Comparative effects of self-assessment, peer-assessment, and teacher assessment on EFL learners' writing performance

Hussein Meihami – Rajab Esfandiari

Abstract
Self-assessment and peer-assessment, as two alternative assessment procedures, have appealed to researchers in recent years and motivated L2 researchers to examine these two techniques. However, most of the studies have used them for summative purposes, and the formative dimension these two methods can have for learning has been neglected. This study was an attempt to find how they contributed to learning gains. To that end, sixty Iranian male and female intermediate language learners at a language institute were randomly assigned into three treatment conditions: Self-assessment and peer-assessment as experimental groups and teacher assessment as a control group. A language proficiency test was used to homogenize language learners, and a posttest was administered to measure the amount of gain language learners achieved after treatment sessions. We analyzed the test data using descriptive and inferential statistics as implemented in SPSS, a general-purpose computer program for data analysis. Results from a one-way analysis of variance showed statistically significant differences between the score means of three treatment groups. Post-hoc analyses revealed that language learners in the peer-assessment group outperformed those in the other two groups. The findings suggest that peer-assessment as a cooperative technique can be used in language classes to help students improve their writing abilities.
Key words: EFL learners, peer-assessment, self-assessment, teacher-assessment

Introduction

In educational systems, assessment is an inevitable ingredient because it may influence learning, and, when made authentic, it provides feedback and revision to improve learning. Furthermore, through meaningful engagement of students in the learning process, assessment can be motivating (Alderson, Banerjee, 2001). Assessment may also enhance instruction by helping teachers recognize students' weaknesses and strengths (Baniabdelrahman, 2010). It is also stated that assessments one can be considered as valid, fair, ethical, and efficient tools measure different attributes (Mousavi, 2012).

Although traditional tests may be impressive to measure some of the skills such as reading, they are not adequate to assess the productive skills of writing and speaking. As Huerta-Macías (1995) pointed out, "the nature of proficiency-oriented language learning asks for a variety of assessment options reflecting the numerous instructional strategies used in the classroom" (p. 8). One important characteristic of new assessment techniques is the involvement of students in the evaluation of their own learning process and product. There are three main methods of assessing the students' writing performance, namely, self-assessment, peer-assessment, and teacher assessment (Esfandiari, Myford, 2013).

Sengupta (1998) asserted that many researchers had studied peer and self-assessments in L1 and L2 writing (e.g. Mangelsdorf, 1992; Mendonea, Johnson, 1994). Consequently, the findings suggest that students used parts of their friends' comments about revising and drafting. Mangelsdorf (1992) reported that peer-assessment is always rated negatively by Asian students and raised the query about the effects of teacher-centered instruction on peer comments.

In recent years, the process approach to writing has been used to help learners become effective communicators. On the other hand, when students are to assess their own writing, developing certain negative effective and psychological filters towards the teaching/learning and testing process will be minimized (Williams, 2003). This is to enable students to produce richer and more developed pieces of writing. How do students think about writing and learning? How do they see themselves as participants in the writing process? How do they provide themselves and their peers with efficient feedback? And how do these perceptions develop or change over the course of the semester? Still more, can students' assessment establish stronger connection between process writing and the assessment of students' ability in this area? (Williams, 2003).

Many studies have considered self-assessment, peer-assessment, and teacher-assessment in EFL students' performance on the language four skills. Among them, very few studies have considered self, peer, and teacher assessment in EFL students' writing performance in Iran. The main feature of this study that distinguishes it from other studies is that this study is a simultaneous investigation through the impact of self, peer, and teacher assessment on the writing performance of the EFL learners at Iranian English language institutes where their main purpose is teaching English systematically and fostering EFL learners and due to its basic rules learners are studying in homogenous classes. This study can be useful for instructors to make the atmosphere of the class friendlier by emphasizing the ease of language learning through cooperative learning. This study looks into whether self, peer, and teacher-assessment have any effect on the improvement of Iranian intermediate EFL learners' writing performance. Additionally, in this study, we are interested in knowing which method of assessment (self, peer, and teacher-assessment) may have a better effect on the improvement of Iranian intermediate EFL learners' writing performance. Thus, this study was an attempt to address the following research question:

What are the comparative effects of self-assessment, peer-assessment, and teacher assessment on Iranian intermediate EFL learners' writing performance?

Literature Review
Teacher assessment

Since assessment is important in the teaching and learning process, teachers assess their students' learning regularly. Most language teachers use traditional tests. According to Rotham (1995), 90 percent of instructors use teacher-made tests at least once a month. In traditional tests, EFL/ESL teachers provide instructions for the students to complete the activities; their work is evaluated by teachers. The decisions concerning the quality of students' responses are also teacher-oriented. When students respond to test items, they meet the standards of teachers. The format of the traditional tests makes students dependent on their teachers for the entire activity. Test results are for teachers as well. They may only show whether a student is a good test taker.

The teacher assessment, or traditional assessment, has some disadvantages. Some of the researchers have pointed out these disadvantages. Buck (1994), for example, enumerated some of the shortcomings of this kind of assessment. He believes this kind of assessment has no psychological or cognitive justification, neither at the level of the item performance nor at the level of the overall test meaning. He also believes these kinds of assessments make particular and often simplistic opportunities about the psychology of the items. These, and other similar criticism, motivated researchers to look for more novel ways to assess students' abilities, as we discuss in the following subsections.

Alternative assessment

Due to the shortcomings of the traditional assessment, alternative assessment emerged. Hamayan (1995) believes the alternative assessment is a kind of

assessment used within the context of instruction. It can be incorporated into different tasks in the classroom. Smith (1999) pointed out that alternative assessment might take place outside the classroom and by different techniques of assessments. In a more recent definition, Alderson and Baneriee (2001) claimed that alternative assessment is less formal than the traditional assessment. They are usually formative rather than summative in function, are often low stakes in terms of consequences, and are claimed to have beneficial washback effects.

