechanics. Writing skill plays a prominent role in recent types of proficiency tests (e.g. TOEFL IBT). In addition, writing has also formed backwash effect for motivating English learners to improve their oral skills (Peng, 2010). Thus, this study aims to broaden the knowledge of implementing portfolio and peer assessment in teaching writing in EFL context by exploring the effectiveness of either portfolio or peer assessments on students' writing skills.

### *Research Questions and Hypotheses*

Since this study aimed to investigate the relationship between the two methods of assessment on students writing skills, the research questions are as following:

- Q1. Do portfolio and peer assessment have the same effect on language use in writing among Iranian EFL learners?
- Q2. Do portfolio and peer assessment have the same effect on organization in writing among Iranian EFL learners?
- Q3. Do portfolio and peer assessment have the same effect on content in writing among Iranian EFL learners?
- Q4. Do portfolio and peer assessment have the same effect on vocabulary in writing among Iranian EFL learners?
- Q5. Do portfolio and peer assessment have the same effect on mechanics in writing among Iranian EFL learners?

According to the nature of the research questions, there are five hypotheses for them. The hypotheses are:

H1. Portfolio and peer assessments do not have the same effect on language use in writing among Iranian EFL learners.

H2. Portfolio and peer assessment do not have the same effect on organization in writing among Iranian EFL learners.

H3. Portfolio and peer assessment do not have the same effect on content in writing among Iranian EFL learners.

H4. Portfolio and peer assessment do not have the same effect on vocabulary in writing among Iranian EFL learners.

H5. Portfolio and peer assessment do not have the same effect on mechanics in writing among Iranian EFL learners.

## II. METHODOLOGY

### *A. Participants*

The study was conducted at Sharif University of Technology. The sampling was a convenience one. Convenience or opportunity sampling is part of non-probability sampling in which we get help from subjects that are available (Dörnyei, 2007). The students were from 2 different classes which were structured with the same teacher and a textbook called 'Active Four'. There were more than 30 students in each class and totally 69 students were engaged in the study. One class experienced portfolio assessment which was named group A and group B experienced peer assessment. Group A (portfolio assessment group) included 37 students and group B (peer assessment group) included 32 learners. The students' proficiency level was deemed as intermediate, however, a PET test was run to test the homogeneity of the two groups and the analysis of the results showed that the two groups were homogeneous. They attended the class two sessions a week, while each session lasted one and a half hours, and the treatment period extended for 16 sessions in a normal semester. They were in the age range of 18 to 22 coming from different colleges such as the colleges of psychology, or Sciences such as Chemistry, Physics, Mathematics, etc. The proportion of male and female was 48/21 in favor of male students. All the students were willing to participate in these classes, for the course was extracurricular and they have been charged in order to attend these classes. Table I shows the descriptive statistics of the participants in the study.

TABLE I.  
PARTICIPANT STATISTICS

	Number	Gender		Age Range	Treatment
		Male	Female		
Group A	37	26	11	18-20	Portfolio assessment
Group B	32	22	10	18-20	Peer assessment
Total	69	48	21		

### *B. Instruments*

The instruments used in this study include: 1) A framework for portfolio assessment including the goal, collection, evaluation, and reflection on the writings of students, 2) The standard PET test used as the tool for testing the homogeneity of the groups and as the post-test, 3) A 5- item pre-established criteria assessment sheet used as a scoring measure for peer assessment, measuring content, language use, organization, vocabulary, and mechanics of the students' writing papers.

The writing papers were scored based on the Jacobs' scoring framework for the end term evaluation; moreover, the scoring of the post test for both groups was conducted based on the same criteria. This scoring framework is divided

into five parts namely as content, organization, vocabulary, language use, and mechanics (See Appendix A). Each part is based on the 4-point Likert scale in which 1 represents the lowest score and 4 represents the highest score.

### C. Data Collection Procedure

The data collection procedure was conducted with 2 different intact classes that were grouped in a) portfolio assessment group and, b) peer assessment group. All the groups took part in a pre-test for the sake of ensuring their homogeneity and writing part of the standard PET test was administered as pretest. During the treatment period, group A experienced portfolio assessment while group B solely experienced peer assessment.

After the introduction of the Portfolio evaluation scheme to the English classes, the teachers proceeded with their usual classroom activities that they had planned for the whole term. After all writing activities were accomplished and completed, the teacher asked their students to collate and compile all their best written output to be marked and analyzed for portfolio assessment. The same writing papers were subjected to peer assessment on the part of the students in the second group. The teacher provided writing prompts to facilitate the writing of the two evaluation essays.

