

The Relationship between the Accuracy of Self- and Peer-assessment of Iranian Intermediate EFL Learners and Their Learning Styles

Parviz Birjandi

Department of English Language, Science and Research Branch, Islamic Azad University, Tehran, Iran

Mina Saghaieh Bolghari

Department of English Language, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract—Within the last few decades, with the growing emphasis on learner-centered curriculum, self-assessment and peer-assessment have become of particular interest in educational assessment. This study aimed at examining the relationship between the accuracy of self- and peer-assessment on the paragraph writing performance of a sample of Iranian intermediate EFL students and their learning styles. To do so, 7 paragraphs during 7 sessions were written and then self- and peer-rated by 62 students from Touba Language Institute in Tehran. Kolb's learning style inventory was used to determine the students' learning styles including diverging, assimilating, converging, and accommodating. The results revealed a significant and positive relationship between the accuracy of self- and peer-assessment of the students and their learning styles. The findings indicated that the students with converging learning style were the most accurate raters of their own performance, while those with diverging style were the least accurate ones. Furthermore, the students with accommodating learning style were the most accurate raters of peer-performance, whereas those with assimilating style were the least accurate ones. The obtained results may offer EFL teachers and educators the opportunity to design alternative assessment methods addressing learners' individual differences including learning styles.

Index Terms—learning style, peer-assessment, self-assessment, writing performance

I. INTRODUCTION

In recent years, language curriculum has witnessed a paradigm shift both in teaching and testing (Farhady, 2003). In testing, there is a shift “from psychometrics to a broader model of educational assessment, from a testing and examination culture to an assessment culture” (Gipps, 1994, p. 1). In line with this movement, the alternative means of assessment are considered as effective tools. These alternative assessments include the use of portfolios, conferencing, diaries and learning logs, teacher checklist, observations, self-assessment, and peer-assessment (Brown & Hudson, 1998; Huerta-Mac ías, 1995). Among the alternative assessments, self- and peer-assessment have gained much more interest owing to growing emphasis on learner independence and learner autonomy (Patri, 2002). Both self- and peer-assessment have significant pedagogical values. Some benefits of using self-assessment in educational contexts are enumerated by Oskarsson (1989) including, the promotion of learning, raised level of awareness, improved goal-orientation, expansion of range of assessment, shared assessment burden, and beneficial post-course effects. With regard to peer-assessment, studies have suggested that peer-assessment might increase reflection and generalization to new situations and promote self-assessment and greater meta-cognitive self-awareness. Furthermore, peer-assessment promotes a sense of ownership, personal responsibility, and motivation (Topping, 1998).

Despite the potential benefits of both self- and peer-assessment, most teachers have doubts about the accuracy of these techniques as measurement tools. This uncertainty stems from the variation in self- and peer-assessment validity which suggests potential difficulty in accurate interpretation (Ross, 1998). In this regard, a number of factors have been identified as being responsible for the variability seen in self- and peer-assessment results, the most important of which are the domain or skill being assessed, students' individual characteristics, and the ways in which questions and items are formulated and delivered (Butler & Lee, 2010). Yet, very few empirical studies have so far investigated the role of learning styles in the accuracy of self- and/or peer-assessment in an Iranian context. The primary objective of this study, therefore, was to examine the relationship between the accuracy of self- and peer-assessment of Iranian EFL learners and their learning styles while assessing their writing performance.

II. REVIEW OF THE LITERATURE

Self- and peer-assessment as a means of advancing learning

Self-assessment and peer-assessment are considered as useful techniques in advancing students' learning. By providing students with opportunities to evaluate their performance as well as giving them feedback based on the results of their assessment, students can become more aware of their own learning process and performance, and in turn they can become more proficient learners (Butler & Lee, 2010). In the literature, as a result, the role of self- and peer-assessment in improving learners' second or foreign language learning has been investigated extensively.

In an Iranian context with regard to writing performance, Javaherbakhsh (2009), for instance, investigated the impact of self-assessment on Iranian advanced EFL learners' writing performance. The results indicated that by using self-assessment as a learning technique, the students' writing performance was improved. In another study, Birjandi and Siyyari (2010) investigated the effect of doing self- and peer-assessment over time on the paragraph writing performance and the self- and peer-rating accuracy of Iranian English major students. The findings indicated that self- and peer-assessment were effective in improving not only the writing performance of the students but also their rating accuracy. Peer-assessment, however, turned out to be more effective in improving the writing performance of the students than self-assessment. Birjandi and Hadidi (2012), also, explored the role of self- and peer-assessment in promoting writing performance of Iranian intermediate TEFL students who participated in four experimental groups and one control group. The first experimental group did journal writing as a self-assessment technique, the second group self-assessed their own writings, the third group employed peer-assessment, and the fourth group had both self- and peer-assessment. The results revealed the maximum improvement in writing in the second and third groups, in which the students employed self-assessment and peer-assessment, together with teacher-assessment.

Self- and peer-assessment as measurement tools

The inherent subjectivity of self- and peer-assessment has traditionally been considered a threat to their validity as measurement tools (Butler & Lee, 2010). As a result, most studies on self- and peer-assessment have predominantly been interested in investigating the validity and reliability of these techniques. The results of such studies have been mixed, however (Bachman & Palmer, 1989; Blanche & Merino, 1989; Ross, 1998). Previous studies have indicated that there are many factors affecting learners' accuracy in self- and peer-assessment. For instance, Blanche (1988) has concluded from a comprehensive literature review that factors such as past academic record, career aspirations, peer group, or parental expectations, and lack of training in self-assessment can affect the subjectivity of learners in self-assessment. These and other factors are briefly explained in what follows.