An alternative assessment has some unique features. For example, alternative formats for assessments are certainly the crucial components of performance-based assessment. Cognitive psychology studies support different aspects of alternative assessment. However, it is not applicable and reliable to produce tasks without understanding whether they can be credibly scored (Linn, et. al., 1991).

Brown and Hudson (1998) summarized some of the most commonly used types of alternative assessment as follows: Conferences, debates, demonstrations, diaries/journals, dramatizations, exhibitions, games, observations, peer-assessment, portfolios, projects, self-assessments, and story retelling. Among the cited alternative types of assessments in Brown and Hudson's study, a couple of these types of assessments, namely self-assessment and Peer-assessment, are considered in this study. The main purpose of self-assessment and peer-assessment is to make students responsible for their learning. To do so, self-assessment and peer-assessment provide contexts for obtaining new experiences. In the following paragraphs, we explain these two procedures.

Self-assessment

Over the past 20 years, self-assessment has been increasingly used in a range of education settings. According to Blanche and Merino (1989), the first reports on self-assessment were published in 1976. Since then, self-assessment has continued to expand as a distinct field of study in L2 learning and education. Self-assessment and other self-reflective activities usually require learners to grade their own performance based on clearly defined task requirements and assessment criteria that are sometimes derived from the learners' input. The types of self-assessment described in the literature are varied. They range from standardized questionnaires (e.g., Bachman, Palmer, 1989) to open-ended, informal activities, e.g., reflective diaries and portfolios.

In the field of education and second language acquisition (SLA), the construct of self-assessment is usually understood as an alternative means to assess learners' abilities. As a result, self-assessment research is primarily quantitative in nature and explores the validity and reliability of student self-ratings (e.g., Bachman, Palmer, 1989) rather than the learning process involved in self-appraisal activities. Variables affecting self-assessment accuracy are diverse, ranging from the domain of assessment to item construction and the learners' individual attributes such as personality traits, affectivity, and level of proficiency (Bachman, Palmer, 1989). These findings suggest that self-assessment procedures are too subjective to be used for summative assessment purposes, particularly in environments where the stakes are high.

However, the construct of self-assessment in the SLA literature is currently shifting away from issues strictly concerning assessment theory to embrace a broader view of self-assessment as a tool for motivating and enhancing learning. Indeed, a renewed interest in self-reflective practices is currently emerging in models and research on motivation and autonomy (e.g., Noels, Pelletier, Clement, Vallerand, 2000), as it appears that the transfer of some of the learning responsibilities from the instructor to the learner increases learner motivation. Self-assessment is a means by which such a transfer can take place, as it provides an opportunity for learners to become increasingly self-regulated. Noels, et. al. (2000) suggested that feedback, self-

reflection, motivation, and autonomy are interrelated in an ongoing, dynamic fashion, a sort of work in progress for the duration of a course. They also found in a subsequent study (Noels, et. al., 2000) that in an autonomy-supportive environment, students were less likely to feel anxious in the learning process and less likely to give up L2 learning. Thus, self-assessment provides a suitable interface between feedback, self-reflection, and increased autonomous learning, enabling both learner and instructor to reflect on the learning process and to give or receive mutual feedback (Noels, et. al., 2000).

Self-assessment seems to be a tool well-suited to helping learners to develop appropriate goals and self-regulate or monitor their efforts accordingly (Warne, 2008). From this perspective, self-reflective activities should not be considered the endpoint of the process (i.e., self-reporting of past performance), as they are traditionally defined in self- assessment research, but rather an ongoing, dynamic tool for reflecting concurrently on past and possible future performance and learning behavior.

The utilization of self-appraisal exercises gives a message to students that such exercises are viewed as a critical part of learning and that is worth spending valuable time on them. How intense this message is depending enormously on how well-designed self-appraisal assignments are and how well they fit the course of which they are a part. If students interpret them as an extra or an additional activity, the significance of self-evaluation will not be imparted (Boud, 1995).

Peer-assessment

Peer assessment is a procedure through which peers assess one another in which it may or may not involve agreed-upon criteria between teachers and students (Falchikov, 1995). Topping (1998, p. 250) described peer assessment as "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status". Brown and Hudson (1998) believed that "peer-assessment is a variant of self-assessment except that in peer-assessment students' language is rated by peers" (p.13). Van Gennip, Segers, and Tillema (2009) believe "peer-assessment involves collaboration in the appraisal of learning outcomes by those involved in the learning process, i.e., students" (P. 41).

The process of peer-assessment encompasses peers' grading and feedback on the works conducted by others (Davies, 2006). Peer-assessment is usually associated with group and teamwork. Patri (2002) stated that students could understand the assessment criteria in an assessment context in which they are assessing the quality of their performances.

A few scholars have faith in the peer-assessment process as a dialogue that should be done between the students. Sadler and Good (2006) pointed out that in peer-assessments, just telling learners what is right and wrong in their piece of writing, and how it might be improved would not on its own enhance learning nor develop deep disciplinary skills. Nicol (2010) points out that feedback ought to be conceptualized as a dialogue instead of as a restricted transmission process and notes that from this viewpoint, both the quality of feedback input and of students' responses to that input are important for productive learning. He points out that if students are to learn from feedback, they must have some chances to make their own meaning from the received message. They must do something with it, analyze it, ask questions about it, discuss it with others and connect it with prior knowledge.

Brown (2005) points out that peer-assessment engages critical standards like cooperative learning. Numerous individuals experience an entire regimen of training from kindergarten up through a graduate degree and never come to appreciate the value of collaboration in learning—the benefit of a community of learners capable of teaching each other something. Lejk and Wyvill (1996) pointed out that peer-assessment contributes to student-centered learning via preparing the students to judge

the nature of the work of others. Studies on the efforts of peer-assessment show that students find it more interesting than traditional methods of assessment, and it increases their motivation and improves their performance.