At the beginning of the semester, the teacher explained the objectives and procedures of portfolio assessment to group A, and the purposes and procedures for peer assessment to group B. In the next session, he discussed issues and concerns of the students regarding the use of portfolio and peer assessment. For instance, Peng (2010) stated that students might be concerned that assigning grades to friends would jeopardize their friendship. Therefore, this issue was dealt with and the students were assured that the peer assessment would not have been used as the base for their final scores. Thus, he assured the students that their names would be kept confidential. In the third session, the instructor had to discuss various components of a good writing. From session four to the end, in each session, students were supposed to write one short essay on a pre-defined topic. Totally, each student was required to fulfill at least three writing assignments for the term. In group A, the collected writings of the participants were included in their own profile. Students were required to submit the final portfolio project to the teacher. Some students were very creative in their projects, and they came up with unique presentations that made their projects very interesting. And in group B, the peer assessment evaluation sheets for each student were collected for further analysis right after each writing task presented. Finally, the post-test was administered to both groups which was writing a composition, and the scores were obtained based on the Jacob's scoring framework to be implemented in the data analyses of the study.

## III. DATA ANALYSIS

Prior to the conduction of statistical procedures for the research questions, the reliability of PET was estimated (table II). In addition, the inter-rater reliability of three raters' evaluation of the participants' writings was computed using Cohen's Kappa test. The resulting Kappa of .85 indicates that raters provided similar information about students' writing performance.

TABLE II.  
RELIABILITY RESULTS OF PET

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.708	.704	6

TABLE III.  
INTER-RATER RELIABILITY RESULTS

	value	Std. Error	Approx. T	Approx. Sig.
Measure of Agreement	.858	0.54	11.163	.000
N of Valid Cases	69			

The major research question of this study was concerned with the effects of peer and portfolio types of assessment on the writing performance of language learners. In order to analyze the data to test the research questions, the statistical procedures have been carried out using statistical package for the social sciences (SPSS, 2012) version 21. First, the scores were analyzed to ensure the assumptions of normality. The results of Kolmogorov-Smirnov tests and the box plots are presented below.

TABLE IV.  
ONE-SAMPLE KOLMOGOROV-SMIRNOV TEST RESULTS

Null hypothesis	test	Sig.	Decision
1 The categories of language use occur with equal probabilities.	One-sample Chi-Square Test	.18	The null hypothesis was retained.
2 The categories of organization occur with equal probabilities.	One-sample Chi-Square Test	.20	The null hypothesis was retained.
3 The categories of content occur with equal probabilities.	One-sample Chi-Square Test	.45	The null hypothesis was retained.
4 The categories of vocabulary occur with equal probabilities.	One-sample Chi-Square Test	.06	The null hypothesis was retained.
5 The categories of mechanics occur with equal probabilities.	One-sample Chi-Square Test	.24	The null hypothesis was retained.

The results of Kolmogorov-Smirnov tests show that the scores are normally distributed ( $p > 0.05$ ). In addition to Kolmogorov-Smirnov, box plots are depicted to ensure the normality of the scores.

### A. Findings of 1st Research Question

Having ascertained the assumptions of independent samples t-test as a parametric test (i.e., the normality of data), the next step was to conduct the t-tests. First, the results of independent samples t-test for the first research question are presented in tables V and VI.

TABLE V.  
DESCRIPTIVE STATISTICS OF LANGUAGE USE ACROSS GROUPS

	Group	N	Mean	Std. Deviation	Std. Error Mean
Language use	Peer assessment	32	3.1716	.59538	.10525
	Portfolio assessment	37	3.2433	.57974	.09531

As the mean and standard deviation scores in table 4.4 show, there are very nuance differences between the peer assessment ( $M = 3.17$ ,  $SD = 0.59$ ) and portfolio assessment ( $M = 3.24$ ,  $SD = 0.57$ ) group learners' performance in language use aspect of writing. However, in order to get more accurate and reliable results, an independent samples t-test was run, the results of which are displayed in table VI.

TABLE VI.  
T-TEST RESULTS OF GROUP DIFFERENCES IN LANGUAGE USE

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Dif.	Std. Error Dif.	95% Confidence Int. of the Dif.	
								Lo		up
Language use	Equal variances assumed	.34	.55	-.50	67	.61	-.07	.14	-.35	.21
	Equal variances not assumed			-.50	65.03	.61	-.07	.14	-.35	.21

The results show that the significance level of Levene's test is  $p = 0.55$ , which means that the variances for the two groups (portfolio and peer) are the same. The results of independent samples t-test show statistically insignificant difference ( $t(67) = -0.50$ ,  $p > 0.05$ ) between the portfolio and peer assessment groups in the language use aspect of writing. Therefore, the first hypothesis is rejected. This hypothesis reiterated the nonexistence of the sameness of the effect of portfolio and peer assessments on language use in writing among Iranian EFL learners.

