Journal of
Language and Translation
Volume 11, Number 4, 2021 (pp. 79-103)

On the Legitimacy of Cooperative Assessment: Examining EFL Learners' Rating Accuracy in Assessing Cooperative Reading

Mina Saghaieh Bolghari[1], Parviz Birjandi[2*], Parviz Maftoon[3]

[1]PhD Candidate in TEFL, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Professor, Science and Research Branch, Islamic Azad University, Tehran, Iran

[3]Associate Professor, Science and Research Branch, Islamic Azad University, Tehran, Iran

Received: November 31, 2020 Accepted: April 20, 2021

Abstract

Cooperative learning has attracted ever-increasing research literature on language studies, among which EFL reading instruction abounds with numerous examples in favor of cooperative learning. However, there is a common belief that if cooperative learning were not assessed, students would be unlikely to take it seriously. Therefore, the main objective of this study was to examine EFL learners' accuracy in assessing cooperative reading. To this end, 60 sophomores majoring in English translation at Islamic Azad University, Tehran Central Branch, participated in this study. Data were collected in the course of 15 weeks. Each week, first, the participants cooperatively read an expository passage in 15 four-member groups in class. Then, each participant was asked to carry out an individual reading assignment at home. For the cooperative reading assessment, both the group performance and the individual homework assignment of each participant was weekly self-, peer-, and instructor-assessed. The results demonstrated the participants' reasonable accuracy in assessing cooperative reading. Furthermore, the qualitative findings obtained from a semi-structured focus group interview showed that the participants perceived cooperative reading assessment positively. The study may produce credible evidence for using cooperative assessment as a reasonably honest approach to evaluate EFL learners' reading performance.

Keywords: Cooperative assessment; Cooperative reading; Rating accuracy; Self- and peer-assessment

* [2]Corresponding Author's email:
birjandi@srbiau.ac.ir

INTRODUCTION

Cooperative learning (CL) has a noticeable effect on education in general and second language learning (Dörnyei, 1997). CL is "the instructional use of small groups so that students work together to maximize their own and one another's learning" (Johnson & Johnson, 1990, p.69). Yet, not every group work is CL. According to Roger and Johnson (1994), there is a difference between simply asking students to work in a group and structuring groups of students to work cooperatively. That is to say, in CL students work together to accomplish shared goals and demand beneficial outcomes to all members (Johnson & Johnson, 1999). In cooperative classrooms, according to Slavin (1995), "students are expected to help each other, to discuss and argue with each other, to assess each other's current knowledge and fill in gaps in each other's understanding" (p. 2). In a similar vein, Dörnyei (1997) refers to three key components of CL which make a learning approach cooperative. First, students spend most of the class time working in groups of three to six members. Second, learning is structured to ensure that their peers have also achieved the instructional goal. Third, both the group's achievement and individual achievement are equally assessed and rewarded.

While reviewing some empirical studies on CL, Falchikov (1993) enumerates some of the benefits of CL. For instance, CL promotes higher achievement in participants compared with that of competitive learners. Moreover, students working cooperatively use higher cognitive strategies for learning than do competitive peers. CL, furthermore, promotes positive attitude toward the subject studied and helps group members improve their interpersonal relations. Students' self-esteem can also be enhanced as a result of participation in cooperative groups.

Despite the potential benefits of CL, assessment of CL has received much less attention (Ballantine & Mccourt Larres, 2007). That is, CL is usually viewed as a means of learning not assessment, at least in EFL settings. The purpose of this study, accordingly, was to explore how CL would be assessed in an Iranian EFL learning context.

CL: Basic Elements

What distinguishes CL is the inclusion of five basic elements. Johnson and Johnson (1999) explain these elements as follows:

1. **Positive interdependence** is furthered simply because one cannot succeed unless other group members succeed. In other words, success depends on the participation and success of all group members.
2. **Individual accountability** is encouraged when the performance of each individual student is assessed, and the results are given back to the group and each of the individuals.
3. **Face-to-face promotive interaction** is promoted when individuals promote each other's success by helping, assisting, supporting, encouraging, and praising one another's effort.

4. **Social skills** are furthered because leadership, decision-making, trust-building, communication, and conflict management skills will be taught to group members.
5. **Group processing** is enhanced when group members discuss how well they achieve their goals and maintain effective relationships.