Learners' language proficiency level is a factor which may affect the accuracy of self- and peer-assessment. In general, studies report that less proficient students tend to overestimate their language abilities, whereas more proficient students tend to underestimate their abilities (Blanche, 1988; Heilenman, 1990; Ross, 1998). Heilenman (1990) suggests the reason for this: "The more experience that learners have in a domain ... the more likely they are to be aware of the limits of their skills and knowledge" (p. 190).

One of the factors which can affect the results of self-assessment and/or peer-assessment is the domain or the skill being assessed. Ross (1998), for instance, conducted a meta-analysis of validation studies on self-assessment and found that self-assessing one's receptive skills has been found to be more accurate than for productive skills at least among adult learners. To Strong-Krause (2000), however, the converse was true; learners were more accurate in oral production than in other skills. He investigated task types which could best predict placement into English courses as a second language program. Results indicated that the speaking self-assessment was the best predictor, while the listening, writing, and reading self-assessments had lower correlation values, respectively.

Another factor that appears to affect the accuracy of self- and peer-assessment is how the items themselves are constructed. In this regard, Bachman and Palmer (1989) investigated self-rating of communicative language ability by implementing different questions types including ability, difficulty, and recognition. The results indicated that among these three types, the most effective appeared to be difficulty items. This suggests that, according to Bachman and Palmer, "foreign/second language users may be more aware of the areas in which they have difficulty than they are of the areas they find easiest." (p. 23) For Heilenman (1990), however, the claim that difficulty questions may be more effective than can-do questions on self-report instrument merits further research. To do so, he investigated the role of response effects (tendencies to respond to factors other than item content) in the self-assessment of second language ability by using can-do items as well as difficulty ones. While can-do items were positively worded, difficulty items were negatively-worded. According to Heilenman, while students paid attention to question content, they may also be responding to question wording. Based on the results, students ranked themselves significantly higher on can-do than on difficulty items. They found it more difficult to disagree with a negative item than to agree with a positive one.

Many scholars have recommended that before the actual assessment, learners should be trained on how to use self- and peer-assessment. According to Oscarson (1989), training can increase the reliability of learners' self-ratings. Similarly, Birjandi and Siyyari (2010) mention that the more practice and training the students have, the more accurate they get in their ratings. Jafarpur and Yamini (1995), however, reported the ineffectiveness of training on self-assessment accuracy and stated that in their study training helped the students to the extent that they were able to make judgments about their peers only. According to Jafarpur and Yamini, "... had the students received more training in the use of ... questionnaires, they would have been in a better position to make sound judgments about their own language abilities." (p. 74)

In the literature, the effective role of teacher or peer feedback in the accuracy of self- and peer-assessment has also reported. Patri (2002), for instance, investigated the agreement amongst teacher-, self- and peer-assessment of oral presentation skills of Chinese students in the presence of peer feedback and concluded that when assessment criteria was firmly set, peer-feedback enabled students to rate their peers more accurately, but the same was not found to be true with self-assessment. To have such influence of peer feedback on the accuracy of peer-assessment, according to Patri, it is necessary to consider students with similar English proficiency and same level of training. Only when it is the case, can teacher- assessment be supplemented with peer-assessment at least in the context of oral skills (Patri, 2002).

Complexity of a task may also affect the accuracy of the assessment. In this respect, Khabiri, Sabbaghan, and Sabbaghan (2011) aimed to investigate whether increasing task complexity increase the accuracy of peer-assessment of L2 oral production among female Iranian EFL learners. Three tasks were used in this study; the first and the simplest task was a descriptive narration, the second was a persuasive speech being more complex than the first task, and finally the third task was a debate being the most complex one. Absolute Mean Deviation (AMD) for each item of peer-assessment rating scale calculated to investigate the degree of agreement among participants who have assessed a particular speaker. Small AMDs were an indication of high degree of agreement among peer assessors. The results indicated that the AMDs of peer assigned scores decreases as task complexity increased. According to the authors, the AMDs decreased because complexity requires more attention and awareness. This increase in attention and awareness allowed the learners to be more accurate in their assessments.

Another affecting factor is the role of affective and psychological variables. Related to this, Alfallay (2004) investigated the role of some selected psychological and personality traits of EFL learners and the accuracy of self- and peer-assessment in oral presentation tasks. The selected traits were motivation types, self-esteem, anxiety, motivational intensity, and achievement. The study concluded that learners possessing the positive side of a trait were more accurate than those who have its negative side, with the exception of students with high classroom anxiety.

Finally, the assessors' severity, bias, attitudes, etc. are also examined as factors affecting the self- and peer-assessment accuracy. Saito and Fujita (2004), for instance, conducted a study on the comparison among peer-, self-, and teacher-assessment of L2 writing products in a Japanese context and reported a striking similarity between peer-assessment and teacher-assessment. About rater severity, results of this study showed that the group of self-raters included both the most lenient and severest raters. As to item difficulty, Saito and Fujita mentioned that for all three groups of raters, items related to mechanics, content, and length were easier on which to earn scores. Moreover, students had positive attitude toward using and receiving peer rating and also regarded themselves as being the least reliable among the three rater groups. In a similar study, Matsuno (2009) reported that many self-raters, particularly high-achieving students, underestimated their own writing performance. Peer-assessors on the other hand were shown to be internally consistent and their rating patterns were not dependent on their own writing performance. They also produced relatively few bias interactions. In addition, peer-raters were the most lenient raters; however, they rated high-achievers lower and low-achievers higher. In general, self-, peer-, and teacher-raters assessed grammar severely and spelling leniently.