### B. Findings of 2nd Research Question

In order to examine the second research question which is concerned with the differences between portfolio and peer assessment groups' performance in the organization aspect of writing, an independent samples t-test was run. First, the results of descriptive statistics are shown.

TABLE VII.  
DESCRIPTIVE STATISTICS OF ORGANIZATION ACROSS GROUPS

	Group	N	Mean	Std. Deviation	Std. Error Mean
Organization	Peer assessment	32	3.1213	.58417	.10327
	Portfolio assessment	37	3.1470	.57101	.09387

As table VII shows, there are not mean differences between the peer ( $M = 3.12$ ,  $SD = 0.58$ ) and portfolio ( $M = 3.14$ ,  $SD = 0.57$ ) group participants' performance in organization. The results of t-test are indicated in table 4.5.

TABLE VII.  
T-TEST RESULTS OF GROUP DIFFERENCES IN ORGANIZATION

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Organization	Equal variances assumed	.27	.60	-.18	67	.85	-.02	.13	-.30	.25
	Equal variances not assumed			-.18	65.11	.85	-.02	.13	-.30	.2593

The results show that the significance level of Levene's test is  $p = 0.60$ , which means that the variances for the two groups (peer and portfolio) are the same. The results of independent samples t-test show statistically insignificant differences ( $t(67) = -0.18$ ,  $p > 0.05$ ) between the experimental participants in the organization scores. Therefore, the researcher could reject the second hypothesis. This hypothesis stated the nonexistence of the sameness of the effect of portfolio and peer assessments on organization in writing among Iranian EFL learners.

### C. Findings of 3rd Research Question

In an attempt to identify the differences between the peer assessment and portfolio assessment groups in terms of their performance in content aspect of writing, an independent samples t-test was run. The results are shown in tables VIII and IX.

TABLE VIII.  
DESCRIPTIVE STATISTICS OF CONTENT ACROSS GROUPS

Group	N	Mean	Std. Deviation	Std. Error Mean
Content	Peer assessment	32	3.1831	.50231
	Portfolio assessment	37	3.1376	.53860

The man scores do not show differences between the peer assessment ( $M = 3.18$ ,  $SD = 0.50$ ) and portfolio assessment ( $M = 3.13$ ,  $SD = 0.53$ ) in the participants' content scores. Results of independent samples t-test are demonstrated in table IX.

TABLE IX.  
T-TEST RESULTS OF GROUP DIFFERENCES IN CONTENT

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Content	Equal variances assumed	.00	.92	.36	67	.71	.04	.12	-.20	.29
	Equal variances not assumed			.36	66.59	.71	.04	.12	-.20	.29

The results show that the significance level of Levene's test is  $p = 0.92$ , which means that the variances for the two groups (peer and portfolio) are the same. The results of independent samples t-test show statistically insignificant differences ( $t(67) = 0.36$ ,  $p > 0.05$ ) between the peer and portfolio types of assessment of learners' content scores in writing. Consequently, the third hypothesis is rejected. This hypothesis reiterated the nonexistence of the sameness of the effect of portfolio and peer assessments on content in writing among Iranian EFL learners.

#### D. Findings of 4th Research Question

In order to examine the differences between the peer assessment and portfolio assessment groups in terms of their performance in vocabulary part of writing, an independent samples t-test was run. The results are shown in tables X and XI.

TABLE X.  
DESCRIPTIVE STATISTICS OF VOCABULARY ACROSS GROUPS

Group	N	Mean	Std. Deviation	Std. Error Mean
Vocabulary	Peer assessment	32	3.5100	.46239
	Portfolio assessment	37	3.5354	.36830

TABLE XI.  
T-TEST RESULTS OF GROUP DIFFERENCES IN VOCABULARY

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Vocabulary	Equal variances assumed	4.83	.03	-.25	67	.80	-.02	.10	-.22	.17
	Equal variances not assumed			-.25	59.04	.80	-.02	.10	-.22	.17

The results of the above table indicate that the significance level of Levene's test is  $p = 0.03$ , which means that the variances for the two groups (peer and portfolio) are not the same. The results of independent samples t-test show statistically insignificant differences ( $t(59.04) = -2.5$ ,  $p > 0.05$ ) between the peer and portfolio types of assessment of learners' vocabulary scores in writing. Therefore, the fourth hypothesis is rejected. This hypothesis reiterated the nonexistence of the sameness of the effect of portfolio and peer assessments on vocabulary in writing among Iranian EFL learners.