Group Assessment: An Approach to Assessing CL

Presumably, evaluation of the group work is not prevalent among teachers (Zhang, Johnston, & Kilic, 2008). Teachers may prefer assessing the students individually after learning collaboratively. Yet, group assessment would appear to be a reasonable way to evaluate students at the time when universities tend to train graduates to deal with real life challenges (Strauss, 2001). However, group assessment suffers from some limitations. If students know that the grade they will get is predominantly the outcome of other students' levels of effort, they do not take it seriously (Gibbs, 2009). Latane´ (1981), writing on the theory of social loafing, claims that the inherent nature of group work is that each individual member contributes to a group less than s/he would work alone. Another problem of group assessment is free-riding which Morris and Hayes describe (as cited in Ballantine & Mccourt Larres, 2007, p. 166) as "the problem of the non-performing group member who reaps the benefit of the accomplishment of the remaining group members with little or no cost to him/herself." Houldsworth and Matthews (2000) also describe a sucker effect in which the most hardworking student gradually reduces his effort in order to avoid being taken advantage of by the free-riders.

To overcome such problems, Falchikov (1986) suggests that group work should be assessed in terms of its process or product. The product of group work usually takes the form of a collective outcome albeit individual student assignments may also be submitted (Spatar, Penna, Mills, Kutija, & Cooke, 2015). The process of group work, according to Spatar et al., is each group member's contribution to group work to achieve shared outcomes. However, the extent to which students should be assessed for either the process of group work or its product is a controversial issue (Johnston & Miles, 2004). To Falchikov (1986, 1993), both the process and product of group learning should be assessed. She distinguishes between assessing the product (through instructor-assessment) and assessing the process (through peer- and self-assessment). As Saito and Fujita (2009) assert "the degree of cooperation in a group should be reflected in the extent to which the group members contributed to the final product by using an array of the group process skills at a moment in need" (p. 152).

Group assessment through self- and peer-assessment. Most methods proposed to assess individual contributions to group work, including CL, rely on self- and peer-assessment (Lejk, Wyvill, & Farrow, 1996). In group-assessment, self- and peer ratings provide a rich source of information to estimate individual contribution (Zhang et al., 2008). In self-

assessment, learners judge their own performance against the assessment criteria, and in peer-assessment they assess the peers' performance (Falchikov, 1986).

In the realm of group assessment, peer-assessment has gained much more interest. In theory, at least, according to Cheng and Warren (2000), peer-assessment allows both the process and product of learning to be assessed more fairly because it is the students themselves who have knowledge about group's activities. However, to Goldfinch (1994) peer-assessment can be more valid by supplementing it with self-assessment. Her assumption is that by including self-assessment, over-generous students in peer-assessment will probably inflate their own total as much as that of their peers.

According to Johnson and Johnson (1999), one way to ensure group dynamics and individual accountability is the assessment of both individual and group performance. By grading learners' within-group performance and individual outcome, either through self- and peer-assessment or instructor-assessment, students are provided with tangible recognition for their own work toward the group process and product, and they may be penalized for not contributing to the group (Davies, 2009; Gibbs, 2009; McWhaw, Schnackenberg, Sclater, & Abrami, 2003; Strauss, 2001).

Very few studies can be found on group assessment in EFL/ESL context. Cheng and Warren (2000), for instance, conducted a study with Chinese students studying English for Academic Purposes (EAP) to examine how individuals assessed the peers' contributions to a group project. The students carried out an integrated group project which was made up of three components: A seminar presentation, an oral presentation, and a written report. The assessment procedures included both teacher-assessment and peer-assessment. It was concluded that peer-assessment is effective to evaluate individual contributions to group work.

In another study, Saito and Fujita (2009) examined the relationship between group cooperation (process) and the quality of group presentations (product) on Japanese EFL students. Results showed that most group members could differentiate the degree of each member's contribution to the group project. In addition, the process of the group work was a statistically significant predictor of the product of the group work.

Self- and peer-assessment and reading. Research pertaining to self- and peer-assessment has primarily focused on their validity (Zhang et al., 2008). In this regard, the students' rating accuracy is concerned. Rating accuracy refers to the extent to which the student-raters are in close agreement with the criterion reference. The criterion reference can be the instructor' ratings, or the results obtained from standardized tests, e.g., TOEFL. To this end, research is devoted to find any correlation coefficient or draw a mean-score comparison between student ratings and the criterion reference. High agreements or non-significant differences between the criterion reference and the student-raters can put an accurate interpretation on the self- and peer-assessment results. That is, students can accurately assess

their own or their peers' performance, when their ratings are compared with those of the instructors'.