As the findings of these studies indicated, self- and peer-assessment practices are affected by many factors, some of which are students' level of proficiency, the domain or skill being assessed, type of the questions, the amount of training and practice as well as teacher or peer feedback, complexity of a task, affective and psychological factors, etc. By controlling the effect of these factors and increase the effectiveness of self-assessment and peer-assessment, the optimal results may be achieved in educational settings.

Learning styles based on Kolb's Experiential Learning Theory

One of the main individual differences among learners is their different ways or styles of learning. There is a vast number of learning style classifications, among which Kolb's (1984) learning style inventory following his *Experiential Learning Theory* (ELT) seems very useful to indicate learners' differences. Like self- and peer-assessment, experiential learning aims to enable the learner to become increasingly self-directed and responsible for his or her own learning (Kohonen, 1992). ELT defines learning as "the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience" (Kolb, 1984, p. 41). The ELT model portrays two dialectically related modes of grasping experience including *Concrete Experience* (CE) and *Abstract Conceptualization* (AC) and two dialectically related modes of transforming experience including *Reflective Observation* (RO) and *Active Experimentation* (AE) (Kolb, 1984; Kolb & Kolb, 2005). Based on these four modes of learning, four learning styles are identified including, *Diverging*, *Assimilating*, *Converging*, and *Accommodating* (Figure 1). The basic characteristics of the learners with these learning styles are described as follows: (adapted from Kolb & Kolb, 2005, p. 5)

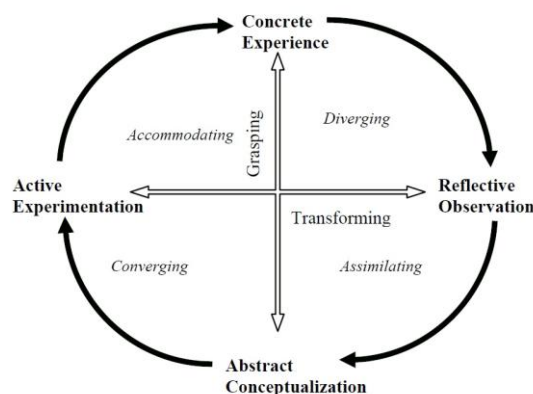


Figure 1. The experiential learning cycle and basic learning styles (Adopted from Kolb & Kolb, 2008, p. 6)

Diverging. An individual with diverging style has CE and RO as dominant learning abilities. People with this learning style are interested in people and tend to be imaginative and emotional. In formal learning situations, divergers prefer to work in groups, listening with an open mind to different points of view and receiving personalized feedback.

Assimilating. An individual with an assimilating style has AC and RO as dominant learning abilities. People with this learning style are less focused on people and more interested in ideas and abstract concepts. In formal learning situations, assimilators prefer readings, lectures, exploring analytical models, and having time to think things through.

Converging. An individual with a converging style has AC and AE as dominant learning abilities. People with this learning style prefer to deal with technical tasks and problems rather than with social interpersonal. In formal learning situations, convergers prefer to experiment with new ideas, simulations, laboratory assignments, and practical applications.

Accommodating. An individual with an accommodating style has CE and AE as dominant learning abilities. People with this learning style have the ability to learn from primarily “hands-on” experience and rely more heavily on people for information than on their own technical analysis. In formal learning situations, accommodators prefer to work with others to get assignments done, to set goals, to do field work, and to test out different approaches to completing a project.

Statement of the Problem and Research Questions

With regard to Kolb’s experiential learning theory, it might be suggested that any assessment should encourage students to apply their knowledge to a new experience, reflect on it, review and adapt strategies and try again (Meier, 2010). Yet, not all activities involve reflection, conceptualization, and experimentation. Although teachers grade or mark each piece of work, in many cases, there is a gaping hole in the learning cycle (Farhady, 2003). In this respect, self- and peer-assessment can provide learners with the opportunity to have a new experience of assessment, to reflect on their weakness and strength points, to conceptualize how to progress in learning, and finally to involve actively in the assessment process leading to a new experience. In the language assessment literature, however, Kolb’s experiential learning cycle is rarely examined.

From a different perspective, it is believed that matching learners’ learning with their preferred learning styles appropriately can greatly influence and improve learners’ learning performance, especially for EFL learners (Reid, 1987). Therefore, when learning styles may have a role in language learning, does it have the same impact on language assessment and more specifically on self- and peer-assessment? Although such a question seems to be of utmost importance, little work has been done in this regard. Hence, the purpose of this study was to investigate whether there was any relationship between the accuracy of self- and peer-assessment of Iranian intermediate EFL students and their learning styles in the context of assessing students’ writing performance. In addition, the study also sought to investigate whether there were significant differences between teacher-rating and student-assessment when the learning styles were taken into consideration. In other words, how did the participants who possess those learning styles rate their own performance and how did they rate the peers’ performance? The research questions, therefore, to which this study aimed to find the answers, are:

RQ₁: Is there any relationship between the accuracy of self-assessment of Iranian intermediate EFL learners and their learning styles?

RQ₂: Is there any relationship between the accuracy of peer-assessment of Iranian intermediate EFL learners and their learning styles?

RQ₃: Are there significant differences between teacher-assessment, self-assessment, and peer-assessment of Iranian intermediate EFL learners with regard to their learning styles?

Given the above research questions, the following null hypotheses were proposed:

RH₀₁: There is not any significant relationship between the accuracy of self-assessment of Iranian intermediate EFL learners and their learning styles.