Brown (2004) argued that both self-assessment and peer-assessment develop students' autonomy and motivation. Cheng and Warren (2005), it is crucial for both teachers and students to understand the underlying rationale for the assessment. Peer-assessment, thus, can have critical pedagogical values since learners can participate in the evaluation process in which they can assess their peers' learning. Furthermore, Patri (2002) echoed the idea that peer-assessment and assessment involvement have critical pedagogical values for peers.

Advantages and disadvantages of applying self-assessment and peer-assessment

In these kinds of assessments, students individually assess their own and other's contributions through a priori criteria. According to Patri (2002) there is a predetermined procedure to conduct the assessment. These methods of assessments encourage student involvement and responsibility; allow students to see and reflect on their peers' assessment of their contribution; focus on the development of students' judgment skills, encourage student involvement and responsibility, and focus on the development of students' judgment skills (Gibbs, 1992).

The advantages of utilizing peer-assessment is an interesting issue in EFL studies. It is believed that peer-assessment enables students to develop abilities and skills to assess the work of other students. In other words, learners can analyze, monitor, and evaluate various issues of learning process and product of their peers. Studies in this domain show that it can work towards developing students' higher order reasoning and higher-level cognitive thought, helping to nurture student-centered learning among undergraduate learners, encouraging active and flexible learning and facilitating a deep approach to learning rather than a surface approach. Peer-assessment can act as a socializing force and increase relevant skills and interpersonal relationships between learner groups (Gibbs, 1992).

Brown and Hudson (1998) believe self-assessment and peer-assessment also have some disadvantages compared to teacher assessment. Subjectivity is a primary obstacle to overcome. Students may be too harsh on themselves or too self-flattering. Students may not have the necessary tools to make an accurate assessment. Students, especially in the case of direct assessment of performance, may not be able to discern their own errors.

Writing assessment

Assessment is defined as a way to improve writing instruction. When assessing, the assessors should take great care since all tests have consequences. The types of assessments that teachers typically undertake influence what and how writing is taught, what kind of feedback students receive about their writing, and which students get extra help from the teacher. Because assessment is evaluative, teacher assessments impact students' grades and perceptions of their writing competence (Andrade, Wang, Du, Akawi, 2008). As a result, the assessment should be taken seriously and must be based on best practices.

It is a critical concern for the teachers that their students' poor written English prevents them from obtaining their potential abilities either at institutes or in their professional lives. Although students have an understanding of their difficulty in writing, they are hindered from improving due to the gap existing between their own and their teachers' recognition of what the criteria for writing assessment are (Miller, 2003). Thus, students should know about the criteria for writing assessment, which may differ from one department to another, be explicit. Nezakatgoo (2011) pointed

out that the quest for learning professional skills has made it critical to teach writing in specialist subject areas.

There has also been research into the relative merits of analytic and holistic scoring schemes (e.g., Bacha, 2001). The latter, though considered easy to apply, are generally deemed to result in less information about individual performance and to have a limited capacity to provide diagnostic feedback, while the former, though they provide more information about individual performances, can be time-consuming to use. An interesting question is whether the assessment criteria provided are used at all. In his investigation of raters' use of a holistic scoring scale, Sakyi (2000) reported that not all raters focus on the scoring guide and he identified four distinct rating scales. Some raters focus on the errors in the text, others on the essay topic and presentation of ideas and yet others simply assign a score depending on their personal reaction to the text. Where raters consciously followed the scoring guide, they tended to depend on one or two particular features to distinguish between different levels of ability.

Studies on self-assessment, peer-assessment, and teacher assessment

There have been some studies done on self, peer, and teacher assessments from different perspectives. Some studies have focused on the specific benefits of self-assessments; some of them believe peer-assessment is the best method; however, some of them believe in teacher assessment. We summarise the findings of these studies in the following paragraphs.

Esfandiari and Myford (2013) compared three types of assessors (self-assessors, peer-assessors, and teacher assessors) to answer this question that if there is any significant difference in the severity of the three groups when rating essays. They analyzed the ratings of 194 assessors who evaluated 188 essays. There was a 6-point analytic scale with 15 assessment criteria. The results of their study indicated that teacher assessors were the most severe assessors. However, self-assessors were the most lenient assessors.

Birjandi and Tamjid (2010) examined the effectiveness of journal writing on developing Iranian EFL learners' motivation. 60 intermediate students who were divided into two groups participated in this study. There were proficiency tests at the start and end of the study to assure their homogeneity. Moreover, the participants filled questionnaires regarding motivation. The results revealed that writing journals on a regular basis had a role in promoting the learners' motivation. In another study, Birjandi and Tamjid (2012) examined the role of self-, peer, and teacher assessment in increasing the writing performance of language learners. The results illustrated that the students who utilized self-assessment and peer-assessment, together with teacher assessment, had the maximum improvement of writing.

Hasani and Rouhollahi (2012) investigated the effect of self-assessment on writing skills and proficiency of Iranian EFL learners. In this regard, prior to the study, 60 participants who were at the advanced level of language proficiency were given a standard pre-test to be homogenized and some instructions on scoring compositions based on grammatical relationships, structural relationships, word choice, and mechanics. Then, ten stages of writing were explained to the experimental group intermittently. After each stage, the participants of the experimental group were supervised. They concluded that the experimental group members did much better than those of the control group and the students could be fairly autonomous, in line with self-assessment and writing, gradually via ten stages of writing since after each stage of self-correction and self-evaluation, the assessment indicators and the related elements can be internalized which can be led to independence.

Brown (2005) found that the Japanese self-assessor in her small-scale study tended to underestimate her own writing ability while overestimating the writing abilities of her peers. Brown (2005) hypothesized that the Japanese self-assessor's

tendency to be overly critical of her abilities might reflect social and/or cultural factors coming into play.