In the EFL/ESL reading literature, for instance, Ross's (1998) examined the validity of self-assessment on second language reading ability. As a result of his meta-analysis on self-assessment of four language skills, Ross reported that the largest number of correlations were recorded for the reading skill. Shokri (2015) also investigated the relationship between students' reading self-assessment, teacher-assessment, and their final exam scores. The results revealed a significant relationship between the intermediate learners' self-ratings and teacher assessments. The same results, however, were not achieved for the elementary students. In another study, Tepsuriwong and Bunsom (2013) designed a rubric for peer-assessment purposes of Thai ESL students assessing their peers' reading performance. The findings surprisingly showed that the correlations between teacher-assessment and peer-assessment with the designed rubric was lower than that of the peer-assessment without the rubric. That is, the students accurately rated their peers' reading performance when the yardstick was the teacher-assessment without any objective criteria.

Purpose of the Study

Affected by the paradigm shift in language assessment, reading assessment has witnessed some innovative approaches to evaluate the students' reading skill. Grabe (2009), for instance, introduces project-performance evaluation as a "newer task format that evaluates test takers as they read texts and then perform in groups to carry out a larger project" (p. 360). To Grabe, group assessment in reading is an interesting alternative but problematic on several validity grounds--giving individual scores based on group interactions and a holistic task. Similarly, Johnston and Miles (2004) assert that group assessment is a complicated task because it is difficult to decide on the individual contributions to a group and report numerical values.

Furthermore, the problem of freeloading in group assessment increases when group members work collaboratively without any individualized task, i.e., there is no division of labor (George, 1992). Accordingly, group assessment within a cooperative learning environment--cooperative assessment--may produce more reliable results since each group member carries out a distinctive task which can be separately assessed.

In the present study, cooperative assessment is defined as one type of formative group assessment, through which each group member's cooperative performance, as well as individual accountability, is weekly assessed by the instructor, self, and peer via related rubrics. Put it another way, both the process and product of cooperative learning are assessed. In this respect, the process of cooperative reading refers to EFL learners' cooperation, division of labor, and collaboration to comprehend readings in a four-member group. The product of cooperative reading refers to four reading tasks, namely summarizing, outlining, graphic organizing, and self-questioning weekly

assigned tasks carried out individually. In other words, the process of cooperative reading includes in-class group participation and cooperative performance, while the products were homework individual assignments. Figure 1 depicts the features of cooperative reading assessment, defined in this study.