RH₀₂: There is not any significant relationship between the accuracy of peer-assessment of Iranian intermediate EFL learners and their learning styles.

RH₀₃: There are not significant differences between teacher-assessment, self-assessment, and peer-assessment of Iranian intermediate EFL learners with regard to their learning styles.

III. METHODOLOGY

Participants

The participants of this study consisted of 62 female intermediate EFL students, making up four classes in Touba language institute. Due to the limited number of the students in one class, the study was conducted in four different classes. There were 13, 16, 16, and 17 students in each class, respectively. All the participants were Iranian female students with mostly Persian language background and were mainly high school students. They ranged in age from 16 to 20 years old, with a few exceptions over the age of 20. In each class, the students were exposed to the same content and instructional method with the same instructor who was an experienced English language teacher, holding a Master degree in TEFL.

Instruments

Proficiency test

The participants' proficiency level was determined by means of the *Preliminary English Test (PET) for Schools* (version 2004). *PET for schools* is a test that shows how a pupil can deal with everyday written and spoken English at an intermediate level. It follows exactly the same format as *the Preliminary English Test (PET)* and the level of the questions is identical. The only difference is that the content and treatment of the topics in *PET for Schools* have been particularly targeted at the interests and experience of teenagers and school pupils.

Learning style inventory

To elicit the learning styles of the participants, Kolb's learning style inventory version 3.1 (2005) was employed after obtaining the official permission. The format of the inventory is a forced-choice format that ranks an individual's relative choice preferences among the four modes of the learning cycle. This is in contrast to the more common normative, or free-choice, format such as the widely used Likert scale, which rates absolute preferences on independent dimensions (Kolb & Kolb, 2005). The participants should read an item and then rank their responses, on a scale of 1-4, to each of the 12 items on the inventory.

Writing scale

The writing scale employed for scoring the paragraphs of the participants was adopted from *ESL Composition Profile* developed by Jacobs et al. (1981). Both raters (teacher and the researcher herself) used this scale to score the participants' paragraphs, and the participants themselves used it for the purpose of self- and peer-assessment. In Jacobs' scale, scripts are rated on five aspects of writing; content, organization, vocabulary, language use, and mechanics with a brief description for each part. The ESL composition profile, however, was not used in its original form since it was too difficult for the participants of this study to use due to their limited experience on English writing performance. Therefore, all the descriptions and the components of the scale were simplified; they were translated to Persian, the wording of the descriptions was made simpler and less technical, and the phrasal structures were changed into complete sentences to reduce the complexity of the descriptions. The Jacobs' scale is a weighted scale with mechanics being the least weighted (5%) and content the most (30%). However, in this study, items were weighted equally with a 4-point rating scale.

Procedures

Data were collected during regular class time and over a period of 4 weeks from 9th to 16th session in an 8-week semester. The reason for collecting the data during these sessions and not right at the beginning of the semester was to ensure that the researcher became familiar with the students and the teacher's teaching style. Furthermore, it should be ensured that all the participants had enough experience to write a paragraph on a given topic in English, for lack of experience in the domain or skill being assessed may cause inaccuracy in the results of self- and peer-assessment (Bachman & Palmer, 1989; Heilenman, 1990; Ross, 1998). The whole procedures were identical for four classes. In order to answer the main research questions, the following procedures were pursued:

Administrating proficiency test. As a first step, the PET for Schools was administrated to all the students to determine their proficiency scores.

Administrating learning style inventory. As a next step, Kolb's learning style inventory was given to the students to identify their learning styles including diverging, assimilating, accommodating, and converging. Official permission was obtained to use version 3.1 (2005) of this inventory in the current study.

Observing paragraph writing instruction. To familiarize with the students, classroom context, and teacher's teaching style, the researcher observed three sessions of instruction in one of the classes. Since the basis for self-/peer-assessment was the learners' writing performance, the focus of the observation was on the writing instruction, in particular. As it is common in many conversation classes, the course offered an integrated approach to learning four language skills and grammar and vocabulary were taught simultaneously. However, none of the participants had received extensive writing instruction before this course. Particularly, during this course there was an attempt to make paragraph writing instruction an integral part of the course in all classes. Paragraph writing instruction was based on

Arnaudet and Barrett's Paragraph Development (1990), based on which several methods of paragraph development were introduced to the students.

During the writing instruction sessions, the students were instructed how to develop the essential parts of a paragraph including a topic sentence, supporting sentences, and a conclusion. In addition, they became familiar with the basic components of the paragraph writing including content, organization, choice of words, grammar usage, etc. Teacher monitoring as well as peer collaboration were the main instructional techniques for paragraph writing instruction in these classes.

Training self-/peer-assessment. After receiving paragraph writing instruction for eight sessions, the participants were introduced to self-assessment and peer-assessment through a training session. The training session lasted for about two hours of class time, the main purpose of which was to explain the assessment criteria (rating scale).

In order to familiarize the participants with the rating scale, the researcher elaborated on the scale items by explaining what they should focus on while assessing their own and their peers' writing performances. After elaborating on the items, the researcher explained the descriptions provided for each item by giving some examples. In order to clearly establish the assessment criteria, each student received a sample paragraph rated before by the researcher based on the same scale. The errors were corrected and the necessary feedback was given by the researcher on the erroneous parts. The participants used this sample as a guideline to rate their own as well as their peers' paragraphs. After becoming familiar with the scale, the participants practiced assessing three samples together in class consisting of a good, an average, and a poor one which had been rated by the raters in advance. All the students reported by rising their hands what score they had given to each item. The researcher questioned some students why a particular score was given and explained why certain scores were more appropriate than the others and tried to prevent too much diversity in the ratings. Having rated all the samples, the students compared their own ratings with those of the raters and the rating ambiguities were tried to be resolved.