Nakamura (2002) studied peer-assessment and teacher assessment in an EFL setting, Japan. In this small-scale study, Nakamura analyzed the ratings that five student assessors and one teacher assessor assigned to the presentations of 12 students in an oral presentation course. The raters used rating scales to evaluate the students' presentations. Nakamura (2002) reported that (a) the teacher assessor was more lenient than the peer-assessors were, but the peer-assessors varied in their levels of severity and leniency; (b) peer-assessment was successful in motivating students to improve their oral skills; and (c) students as peer-assessors could be reasonably reliable raters of their peers.

Meihami and Razmjoo (2016) investigated the challenges and solutions of utilizing self- and peer-assessment in writing courses. By conducting a series of interviews and open-ended questions, Meihami and Razmjoo (2016) delved into the perspectives of 11 ELT teachers and 56 ELT students to find out their challenges and probable solutions to applying self- and peer-assessment in the writing classes. The results of their study indicated that ELT students thought of self and peer's subjective judgment, learners' lack of assessment literacy, and instructional problems as the main challenges of self- and peer-assessment in writing classes. Moreover, ELT teachers believed that the challenges were the feasibility to do these types of assessment and ELT teachers' lack of assessment literacy. The common solution posited by both ELT teachers and students was development in teachers' assessment literacy.

As it can be understood from the above-mentioned empirical studies, there is a paucity of literature on the comparative effects of self-assessment, peer-assessment, and teacher assessment on intermediate EFL learners' writing performance. By doing this study, thus, we (1) fill the literature gap in this regard, and (2) answer the question that which type of assessment (self, peer, teacher) can be facilitative to develop Iranian intermediate language learners' writing performance.

Method
Participants
Eighty intermediate male and female Iranian EFL learners were selected to participate in the study. They were studying English at an English language institute. The participants' age ranged from 18 to 20, and their mother tongue was Persian. The lead researcher talked to the head of the institute about the purpose of the research and took his permission for carrying out the study. The students taking part in this study were of two types: Those newly registered into the English language classes and those who had been promoted from lower levels.

After the administration of the Oxford quick placement test of English language and taking the results into account, 20 participants were excluded from the study because of either a different language proficiency test score or not writing their names on the papers. There remained 60 learners to take part in the study. They were divided into three groups, and each group was randomly assigned to one type of groups: group A and group B acted as experimental groups, but group C as the control group: Group A: self-assessment; Group B: peer-assessment; and Group C: teacher-assessment.

Instruments and materials
To conduct the present study, we employed the following instruments. More explanation about these instruments is given below.
1. Oxford quick placement test of English language

2. The main book, American English File 3, at intermediate level introduced by the institute.
3. Jacob's ESL composition profile
4. Writing pre-test
5. Writing post test
6. Writing tasks

Oxford placement test of English language (OPT). In the present study, the Oxford placement test of English language was administered to homogenize the participants. Oxford placement test (OPT) is one of the most famous and standard tests for specifying ESL or EFL learners' level of language proficiency. The sample of Oxford placement test used in this study is divided into two parts: Part A: 60 items, part B: 20 items. Part A contained 25 multiple-choice pictorial items, 15 multiple choice cloze items, and 20 grammar multiple choice items. Part B contained 10 multiple choice cloze items, and 10 multiple choice vocabulary items. It took language learners 50 minutes to complete the test. The reliability of 60 items of the OPT test was estimated through a pilot study. The estimated value of Cronbach's Alpha for OPT test was .846 which was higher than the least possible value required (i.e. .70).

American English file book. American English File by Oxenden and Latham-Koeing (1997) is a four-skill course book with a communicative methodology, engaging texts, and a strong pronunciation syllabus which is designed to improve the English language of students. It was published by Oxford University Press.

Jacob's ESL composition profile. According to Farhady et al. (1994), the analytic method as one of the common scoring methods is the most valid and objective one in which the learners' writing components are scored based on some criteria. In this study, the formative analytical rating scale is an adaptation of the ESL composition profile in Jacob's et. al. (1981), which contains five basic criteria or main traits: content, organization, vocabulary, language use, and mechanics. Each trait was broken down into sub-traits. The levels of students' writing tests were rated based on this scale. Each paper was rated on these components. Two qualified raters, who were trained at the institute, scored the papers, and the results were analyzed to estimate the inter-rater reliability.

Writing pre-test and posttest. The pre-test and posttest of this study were conducted through essay writing. The sample tests were taken from IELTS sample tests published by Cambridge University Press. In each test, there were two tasks. The first task was about describing two charts in 20 minutes. The second one was about writing on a topic in 250 words in 40 minutes.
To assess each individual's writing for the pre-test and post-test of the three groups, the lead researcher used criteria that were adopted from Jacob's et. al.'s (1981) scoring profile. The writing performance of the participants was scored (based on the determined criteria) by two different scorers who were experienced foreign language teachers. The consistency of the raters' evaluations was examined using correlation analysis that showed a relatively high level of inter-rater reliability for both pre and post writing test scores. The results of the correlation coefficient between two raters turned out to be .719.

Writing tasks. The writing tasks of this study were taken from TOEFL iBT written by Macgillivray, Yancey, and Malarcher (2002). It was published by Compass Publishing in 2002. This book consists of four parts based on four skills. This is a preparation course book for TOFEL iBT. Four kinds of tasks were used during

treatment. These tasks were chosen under four categories: Paraphrasing, brainstorming, making connections, and organizing information.

Procedures

In order to achieve the aim of the study, the following procedures were followed. Initially, a total number of 80 participants were selected. To homogenize the participants, an Oxford Proficiency Test (OPT) was administered. As a result, 20 participants, who had scored more than one standard deviation away from (above or below) the mean, were excluded from subsequent statistical analyses, and there remained 60 homogenous participants to take part in this study. The participants were divided into three groups. Each group was randomly assigned to one of the treatment conditions.