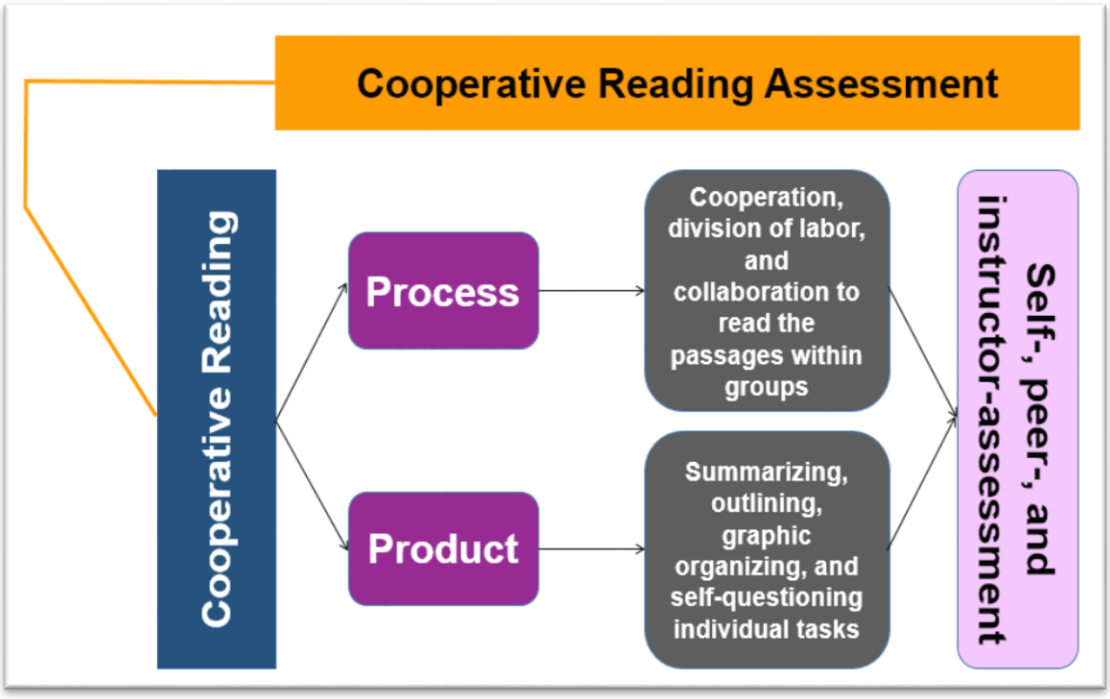

Figure 1. Features of Cooperative Reading Assessment.

Cooperative assessment would be different from previous studies (e.g., Falchikov, 1986, 1993; Johnston & Miles, 2004) in that it focuses on both within-group performance and individual learning, each is assessed by the instructor, self, and peers. In addition, it is a formative assessment, through which weekly ratings and comments would be reported to the students. Thus, they can learn from the instructor's feedback and improve their reading performance.

Notwithstanding the popularity of self-assessment and peer-assessment in EFL studies, group assessment appears to arouse less interest. More specifically, no study has so far, to the best of the researchers' knowledge, been carried out on cooperative reading assessment in Iranian EFL settings although there are ample pieces of evidence to support the positive effect of CL on Iranian EFL students' reading comprehension (e.g., Azizinezhad, Hashemi, & Darvishi, 2013; Jalilifar, 2010; Marzban & Alinejad, 2014). The present study, consequently, aimed at examining Iranian EFL learners' rating accuracy in assessing cooperative reading.

Learners' perception toward cooperative assessment was also determined to cross-validate the quantitative data. The following

questions were raised to fulfill the research purpose:

Q_1: *Is there any statistically significant relationship between self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading products?*

Q_2: *Is there any statistically significant difference between the ratings of self-, peer-, and instructor-raters assessing Iranian EFL learners' cooperative reading products?*

Q_3: *Is there any statistically significant relationship between self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading process?*

Q_4: *Is there any statistically significant difference between the ratings of self-, peer-, and instructor-raters assessing Iranian EFL learners' cooperative reading process?*

Q_5: *How do Iranian EFL learners evaluate cooperative assessment?*

METHODS

Participants

Sixty Iranian sophomores, aged between 19 and 24, participated in this study. The participants included 14 male and 46 female EFL undergraduates studying at Islamic Azad University, Tehran Central Branch. They attended Reading Comprehension III class which was a four-credit course. Therefore, during the semester the participants received 32 sessions of instruction, each of which lasted for 90 minutes.

The participants were selected based on their performance on the Oxford Placement Test (OPT) and the Reid's (1987) perceptual learning style preference questionnaire. On the basis of the OPT results, only those students whose scores fell between one standard deviation above or below the mean were chosen. Moreover, based on the results of the perceptual learning style questionnaire, those students who did not favor group learning were excluded from the research population. Due to the university registration norms, it was practically impossible for the researchers to disrupt the schedules by randomizing the participants. Thus, an intact group was used as the research sample.

Instruments

OPT. The OPT was used in the present study as an English language proficiency test. The test has two main sections, each with 100 items. Listening skills, grammar, vocabulary, and reading skills are targeted at the OPT.

Perceptual Learning Style Preference Questionnaire. Since reading instruction was geared to group learning, it was critical to identify the students who did not prefer group work. To do so, Reid's (1987) learning style questionnaire was used. Reid has classified perceptual learning styles into six types: visual, auditory, kinesthetic, tactile, individual, and group, the last of which was taken into account in the current study.

The results obtained from the questionnaire were classified into major, minor, and negligible preferences. Major preference is a preferred leaning style, minor preference is one

in which learners can still function well, and negligible preference is the one that makes the learning difficult for students (Reid, 1987). For the purpose of this study, six students with negligible group learning style were discarded.

Reading passages. The main reading materials included 12 expository passages taken from electronic magazines and the Internet pages. The researchers benefited from Nuttall' (1996) criteria for choosing EFL/ESL reading materials. For piloting the passages, factors such as readability (13 as measured by Fry formula), length (1400 words on the average), authenticity, date of publication, American English style, rhetorical organization of the passages, and learners' topic of interest were taken into account.