Practicing self- and peer-assessment. Having been introduced to the scale and practiced the paragraph rating, the participants were asked to write a paragraph (100-150 words) on a topic chosen by the instructor based on the students' familiarity with it. That is, those topics were chosen for paragraph writing with which the students had become familiar during the course. Owing to the time-restriction, the students were asked to write the paragraph as a homework assignment and bring it back the next session. Some copies of each paragraph were made by the participants to subject them to peer and teacher assessments.

To do peer-assessment, it was necessary to assign the students into some style-alike groups. Based on the results of the learning style inventory, there were 22 accommodators, 18 assimilators, 13 divergers, and 9 convergers. Table 3.1 shows the distribution of the students with different learning styles in each class.

TABLE 1
DISTRIBUTION OF THE STUDENTS WITH DIFFERENT LEARNING STYLES IN EACH CLASS

| Learning style | Class 1 | Class 2 | Class 3 | Class 4 |
|----------------|---------|---------|---------|---------|
| Accommodators | 7 | 5 | 7 | 3 |
| Assimilators | 7 | 4 | 4 | 3 |
| Divergers | 3 | 4 | 3 | 3 |
| Convergers | 0 | 3 | 2 | 4 |

To do peer-assessment more easily, the groups whose members were more than four (five and seven) were divided into two sub-groups; seven-member groups were divided into groups three and four members and five-member groups were divided into groups two and three members. Therefore, each group consisted of two to four members.

Every session, the students were asked to rate their own as well as their peers' paragraphs based on the scale and give the necessary feedback to peer-performance in English or even in Persian. First, they self-rated their own paragraphs. Then, each student handed a copy of her own writing to all group members to do peer rating. Peer-assessment practice was repeated for all group members. That is, in a four-member-group, for example, each student rated three peers' paragraphs. Depending on the number of the students in each group, it lasted for about 15-30 minutes for the learners to rate their own and their peers' paragraphs. Having assessed their own and their peers' performances, the students handed all the copies to the researcher. Both the researcher and the teacher rated the writings and gave them back to the students in the next session. Every session, the students received all the ratings and comments of the previous session and could compare their own self-assessment score with those of the other raters. The same procedure was followed for seven sessions in all classes. During this time, the students were gradually liberated from the researcher's guidance and support and could rate the writings more independently.

IV. DATA ANALYSIS

Before answering the research questions and testing the hypotheses, the results of the proficiency test as well as the learning style questionnaire were analyzed.

Analyzing data obtained from proficiency test. At the beginning of the study, the PET for Schools was administrated to all the students to determine their proficiency scores. Table 2 provides the descriptive statistics for the participants' proficiency raw scores out of 62.

TABLE 2
DESCRIPTIVE STATISTICS ON PARTICIPANTS' PROFICIENCY SCORES

| Classes | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|----|---------|---------|---------|----------------|
| Class 1 | 17 | 70.00 | 90.00 | 79.7059 | 5.64319 |
| Class 2 | 16 | 72.00 | 87.00 | 77.5000 | 7.71146 |
| Class 3 | 16 | 70.00 | 86.00 | 75.1250 | 6.50000 |
| Class 4 | 13 | 73.00 | 89.00 | 77.7692 | 6.15296 |
| Valid N (listwise) | 13 | | | | |

Since the level of English language proficiency was a relevant factor to the self-/peer-assessment skill of the participants (Heilenman, 1990; Ross, 1998), the proficiency means of the groups were compared to see how different they were from each other. To start with, a between-groups analysis of variance (ANOVA) was run in order to check the homogeneity of the groups in terms of their English language proficiency. The mean scores of the four groups on the language proficiency test were compared. The result showed that the difference among the groups was not statistically significant [$F(3, 58) = .464, p > .05$]; therefore, the groups could be considered as homogeneous in terms of their English language proficiency (see Table 3).

TABLE 3
ONE WAY ANOVA FOR PET SCORES

| Classes | Sum of squares | df | Mean square | F | Sig. |
|----------------|----------------|----|-------------|------|------|
| Between groups | 59.768 | 3 | 19.923 | .464 | .708 |
| Within groups | 2489.587 | 58 | 42.924 | | |
| Total | 2549.355 | 61 | | | |

Analyzing data obtained from learning style inventory. As a second step, Kolb learning style inventory was used to identify one's learning style. Table 4 shows the number and percentage of each learning style within 62 students. Table 5 shows the descriptive statistics of the four learning orientations for the whole sample.

TABLE 4
NUMBER AND PERCENTAGE OF THE STUDENTS WITH DIFFERENT LEARNING STYLES

| Learning styles | N | % |
|-----------------|----|-------|
| Accommodator | 22 | 35.48 |
| Assimilator | 18 | 20.96 |
| Diverger | 13 | 29.05 |
| Converger | 9 | 14.51 |

TABLE 5
DESCRIPTIVE DATA FOR LEARNING STYLE ORIENTATIONS

| Learning orientations | N | Minimum | Maximum | Mean | Std. Deviation |
|-----------------------|----|---------|---------|---------|----------------|
| AE | 62 | 17.00 | 44.00 | 30.7419 | 5.82514 |
| RO | 62 | 12.00 | 40.00 | 25.0161 | 7.38073 |
| AC | 62 | 13.00 | 42.00 | 29.2903 | 6.92675 |
| CE | 62 | 15.00 | 43.00 | 25.8226 | 7.29591 |
| Valid N (listwise) | 62 | | | | |

Reliability analyses. The internal consistency of the learning style inventory was computed by using Cronbach's Alpha. Furthermore, the inter-rater reliability between two raters was calculated by using Pearson correlation coefficient. The learning style inventory was pre-tested with 20 students similar to the population of the main study (Table 6). The results of this study are lower than the studies reviewed by Kolb and Kolb (2005, p. 14), but lie within the acceptable range.