#### E. Findings of 5th Research Question

In order to provide an answer to the last research question about the differences between the peer assessment and portfolio assessment groups in terms of their performance in mechanics of writing, an independent samples t-test was run. The results are shown in tables XII and XIII.

TABLE XII.  
DESCRIPTIVE STATISTICS OF MECHANICS ACROSS GROUPS

	Group	N	Mean	Std. Deviation	Std. Error Mean
Mechanics	Peer assessment	32	3.6463	.30080	.05317
	Portfolio assessment	37	3.4832	.26948	.04430

TABLE XIII.  
T-TEST RESULTS OF GROUP DIFFERENCES IN MECHANICS

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mechanics	Equal variances assumed	.17	.68	2.37	67	.02	.16	.06	.02	.30
	Equal variances not assumed			2.35	62.88	.02	.16	.06	.02	.30

The results demonstrate that the significance level of Levene's test is  $p = 0.68$ , which means that the variances for the two groups (peer and portfolio) are the same. The results of independent samples t-test show statistically significant differences ( $t(67) = 2.37$ ,  $p < 0.05$ ) between the peer and portfolio types of assessment of learners' mechanics use in writing. The results confirm those of descriptive statistics, exhibiting that the peer assessment group ( $M = 3.64$ ,  $SD = 0.30$ ) obtained a higher mean score compared to the portfolio assessment group ( $M = 3.48$ ,  $SD = 0.26$ ). Therefore, the fifth hypothesis is not rejected. This hypothesis stated the nonexistence of the sameness of the effect of portfolio and peer assessments on mechanics of writing among Iranian EFL learners.

#### IV. DISCUSSION

The present study was conducted to examine the effectiveness of two methods of writing skill assessment, namely peer and portfolio assessment. In order to have a clearer and detailed insight about the efficiency of each method, five aspects of writing were examined: language use, organization, content, vocabulary and mechanics. The results of statistical analysis of the learners' writing pieces showed that although the treatments were not different in the first four categories of language use, organization, content, vocabulary and mechanics, there was significant difference for the mechanics of writing, with the peer group outperforming the portfolio group learners. The summary of the findings of the statistical analyses of the present study is reported in table XVI below.

TABLE XIV.  
SUMMARY OF MAIN FINDINGS

hypothesis	Test	Sig.	Decision
1 Portfolio and peer assessment do not have the same effect on language use.	Independent samples t-test	.61	The hypothesis was rejected.
2 Portfolio and peer assessment do not have the same effect on organization.	Independent samples t	.85	The hypothesis was rejected.
3 Portfolio and peer assessment do not have the same effect on content.	Independent samples t	.71	The hypothesis was rejected.
4 Portfolio and peer assessment do not have the same effect on vocabulary.	Independent samples t	.80	The hypothesis was rejected.
5 Portfolio and peer assessment do not have the same effect on mechanics.	Independent samples t	.02	The hypothesis was retained.

#### V. CONCLUSION

It is claimed that portfolio and peer assessment is not as accurate as teacher assessment. But it must be noted that reaching high reliability and validity is not the primary objective of portfolio and peer assessment. As mentioned by Devenney (1989), the goals and functions of portfolio, peer, and teacher assessments are different; teacher assessment is a summative assessment technique which is mostly used for evaluation at the end of the courses while portfolio and peer assessment are formative assessment techniques which aimed at ongoing learning processes. One of the main advantages of having several input samples of students is to help teachers understand learning processes and their outcomes. In other words, portfolio and peer assessment are accompanying tools for students' engagement and empowerment which should be used along with teacher assessment. Shohamy (2001) also stipulated that portfolio and peer assessment should be used as tools for gathering samples of language from learners. All in all, with careful training,

monitoring and utilization, portfolio and peer assessment can be beneficial as good as teacher assessment in “cognitive, social, affective, transferable skill and systemic domains” (Topping, 1998, p. 269).

Since there were not any doubts regarding the benefits of alternative assessments over traditional testing, there was no need to have a control group in the study. Therefore, the comparison of traditional and alternative assessments was not the aim, and the design of the study was set to have two experimental groups to have portfolio assessment verses peer assessment on the improvement of writing skill.