In the next phase, the treatment began. Each group of participants received their treatment under one of the following conditions: Group A: self-assessment; Group B: peer-assessment; and Group C: teacher-assessment. In the first session of treatment in all the groups pre-test was given. Then, the teachers informed participants of all groups about the aims of the research and explained to students how to score the essays, using Jacob's ESL Composition Profile. In group A (self-assessment), students scored their own papers. In group B (peer assessment), students scored their classmates' papers. After that, the teacher checked students' scoring in group A and B and gave feedback to students. In group C (teacher's assessment), which was considered control group in this study, the teacher scored students' papers.

The experimental period lasted for 13 weeks, of which 10 weeks were allocated to the treatment, 2 weeks to the Oxford Placement Test and the pre-test, and 1 week to the posttests. It needs to be noted, however, that not all the class time was used for the treatment in each session. Since the learners were taking their general English course, only a third of each class time every week (about 45 minutes) was allocated to the experiment. At the end of the experimental period, two post-tests were administered.

Data analysis

The dependent and independent variables of the study are the learner's writing performance and type of assessment (self-assessment, peer-assessment, and teacher assessment), respectively. To gather data on the variables under investigation, two instruments were utilized. They were the general proficiency test and writing tests. The data gathered on these variables were analyzed by following methods through SPSS software

One-Way ANOVA procedure produced a one-way analysis of variance to check the homogeneity of the three groups regarding their general proficiency and writing performance at the outset of the study. To determine the mean and standard deviation, the researchers used descriptive statistics of each group on the pre and posttests. One-way ANOVA was used to find the differences of the participants in three groups on the posttest writing test. Analysis of variance was used to test the hypothesis that the means of the three groups were equal on writing posttest. Finally, to explore the participants' possible involvement, paired samples t-test were run on the results of pre and posttests of each group.

Results

To answer the research question, we used a one-way ANOVA procedure to test the hypothesis that the means of the three groups were equal on writing pre-test. Since Analysis of variance is sensitive to deviation from normality, the equality of the variances for the three groups was examined for the results of the posttest, too. It was assumed that each group was an independent random sample from a normal

population. To test this assumption, Levene's homogeneity of variance test was run for the results of the pre-test (Table 1).

Table 1: Test of homogeneity of variances for the pre-test scores

Levene Statistic	Df	df	Sig.
2.216	2	57	.118

Thus, the important first step in the analysis of variance indicated that the variances of the three groups were equivalent for the pre-test of writing (sig =.118, $\geq .05$). After confirming the normality assumption, ANOVA was run to the results of the writing posttest. The descriptive statistics for the writing posttest are presented in the Table 2:

Table 2: Descriptive statistics for the posttest scores

Groups	N	Mean	SD	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Group A (self-assessment)	20	16.05	1.82	.40	15.19	16.90	13.00	19.00
Group B (peer-assessment)	20	16.85	1.23	.27	16.27	17.42	15.00	19.00
Group C (teacher assessment)	20	15.15	1.08	.24	14.64	15.65	12.00	17.00
Total	60	16.01	1.55	.20	15.61	16.41	12.00	19.00

The results of the posttest showed that the mean of group A = 16.05, mean of group B = 16.85, and mean of group C = 15.15 differed significantly. The significance value of the *F* test in the ANOVA table (Table 2) was less than .05. Thus, the hypothesis that average assessment scores of the writing test (posttest) were equal across the three groups was rejected ($F_{(2, 57)} = 7.199$, $p < .05$), as can be seen in Table 3.

Table 3: ANOVA for the results of the writing test posttest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28.933	2	14.467	7.199	.002
Within Groups	114.550	57	2.010	-	-
Total	143.483	59	-	-	-
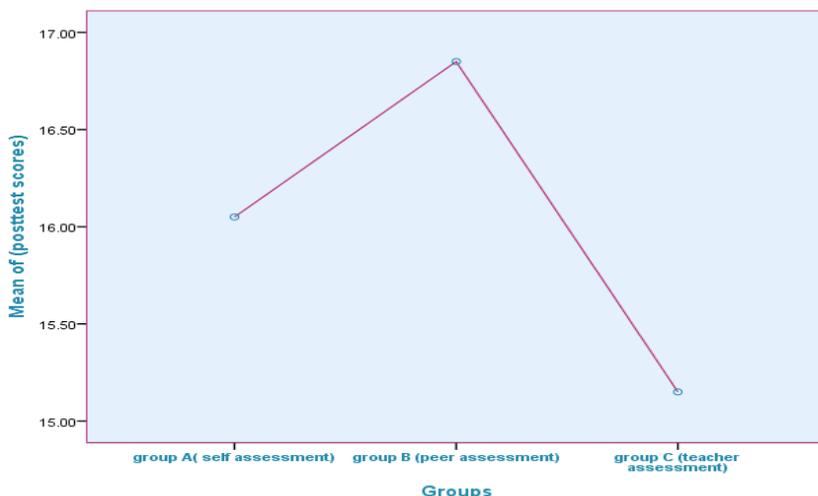
**Figure 1:** Mean plot for the results of the writing posttest

After it was revealed that the groups differed in some way, post-hoc test (Scheffe) (Table 4) disclosed more about the structure of the differences. The detailed structure of the differences was investigated through doing multiple comparisons. Post-hoc test (Scheffe) was employed for comparing the means of the three groups.

Table 4: Multiple comparisons for the results of the posttest

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Group A	group B	-.80000	.44829	.212	-1.9268	.3268
	group C	.90000	.44829	.143	-.2268	2.0268
Group B	group A	.80000	.44829	.212	-.3268	1.9268
	group C	1.70000*	.44829	.002	.5732	2.8268
Group C	group A	-.90000	.44829	.143	-2.0268	.2268
	group B	-1.70000*	.44829	.002	-2.8268	-.5732

Dependent Variable: Posttest scores
Scheffe
*. The mean difference is significant at the 0.05 level.