TABLE 6
RELIABILITY COEFFICIENTS FOR LEARNING STYLE INVENTORY VERSION 3.1

| Measure | N. of Items | Alpha Coefficient |
|---------|-------------|-------------------|
| AC | 12 | .65 |
| CE | 12 | .62 |
| AE | 12 | .70 |
| RO | 12 | .66 |

To ensure the learners' accuracy in rating self- and peer-performances, the correlations are calculated between self- and peer-assessment scores and scores obtained through various types of external measurements such as objective tests, final grades, and teachers' ratings (Butler & Lee, 2010). In this study, teacher-assessment was the criterion measure to check the accuracy of the students in assessing self- and peer-performances. Both the teacher and the researcher herself rated the paragraphs by following the same rubric. Inter-rater reliability was calculated for the average scores of both raters rating the paragraphs during seven sessions. The degree of association between the ratings provided by two raters was $r = .93$, which is significant at the 0.01 level (Table 7).

TABLE 7
PEARSON PRODUCT-MOMENT CORRELATION FOR INTER-RATER RELIABILITY

| | R1 | R2 |
|------------------------|--------|--------|
| R1 Pearson correlation | 1 | .934** |
| Sig. (2-tailed) | | .000 |
| N | 62 | 62 |
| R2 Pearson correlation | .934** | 1 |
| Sig. (2-tailed) | .000 | |
| N | 62 | 62 |

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

Testing the research hypotheses. The numerical data for answering the research questions 1 and 2 came from paragraph writing performance scores given by two raters and the students themselves across seven sessions. The final scores were the average of self-assessment, peer-assessment, and teacher-assessment rounded to the closest integer. Hence, research hypotheses 1 and 2 were tested by correlating the average scores of self-assessment (SA) and peer-assessment (PA) with the average score of teacher-assessment (TA) with regard to the students' different learning styles. The results are presented in Table 8 for accommodating, assimilating, diverging, and converging groups.

TABLE 8
CORRELATION COEFFICIENTS BETWEEN TA, SA, AND PA OF ALL GROUPS

| Assessment Variable | SA-TA | PA-TA | SA-PA |
|-----------------------------|--------|--------|--------|
| Accommodators (N= 22) | | | |
| Correlation coefficient (r) | .735** | .879** | .556** |
| Significance | .000 | .000 | .007 |
| Assimilators (N= 18) | | | |
| Correlation coefficient (r) | .668** | .594** | .619** |
| Significance | .002 | .009 | .006 |
| Divergers (N= 13) | | | |
| Correlation coefficient (r) | .613* | .751** | .262 |
| Significance | .003 | .003 | .388 |
| Convergers (N= 9) | | | |
| Correlation coefficient (r) | .836** | .803** | .809** |
| Significance | .005 | .009 | .008 |

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

* . Correlation is significant at the 0.05 level (2-tailed).

In all groups, the correlations were statistically significant ($P < .01, .05$), with the exception of the correlation between SA and PA of the diverging group. In the case of the accommodating and diverging groups, the correlation between PA and TA was higher than that of SA-TA. In the assimilating and converging groups, however, the correlation between SA and TA was higher than that of PA-TA. The highest coefficient obtained in this study was between TA and PA ($r = .87$) for the accommodative learners. In addition, the lowest coefficient was between TA and PA for the assimilative learners ($r = .59$). Therefore, the accommodators were the most accurate raters of peer-performance, whereas the assimilators were the most inaccurate ones. Regarding self-assessment, the highest coefficient was that of convergers ($r = .83$), while the lowest one was that of divergers ($r = .61$). Hence, the most accurate self-raters were the convergers, whereas the least accurate ones were the divergers.

As the results indicated, the students with converging learning style were the most accurate raters of their own writing performance, whereas those with accommodating style were the most accurate raters of their peers' performance. Therefore, the null hypotheses 1 and 2 were rejected, and it could be claimed that there were significant and positive relationships between the accuracy of self- and peer-assessment of the Iranian intermediate EFL learners and their learning styles.

Before answering the last question, it was revealed that the data were not normally distributed ($p < .05$). Table 9 shows the results of Kolmogorov-Smirnov and Shapiro-Wilk tests. Consequently, the non parametric Mann-Whitney U test was used to compare the means of SA-TA, PA-TA, and SA-PA in all groups. The results are presented in Table 10 for all learning style groups.