Rating scales for assessing the reading performance. Both the process and product of cooperative reading performance were assessed using appropriate rubrics--having been validated in a forthcoming study by the same researchers. The researchers developed four rating sub-scales to assess the products of cooperative reading--summarizing, outlining, graphic organizing, and self-questioning tasks. Each sub-scale included three components that would be assessed according to the four levels of performance, namely, very good (4), good (3), fair (2), and poor (1), ranging in numerical value from 4 to 1. Therefore, each student could obtain a maximum total score of12. The researchers validated the scales through running factor analysis.

In addition, a five-point Likert scale was designed based on five basic elements of CL (Johnson & Johnson, 1990) to measure the process of cooperative reading. Accordingly, the maximum total score obtained through the scale was 20. After being developed, both scales were used for the purpose of instructor-, self-, and peer-assessment during the treatment.

Semi-structured focus group interview. In order to gather the qualitative data from the participants, a semi-structured focus-group interview was conducted with 20 students after the treatment. In other words, five four-member groups were interviewed by the instructor/researcher separately. The interview questions aimed at eliciting the students' evaluation of cooperative assessment. The students were asked whether they were satisfied with their group' cooperative performance, self- and peer-assessment, instructor's evaluation criteria, and so forth. Each group's interview was separately audio-recorded to be carefully analyzed.

Data Collection Procedures

The current study was a mixed-methods research collecting and analyzing both quantitative and qualitative data through an embedded design. Data were collected in regular class time and over a period of 15 weeks. To this end, the following procedures were pursued:

Homogenizing the participants. Before the onset of the study, the OPT was administered to

all participants to homogenize their English language proficiency. Then, the Reid' (1987) learning style preference questionnaire was given to the participants to cross out those students who did not prefer working in a group.

Training cooperative assessment. At the beginning of the semester, the instructor, the first author of this article, thoroughly explained the principles of CL serving as a basis for reading instruction. After being familiar with cooperative reading, the participants were asked to form four-member groups based on their own preferences.

Later, the instructor modeled the four reading tasks defined as the products of cooperative reading. These tasks were chosen from a list of effective reading strategies, compiled by Grabe (2009, p. 209), which have been empirically supported by numerous studies (e.g., Brown & Day, 1983; Klingner & Vaughn, 1999; Palincsar & Brown, 1984).

Then, the participants received training in self-assessment and peer-assessment for 90 minutes. They were briefly instructed to use the intended scales for the purpose of self- and peer-assessment. Then, in order to clearly establish the assessment criteria, the instructor provided the participants with some rated samples shown via a video projector. After becoming familiar with the reading scales, the groups practiced assessing four student samples together in class. Having rated the samples, the participants compared their own scores with those of the instructor to come to an agreement on the ratings.

Assessing cooperative reading. After receiving the two training weeks, the participants read 12 unseen passages cooperatively within their own groups for 12 sessions. All passages should be read paragraph by paragraph by the group members who were required to carry out four assigned tasks, including clarifying any unclear points of a paragraph, identifying the paragraph structure and transition markers, choosing the best title for a paragraph, and restating the gist. In other words, each group member took a different role. The group members should assist each other in achieving the intended outcome--comprehending the passage. To this end, the members shared their findings with the group and received their feedback.

The seats in the class were arranged in such a way that the group members could easily interact with one another. The instructor carefully observed the groups' cooperative performance using the designed scale to rate the participants' level of group contribution, i.e., the process of cooperative reading in class. Similarly, the group members were required to rate the peers' contributions to the group via the identical scale. The results of instructor, self-, and peer-assessment of group performance were reported to the individuals the following week.

Having read the passages in class, the participants were required to do their weekly assignments at home and submitted them to the instructor the subsequent week. That is, the products of cooperative reading were prepared. To do this, the members of a group shared the four tasks of summarizing, outlining, graphic

organizing, and self-questioning among themselves and exchanged them every week. In other words, each group member did a different task each session. During 12 sessions, every student carried out the four tasks three times. Division and rotation of the assigned tasks were checked by the group leader, chosen by the instructor. After doing the assignment, each student sent his/her own piece of work to the group members through either email or any smart phone messaging services. The participants then assessed their own, as well as their group members' assignments, using the same rating scales. Each week, the instructor assessed the students' individual assignments and brought back them to the learners so that they could receive the instructor' ratings and comments on their own piece of work. In addition, an external rater, who was a Ph.D. candidate with previous experience of teaching reading at universities, was asked to rate the student samples in order to achieve the reliable results.