The highest mean difference was found between group (B) and group (C) (mean difference = 1.70). On the other hand, the lowest mean difference was reported for group (B) and group (A) (mean difference = .80). As it is shown in Table 4, group (B) outperformed the other two groups in terms of their writing performance (mean $_{group\ B}$ = 16.85; SD = 1.23). In the second place, group (A) performed better than group (C) (Mean $_{group\ A}$ = 16.05; SD = 1.82). Finally, the group (C) performance was lower than the other two groups (mean $_{group\ C}$ = 15.15; SD= 1.08). Figure 4 illustrates the performance of the three groups in the posttest.
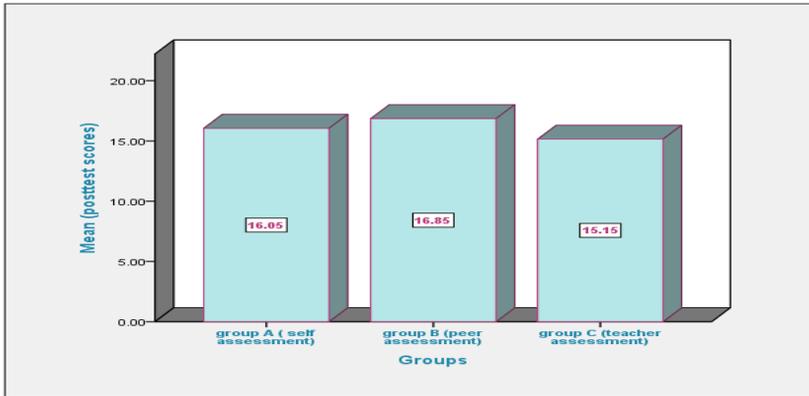
Figure 2: The three groups' performances on writing posttest

In general, *F* statistics established that there was a statistically significant difference between the three groups' means, and means plots showed the location of these differences. Participants of the experimental group B (peer-assessment) outperformed their counterparts namely group A (self-assessment) and group C (teacher-assessment).

In order to investigate students' performance within groups, three paired t-tests were also run, which showed the subjects' performance in pre-test and posttest that are shown in Table 5 and Table 6.

Table 5: Statistics for the pre and post-test scores of writing test

		Pretest group A	Posttest group A	Pretest group B	Posttest group B	Pretest group C	Posttest group C
N	Valid	20	20	20	20	20	20
Mean		14.05	16.05	14.15	16.85	14.02	15.15

The mean scores of group (A) improved from (14.05) in pre-test to (16.05) in posttest; that of the group (B) has changed from (14.15) in pre-test to (16.85) in posttest; and finally, the mean of the group (C) changed from (14.02) in pre-test to (15.15) in posttest on the writing test.

Table 6: Paired samples test for the pre and post-tests

Pair/pre-posttest		Paired Differences					T	df	Sig. (2-tailed)
		Mean	SD	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest group A –posttest group A	-2.00	.62	.14	-2.29	-1.70	-14.2	19	.00
Pair 2	Pretest group B – posttest group B	-2.70	.49	.11	-2.93	-2.46	-24.2	19	.00
Pair 3	Pretest group C – posttest group C	-1.12	.45	.10	-1.33	-.91	-11.0	19	.00

As depicted in the Table 5 and Table 6, all the three groups progressed in the posttest. Based on the results of paired t-tests, this progress was statistically significant for all

the three groups (*p*. <.05). In other words, all the three groups made a substantial progress on the writing posttest. This progress for group (B) was higher than the other groups. The lowest progress between the results of pre and posttests was observed for the third group that received instruction on teacher-assessment.

Discussion

The results of this study showed that peer-assessment outperformed the other two groups. Self-assessment gained the second position and teacher-assessment gained the lowest position among three groups. These findings are in line with those of Birjandi and Tamjid (2012) in which they examined the role of self-assessment, peer-assessment, and teacher assessment in increasing the writing performance of language learners. The results illustrated that the students who utilized self-assessment and peer-assessment together with teacher-assessment had the maximum improvement of writing. Abolfazli and Sadeghi (2013a) also examined the Iranian EFL students' attitudes toward self-, peer-, and teacher-assessment experiences. They found that students show positive attitudes toward self-, peer-, and teacher-assessment practices, with the peer-assessment group expressing significantly more positive than negative attitudes in this regard.

Abolfazli and Sadeghi (2013b) found that when self-assessment is contrasted with peer-assessment regarding their impact on students' scores, peer-assessment turned out to be more powerful. The same results were found by Chang, et. al. (2012) on portfolio-assessment in which they found that peer-assessment had the highest mean scores followed with the lowest scores by self-assessment and teacher-assessment. Sadler and Good (2006) concluded that peer assessors were stricter than self-assessors. Peer-assessment of writing has additionally been found to have a beneficial effect on the student writers' improvement (Brown, 2001; Patri, 2002).

Although the results of this study confirmed the prominent role of peer-assessment, there was an improvement in writing performance of self-assessment group, too. This finding is in line with that of Birjandi and Tamjid (2010) in which they explored the role of journal writing as a self-assessment technique in promoting Iranian EFL learners' motivation. The results revealed that writing journal on a regular base had a role in promoting the learners' motivation.

In another study, Hasani and Rouhollahi (2012) investigated the effect of self-assessment on writing skill and proficiency of Iranian EFL learners. They concluded that the experimental group did much better than the control group. With respect to self-assessment, Warne (2008) and Birjandi and Tamjid (2010) agree on the positive effect of self-assessment on the students' performance.

This study showed that peer-assessment and, to a lesser extent, self-assessment were two effective techniques in improving learners' writing performances. They encouraged students to have active participation in their own and their peers' learning. As learners knew they would not be evaluated by teachers, they could make more accurate and better pieces of writing and some psychological barriers such as tension and anxiety would hardly affect their results and this made learners more motivated and involved in the learning process. Since after each stage of correction and evaluation by self-assessment or peer-assessment, the assessment indicators and the related elements could be internalized.