TABLE 9
TEST OF NORMALITY

| Classes | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|---------|---------------------------------|----|------|--------------|----|------|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| TA | .176 | 62 | .000 | .907 | 61 | .000 |
| SA | .132 | 62 | .008 | .902 | 63 | .000 |
| PA | .123 | 62 | .021 | .948 | 62 | .011 |

a. Lilliefors Significance Correction

TABLE 10
MANN-WHITNEY TEST STATISTICS FOR ALL GROUPS

| Assessment Variable | SA-TA | PA-TA | SA-PA |
|-----------------------|--------|--------|--------|
| Accommodators (N= 22) | | | |
| Z-Value | -4.312 | -2.512 | -2.000 |
| Significance | .065 | .079 | .260 |
| Assimilators (N= 18) | | | |
| Z-Value | -6.602 | -1.629 | -2.096 |
| Significance | .547 | .103 | .036 |
| Divergers (N= 13) | | | |
| Z-Value | -1.796 | -2.827 | -.601 |
| Significance | .072 | .005 | .548 |
| Convergers (N= 9) | | | |
| Z-Value | -1.927 | -.526 | -.882 |
| Significance | .054 | .526 | .378 |

As seen in the Table, all differences between the mean scores of the accommodators are statistically non-significant ($Z = -4.31, -2.51, \text{ and } -2.00; p < .05$). For the assimilators, the only significant difference is between the means of SA and PA ($Z = -2.096, P < .05$). With regard to the divergers, the only significant difference is between the means of PA and TA ($Z = -2.827, P < .05$). There were no significant differences between TA, SA, and PA in the case of the convergers.

The only significant difference between PA and TA was found for the divergers. Despite the high coefficient obtained between PA and TA ($r = .75$), the significant difference between the means of TA and PA implies that the divergers rated peer-performances differently from those of the teachers. That is, they overrated peer-performance much more than the other groups. Moreover, the only significant difference between SA and PA obtained for the assimilators which indicated that they rated self-performance differently from that of peer-performance. That is, they overrated self-performance much more than peer-performance comparing to other groups. Hence, it could be suggested that self- and peer-assessment of the accommodators and the convergers whose differences were statistically non-significant were consistent with teacher-assessment. It means that their given scores were much closer to those of the teachers.

The results of Mann-Whitney U-test indicated that there were some significant differences between the mean scores of PA and TA, and SA and PA in diverging and assimilating learning style groups. Therefore, the null hypothesis 3 was rejected in these two groups but supported in the converging and accommodating groups in which there were not significant differences between self-, peer-, and teacher-assessment.

V. DISCUSSION

The study revealed the role of learning styles in the accuracy of self- and peer-assessment. This is in line with the studies of Cassidy and Eachus (2000) and Cassidy (2006) who reported the relevance of learning styles to the students' self-assessment skill. In the current study, Kolb's learning style inventory was used to identify the participants' different learning styles. Based on the results, it seems that learners with active-experimentation mode of learning (convergers and accommodators) were more accurate than those with reflective observation (assimilators and divergers) in estimating of their own and their peers' performances since they indicate an active orientation to learning that relies on experimentation. As Kolb and Kolb (2005) describe the characteristics of the individuals with active-experimentation mode of learning, they enjoy problem solving, small group discussion, peer feedback, and self directed assignments and so for these learners the practice of self- and peer-assessment can be beneficial to actively engage in learning and assessment. Furthermore, learners with abstract conceptualization (convergers and assimilators) were more accurate than those with concrete experience (accommodators and divergers) in rating self-performance since they indicate an analytical, conceptual approach to learning that relies on logical thinking and rational evaluation (Kolb & Kolb, 2005). In addition, learners with concrete experience were better in peer-performance estimation (accommodators vs. convergers and divergers vs. assimilators) since they are extroverted, emphatic, and people-oriented individuals who enjoy peer works and may be lenient in rating peer-performance. For these reasons the convergent learners were the most accurate raters in estimating of their own performance, whereas the divergent learners were the least accurate ones. Moreover, the accommodative learners were the most accurate raters in rating peer-performance, while the assimilative learners were the least accurate ones.

As the results indicated, the highest coefficient was found between teacher-assessment and peer-assessment ($r = .87$). This is in line with the studies of AlFallay (2004), Patri (2002), and Saito and Fujita (2004) who reported the high coefficients between teacher-assessment and peer-assessment ($r = .92, .85, \text{ and } .72$, respectively). In general, all correlations obtained in this study were significant and relatively high (ranged .55 to .87). One possible explanation could be that the assessment criterion (questionnaire), was clear enough for the students to rate the samples. According to Oscarson (1989), if the items are delivered in the students' first language, they can assess their own or their peers' performances more accurately compared to when the items are delivered in their target language. Another reason may be related to the presence of the training session (AlFallay, 2004; Birjandi & Siyyari, 2010, Patri, 2002) although Jafarpur and Yamini (1995) reported the effect of the training on only peer-assessment accuracy. Number of the practice sessions may also affect the results. It was expected that each assessment session would add to the accuracy of

ratings assigned. This is in line with the study of AlFallay who believes that long period of practice yields homogeneous groups with respect to rating their own performances and those of their peers. In addition, it familiarizes students with their peers' language ability. Finally, it seems that the presence of teacher and peer feedback can lead to the accurate estimation of students in self- and peer-rating (AlFallay, 2004; Patri, 2002).

VI. CONCLUDING REMARKS

As the findings of this study indicated, the accuracy of self- and peer-assessment of the learners was affected by their learning styles. This conclusion revealed the fact that language assessment is a multifaceted process affected by various psychological and personality traits of the raters (AlFallay, 2004). In this regard, students' learning styles should be emphasized as well. Teachers should design a variety of assessment methods that cater for the learning style preference favored by the majority of the students in the class. The findings of this study showed that the majority of the students preferred accommodating and assimilating learning styles. This means that both hands-on and reflective activities should be emphasized in the class to attune to the diversity of students' learning preferences. In addition, the study suggests that peer- and self-rating can be valid and reliable assessment tools. The high correlations obtained here between teacher-, self-, and peer-assessment support this claim. This reemphasizes the need to integrate self- and peer-assessment into language teaching and testing.