According to the results of the study, it can be concluded that there were not any significant differences between the two methods of assessment namely as portfolio and peer. The only significant difference was in mechanics of writing between the two groups. Since the two groups acted similarly for four major components of writing, it can be concluded that portfolio and peer assessments have the same effect on the writing skill of Iranian English learners.

#### APPENDIX A. JACOBS ET AL.'S ESL COMPOSITION PROFILE

Students	Date	Topic	
	Score level Criteria		Comments
<b>Content</b>	4	Excellent to very good: knowledgeable. Substantive. Thorough development of thesis. Relevant to assigned topic	
	3	Good to average: some knowledge of subject. Adequate range. Limited development of thesis. Mostly relevant to topic, but lacks detail	
	2	Fair to poor: limited some knowledge of subject. Little substance. Inadequate development of topic.	
	1	Very poor: does not show knowledge of subject. Non- substantive. Not pertinent. Or not enough to evaluate	
<b>Organization</b>	4	Excellent to very good: fluent expression. Ideas clearly stated/ supported. Succinct. Well- organized. Logical sequencing. cohesive	
	3	Good to average: somewhat choppy. Loosely organized but main ideas stand out. Limited support. Logical but incomplete sequencing.	
	2	Fair to poor: on fluent. Ideas confused or disconnected. Lacks logical sequencing and development	
	1	Very poor: does not communicate. No organization. Or not enough to evaluate	
<b>Vocabulary</b>	4	Excellent to very good: sophisticated range. Effective word/ idiom choice and usage. Word form mastery. Appropriate register	
	3	Good to average: adequate range. Occasional errors of word/idiom form, choice, usage but meaning not obscured	
	2	Fair to poor: limited range. Frequent errors of word/idiom form, choice, usage. Meaning confused or obscured	
	1	Very poor: essentially translation. Little knowledge of English vocabulary, idioms, word form. Or not enough to evaluate.	
<b>Language use</b>	4	Excellent to very good: effective complex constructions. Few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions	
	3	Good to average: effective but simple constructions. Minor problems in complex constructions. Several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning never obscured.	
	2	Fair to poor: major problems in simple/complex constructions. Frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run- ons, deletions. Meaning confused or obscured	
	1	Very poor: virtually no mastery of sentence construction rules. Dominated by errors. Does not communicate. Or not enough to evaluate.	
<b>Mechanics</b>	4	Excellent to very good: demonstrates mastery of conventions. Few errors of spelling, punctuation, capitalization, paragraphing.	
	3	Good to average: occasional errors of spelling, punctuation, capitalization, paragraphing, but meaning not obscured	
	2	Fair to poor: frequent errors of spelling, punctuation, capitalization, paragraphing. Poor handwriting. Meaning confused or obscured	
	1	Very poor: no mastery of conventions. Dominated by errors of spelling, punctuation, capitalization, paragraphing. Handwriting illegible. Or not enough to evaluate.	

#### REFERENCES

- [1] Brown, J. D., & Hudson, T. (2002). Criterion-referenced language testing. Cambridge: Cambridge University Press.
- [2] Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31.
- [3] Devenney, R. (1989). How ESL teachers and peers evaluate and response to students writing. *RELC Journal*, 20(1), 77-90.

- [4] Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co- assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350.
- [5] Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford: Oxford University Press.
- [6] Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.
- [7] Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education*, 21(3), 239-250.
- [8] Oscarson, M. (1989). Self-assessment of language proficiency: Rational and applications. *Language Testing*, 6(1), 1-13.
- [9] Patri, M. (2002). The influence of peer feedback on self-and peer assessment of oral skills. *Language Testing*, 19(2), 109-131.
- [10] Peng, J. C. (2010). Peer Assessment in an EFL Context: Attitudes and Correlations. In T. P. Matthew, Y. Watanabe, and S. K. Lee. (Eds.), *Selected proceedings of the 2008 second language research forum* (pp. 89-107). Somerville, MA: Cascadilla Proceedings Project.
- [11] Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research and Evaluation*, 11(10), 1-13.
- [12] Satio, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classroom. *Language Teaching Research*, 8(1), 31-54.
- [13] Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- [14] Topping, K. (1998). Peer Assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.

**Ahmad Yaghoubi**, assistant professor in TEFL, is the vice dean of Faculty of Persian Literature and Foreign Languages at Islamic Azad University, Roudehen Branch, Iran. His areas of interest include issues in foreign/second language teaching and learning.

**Majid Mobin**, received his MA in TEFL from Islamic Azad University, Roudehen Branch, Iran. His areas of interest include issues in foreign/second language teaching and learning.