Conclusion and Implications

In the present study, the efficacy of three kinds of assessment techniques (self-assessment, peer-assessment, and teacher assessment) on Iranian EFL learners' writing performance were examined. As the results showed, peer assessment gained the best position among the other groups. This finding confirmed the key role of

cooperative techniques in language teaching, specifically writing performance. The lowest position among these groups was for teacher assessment. It indicates that students are not so optimistic about the dominant role of teachers in the classroom, and they will have better improvement in cooperative atmosphere of the classroom.

The findings of this study can be useful for teachers to make the atmosphere of the class friendlier by emphasizing the ease of language learning through cooperative learning. The results of this study can also be used for situations in which students have problems due to affective factors. For instance, because of stress, they cannot talk with their teachers easily; consequently, they can work on their weaknesses by the help of their friends. This study can also be of great benefit to syllabus designers and curriculum developers as self-assessment and peer-assessment are applicable to both the design of the course books traditionally structured around these categories and to supplementary textbooks focusing on specific aspects of language skills or knowledge. If we provide opportunities for the students to assess their own and their peers' pieces of work, they will become more responsible for their own learning and meaningful learning will occur.

One of the limitations of this study was its difficulty to run at the outset. Due to the students' unfamiliarity with the concept of self-assessment and peer-assessment, most of the participants were reluctant to take part, but teacher's further clarification about self-assessment and peer-assessment and their objectives helped students to be used to and interested in it. The second limitation was the level of the students in this study. Only intermediate learners participated in this study. The third limitation of this study was that attitudes of participants were not investigated on self-assessment, peer-assessment, and teacher assessment.

Bibliographic references

ABOLFAZLI, Z. – SADEGHI, K. 2013a. Self-, peer-, and teacher-assessment: An investigation into Iranian EFL students' attitudes. In: Studies in Second Language Learning and Teaching. vol. 3, n. 1, pp. 87-107. ISSN: 2083-5205.

ABOLFAZLI, Z. – SADEGHI, K. 2013b. The effect of assessment type (self vs. peer) on Iranian university EFL students' course achievement. In: Social and Behavioral Sciences. vol. 70, n.1, pp. 1552-1564. ISSN: 1877-0428.

ALDERSON, C. – BANERJEE, J. 2001. Language testing and assessment. In: Language Teaching. vol. 34, no. 4, pp. 213-236. ISSN: 0261-4448.

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. In: Educational Measurement: Issues and Practice. vol. 27, n. 3, pp. 3-13. ISSN: 0731-1745.

BACHA, N. 2001. Writing evaluation: What can analytic versus holistic essay scoring tell us? In: System. vol. 29, n. 3, pp. 371-383. ISSN: 2079-8954

BACHMAN, L. – PALMER, A. S. 1989. The construct validation of self-ratings of Communicative language ability. In: Language Testing. vol. 6, n. 1, pp. 14-29. ISSN: 0265-5322.

BANIABDELRAHMAN, A. 2010. The effect of the use of self-assessments' performance in reading comprehension in English. In: The Electronic Journal for English as a Second Language. vol. 14, n. 2, pp. 147-156. ISSN: 1072-4303.

BIRJANDI, P. – TAMJID, N. H. 2010. The role of self-assessment in promoting Iranian EFL learners' motivation. In: English Language Teaching. vol. 3, n. 3, pp. 211-220. ISSN: 1916-4742

BIRJANDI, P. – TAMJID, N. H. 2012. The role of self-assessment in promoting Iranian EFL learners' writing performance. In: Assessment and Evaluation in Higher Education. vol. 37, n. 5, pp. 513-533. ISSN: 0260-2938.

BLANCHE, P. – MERINO, B. 1989. Self-assessment of foreign-language skills: Implications for teachers and researchers. In: Language Learning. vol. 39, n. 3, pp. 313- 338. ISSN:1467-9922.

BOUD, D. 1995. Enhancing learning through self-assessment. London: Kogan page limited. ISBN: 1135354413.

BROWN, A. 2005. Self-assessment of writing in independent language learning programs: The value of annotated samples. In: Assessing Writing. vol. 10, n. 3, pp. 174–91. ISSN: 1075-2935.

BROWN, D. 2004. Language assessment: Principles and classroom practice. New York: Longman. ISBN: 0130988340.

BROWN, H. D. 2001. Teaching by principles: An interactive approach to language pedagogy (2nd Ed.). San Francisco State University: Longman. ISBN: 0133282201.

BROWN, J. D. – HUDSON, T. 1998. The alternatives in language assessment. In: TESOL Quarterly. vol. 32, n. 4, pp. 653-675. ISSN:1545-7249.

BUCK, G. 1994. The appropriacy of psychometric measurement models for testing second language listening comprehension. In: Language Testing. vol. 11, n. 2, pp. 145-170. ISSN: 0265-5322.

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. vol. 58, no. 1, pp. 303-320. ISSN: 0360-1315

CHENG, W. – WARREN, M. 2005. Peer assessment of language proficiency. In: Language Testing. vol. 22, n. 1, pp. 93-121. ISSN: 0265-5322.

DAVIES, P. 2006. Peer assessment: Judging the quality of students' work by comments rather than marks. In: Innovations in Education and Teaching International. vol. 43, n. 1, pp. 69–82. ISSN: 1470-3297.

ESFANDIARI, R. – MYFORD, C. M. 2013. Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. In: Assessing Writing. vol. 18, n. 2, pp. 111-131. ISSN: 1075-2935.

FALCHIKOV, N. 1995. Peer feedback marking: Developing peer assessment. In: Innovations in Education and Training International. vol. 32, n. 2, pp. 175-187. ISSN: 1470-3297.