The present study focused on Kolb's learning style inventory to elicit the participants' different learning styles. Future research is needed to examine whether the application of other learning style instruments can affect the accuracy of self- and peer-assessment. Moreover, this study investigated only the students' self- and peer-assessment on their paragraph writing performance and so in other skills or sub-skills it needs further research. In addition, this study was conducted with the students of intermediate level. As such, generalizations to elementary or advanced levels would not be appropriate without further research. Finally, owing to the access of sampling only 62 female students were selected to participate in the present study, which implies follow-up research with more participants.

As a concluding remark, it should be noted that self- and peer-assessment are valuable learning activities even in the absence of high coefficients between student-assessment and teacher-assessment. Although the unreliability of self- and peer-assessment limits some teachers to use them only for developmental purposes (Saito & Fujita, 2004), one should keep in mind that even experienced teachers are not always reliable markers and so it seems unreasonable to expect inexperienced students always to demonstrate reliability (Falchikov & Boud, 1989). This means that the limitations should not downgrade the pedagogical values of self-assessment and peer-assessment in language education.

REFERENCES

- [1] AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32, 407-425. doi:10.1016/j.system.2004.04.006.
- [2] Arnaudet, M. L., & Barrett M. E. (1990). *Paragraph development: A guide for students of English* (2nd ed.). New Jersey: Prentice Hall.
- [3] Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6(1), 14-29. doi: 10.1177/026553228900600104.
- [4] Birjandi, P., & Hadidi, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533. doi:10.1080/02602938.2010.549204.
- [5] Birjandi, P., & Siyyari, M. (2010). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *IJAL*, 13, 23-45.
- [6] Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, 19(1), 75-93. doi: 10.1177/003368828801900105.
- [7] Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313-340. doi: 10.1111/j.1467-1770.1989.tb00595.x.
- [8] Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675. doi: 10.2307/3587999.
- [9] Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among learners of English. *Language Testing*, 27(1), 5-31. doi: 10.1177/0265532209346370.
- [10] Cassidy, S. (2006). Learning style and student self-assessment skill. *Education + Training*, 48(2), 170-177.
- [11] Cassidy, S., & Eachus, P. (2000). Learning style, academic belief systems, self-report student proficiency and academic achievement in higher education. *Educational Psychology*, 20(3), 7-22.
- [12] Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395-430. doi: 10.3102/00346543059004395.
- [13] Farhady, H. (2003). Classroom assessment: A plea for change. Paper presented at the Second Conference on Issues in English Language Teaching, Tehran, Iran.
- [14] Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Fulmer Press.
- [15] Heilenmann, K. L. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7(2), 174-201. doi: 10.1177/026553229000700204.
- [16] Huerta-Mac ís, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1), 8-11.
- [17] Jacobs, H. J., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Massachusetts: Newbury House.

- [18] Jafarpur, A., & Yamini, M. (1995). Do self-assessment and peer-rating improve with training? *RELC Journal*, 26(1), 63-85. doi: 10.1177/003368829502600104.
- [19] Javaherbakhsh, M. (2008). The impact of self-assessment on Iranian EFL learners' writing skill. (Unpublished master's thesis). Science and Research Branch of Islamic Azad University, Tehran, Iran.
- [20] Khabiri, M., Sabbaghan, S., & Sabbaghan, S. (2011). The relationship between peer assessment and the cognition hypothesis. *MJAL*, 3(2), 278-299.
- [21] Kohonen, V. (1992). Experiential language learning: Second language learning as cooperative learner education. In D. Nunan (Ed.), *Collaborative language learning and teaching* (pp. 14-39). Cambridge: Cambridge University Press.
- [22] Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- [23] Kolb, A. Y., & Kolb, A. D. (2005). The Kolb learning styles inventory- version 3.1: Technical specifications. Boston, MA: Hay Resources Direct.
- [24] Kolb, A. Y., & Kolb, D. A. (2008). Experiential learning theory: A dynamic, holistic approach to management learning, education and development. In S. J. Armstrong & C. Fukami (Eds.), *Handbook of management learning, education and development* (pp. 1-59). London: Sage Publications.
- [25] Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100. doi: 10.1177/0265532208097337.
- [26] Meier, G. (2010). Review of the assessment of the year abroad in the modern language degrees at Bath: Assessment for experiential and autonomous learning based on the continuity model. <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2> (accessed 3/15/2012).
- [27] Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and implications. *Language Testing*, 6(1), 1-13. doi: 10.1177/026553228900600103.
- [28] Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19(2), 109-131. doi: 10.1191/0265532202lt224oa.
- [29] Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, 21(1), 87-110.
- [30] Ross, S. (1998). Self-assessment in second language testing: A meta- analysis of experimental factors. *Language Testing*, 15(1), 1-19. doi: 10.1177/026553229801500101.
- [31] Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8(1), 31-54. doi: 10.1191/1362168804lr133oa.
- [32] Strong-Krause, D. (2000). Exploring the effectiveness of self-assessment strategies in ESL placement. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 49-74). New Jersey: Lawrence Erlbaum Associates, Inc.
- [33] Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276. doi: 10.3102/00346543068003249.

Parviz Birjandi is a full professor holding an MA in applied linguistics and a PhD in English education (Research methods and statistics) from the University of Colorado. He has published a large number of research articles and books and is also the author of English textbooks for high school and pre-university levels, used nationwide, and many university textbooks. His major interests concern assessment, research methodology, and first language acquisition.

Mina Saghaieh Bolghari is currently a Ph.D. candidate of TEFL at Tehran Islamic Azad University, Science and Research Branch. Her main areas of research interest include classroom-based assessment and foreign language teaching issues and methodologies.