FARHADY, H. – JAFARPOUR, A. – BIRJANDI, P. 1994. Language testing: from theory to practice. Tehran: SAMT Publication. ISBN: 9645304644.

GIBBS, G. 1992. Teaching more students. Assessing more students. Oxford: The Polytechnic & Colleges Funding Council. ISBN: 1873576102.

HAMAYAN, E. V. 1995. Approaches to alternative assessment. In: Annual Review of Applied Linguistics. vol. 15, pp. 212-226. ISSN: 0267-1905.

HASANI, M. – ROUHOLLAHI, C. 2010. The effect of self- Assessment on Iranian EFL learners' writing skills. In: Iranian EFL Journal. vol. 8, n. 5, pp. 371-388. ISSN: 1836-8751.

HUERTA-MACIAS, A. 1995. Alternative assessment: Responses to commonly asked questions. In: TESOL Journal. vol. 5, n. 1, pp. 8-11. ISSN:1949-3533

LEJK, M. – WYVILL, M. (1996). A survey of methods of deriving individual grades from group assessments. In: Assessment & Evaluation in Higher Education. vol. 21, n. 3, pp. 267-280. ISSN: 0260-2938.

LINN, R. L., BAKER, E. L. – DUNBAR, S. B. 1991. Complex performance-based assessment: Expectations and validation criteria. In: Educational Researcher. vol. 20, n. 8, pp. 15–21. ISSN: 0013-189X.

MANGELSDORF, K. 1992. Peer reviews in the ESL composition classrooms: What do students think? In: ELT Journal. vol. 46, n. 3, pp. 274-284. ISSN: 0951-0893.

MACGILLIVRAY, M. – YANCEY, P. – MALARCHER, C. 2002. Mastering skills for the TOEFL iBT, advanced writing. Petersburg: Compass Publishing. ISBN: 1599660113.

MEIHAMI, H. – RAZMJOO, S. A. 2016. An emic perspective toward challenges and solutions of self- and peer-assessment in writing courses. Asian-Pacific Journal of Second and Foreign Language Education. vol. 1, n. 1, pp. 1-20. ISNN: 2363-5169.

MENDONCA, C. O. – JOHNSON, K E. 1994. Peer review negotiations: revision Activities in ESL writing instruction. In: TESOL Quarterly. vol. 28, n. 4, pp. 745-769. ISSN:1545-7249.

MILLER, P. J. 2003. The effect of scoring criteria specificity on peer and self-assessment. In: Assessment and Evaluation in Higher Education. vol. 28, n. 4, pp. 383-394. ISSN: 0260-2938.

MOUSAVI, S. A. 2012. An encyclopedic dictionary of language testing. Tehran: Rahnama Press. ISBN: 9643673197

NAKAMURA, Y. 2002. Teacher assessment and peer assessment in practice (Educational Studies 44). Tokyo, Japan: International Christian University. (ERIC Document Reproduction Service No. ED464483).

NEZAKATGOO, B. 2011. Portfolio as a viable alternative in writing assessment. In: Journal of Language Teaching and Research. vol. 2, n. 4, pp. 747-756. ISSN: 1798-4769.

NICOL, D. 2010. From monologue to dialogue: Improving written feedback in mass higher education. In: Assessment & Evaluation in Higher Education. vol. 35, n. 5, pp. 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. In: Language Learning. vol. 50, n. 1, pp. 57-85. ISSN:1467-9922.

PATRI, M. 2002. The influence of peer feedback on self- and peer-assessment of oral skills. In: Language Testing. vol. 19, no. 2, pp. 109-131. ISSN: 0265-5322.

ROTHMAN, R. 1995. Measuring up: Standards assessment, and school reform. San Francisco: Jossey-Bass Publishers. ISBN: 0787900559.

SADLER, D. R. – Good, E. 2006. The impact of self- and peer-grading on student learning. In: Educational Assessment. vol. 11, n. 1, pp. 1-31. ISSN: 1062-7197.

SAKYI, A. A. 2000. Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. In A.J. KUNNAN (Ed.), Fairness and validation in language assessment. Selected papers from the 19th Language Testing Research Colloquium, pp. 129-152. Orlando, Florida. ISBN: 0521651034.

SENGUPTA, S. 1998. Peer evaluation: I am not the teacher. In: ELT Journal. vol. 52, n. 1, pp. 19-28. ISSN: 0951-0893.

SMITH, K. 1999. Language Testing: Alternative Methods. In B. SPOLSKY (Ed), Concise encyclopedia of educational linguistics pp. 703-706. Amesterdam: Elsevier. ISBN: 0080431631.

TOPPING, K. 1998. Peer assessment between students in colleges and universities. In: Review of Educational Research. vol. 68, n. 3, pp. 249-276. ISSN: 1935-1046.

VAN GENNIP, N. A. E. – SEGERS, M. – TILLEMA, H. H. 2009. Peer assessment for learning from a social perspective: The influence of interpersonal and structural features. In: Learning and Instruction. vol. 4, n. 1, pp. 41-54. ISSN: 0959-4752.

WARNE, B. M. 2008. Writing steps: A recursive and individual experience. In: English Journal. vol. 97, n. 5, pp. 23-27. ISSN: 0013-8274.

WILLIAMS, J. C. 2003. Providing feedback on ESL students' written assignments. In: The Internet TESL Journal. vol. 9, n. 10, pp. 125-139. ISSN: 1072-4303

Words: 8026
Characters: 53 673 (23,82 standard pages)

Assistant Professor, Hussein Meihami, Ph.D.
Imam Khomeini International University
Department of English Language, Faculty of Humanities
Qazvin,
Iran
meihami@hum.ikiu.ac.ir

Assistant Professor, Rajab Esfandiari, Ph.D.
Imam Khomeini International University
Department of English Language, Faculty of Humanities
Qazvin,
Iran
esfandiari@hum.ikiu.ac.ir