Conducting the semi-structured focus group interview. After the intervention, five group of students were randomly invited for the focus group interview. Each group was interviewed for approximately 15-20 minutes by the instructor who asked the questions in English, but the participants could answer them in either English or Persian.

RESULTS

Quantitative data analysis. Before the null hypotheses were tested, reliability analyses were computed for estimating intra-rater and inter-rater consistency. Table 1 represents inter-rater reliability coefficients for the four intended reading tasks--cooperative reading products. The results indicate substantially great consistency between the two raters for all tasks. The first rater, the instructor, randomly selected 20 samples out of a pool of 60 and rated them once again after the treatment. Correlation coefficients were used between the two set of scores to determine her intra-rater consistency for each task (Tables 2). The results indicate highly acceptable reliability values confirming intra-rater consistency.

Table 1

Inter-Rater Reliability Coefficients for Cooperative Reading Products

	Pearson Correlation	**Sig.**
Summarizing	.959**	.000
Outlining	.974**	.000
Graphic Organizing	.971**	.000
Self-questioning	.962**	.000

**. Correlation is significant at the 0.01 level.

Table 2

Intra-Rater Reliability Coefficients for Cooperative Reading Products

	Pearson Correlation	Sig.
Summarizing	.985**	.000
Outlining	.970**	.000
Graphic Organizing	.937**	.000
Self-questioning	.947**	.000

**. Correlation is significant at the 0.01 level.

Testing the first null hypothesis. The numerical data for answering the first research question came from the average scores of each reading task given by the two raters and the participants themselves across 12 sessions. The following null hypothesis was formulated to this end:

H_{01}: There is not any statistically significant relationship between self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading products.

The hypothesis was tested using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity for all the correlation coefficients. Three average scores were calculated for each participant for self-assessment, mean of peer-assessment, and mean of instructor-assessment, respectively. These scores are obtained from the designed rating scales, which were identical for both the participants and the instructors.

As Table 3 shows, there is a relatively strong and positive relationship between instructor-assessment and self-assessment, $r= 0.70$, $p < 0.01$, because high scores of self-assessment associate with high scores of instructor-assessments. There is a strong and positive association between instructor-assessment and peer-assessment in a similar vein, $r= 0.73$, $p < 0.01$. The correlation between self- and peer-assessment is also positive and acceptable ($r= 0.60$, $p < 0.01$). To Pallant (2013), the strength of the relationship between 0.50 and 1.0 is large, indicating a strong relationship between the two variables. Therefore, the first research null hypothesis was rejected, suggesting a statistically significant relationship between self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading products.

Table 3

Correlations Coefficients between the Instructor-, Self-, and Peer-assessment on the Products of Cooperative Reading

		Instructor	Self	Peer
Instructor	Pearson Correlation	1	.707**	.736**
	Sig. (2-tailed)		.000	.000
	N	60	60	60
Self	Pearson Correlation	.707**	1	.601**
	Sig. (2-tailed)	.000		.000
	N	60	60	60
Peer	Pearson Correlation	.736**	.601**	1
	Sig. (2-tailed)	.000	.000	
	N	60	60	60

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

Testing the second null hypothesis. The mean scores of the student-raters were compared with those of the instructor-raters to confirm the participants' rating accuracy in cooperative reading assessment. Regarding the product of cooperative reading, the following null hypothesis was formulated:

H_{02}: There is not any statistically significant difference between the ratings of self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading product.

The hypothesis was tested using One-way ANOVA since the data was typically distributed based on Kolmogorov-Smirnov and Shapiro-Wilk tests results ($p < .05$). In addition, the results of Levene's test showed that the assumption of homogeneity of variances was not violated. The results of descriptive statistics (Table 4) show that the students overrated their own, as well as their peers', reading task performance comparing to the instructor-assessment as the criterion measurement. However, One-way ANOVA was used to check if there was any statistically significant difference among the ratings of the three group of raters (Table 5). As Table 5 demonstrates, there is not any statistically significant difference at the $p< .05$ level in the reading task scores for the three groups when the effect size is also too small (Eta squared=.03). Hence, the second null hypothesis was confirmed.

Table 4

Descriptive Statistics for the Products of Cooperative Reading Assessment

	Minimum	Maximum	Mean	Std. Deviation
Instructor	2.83	11.20	8.1688	1.76896
Self	5.91	10.66	9.0797	1.97726
Peer	6.58	11.41	9.8945	1.14731
Valid N (listwise)				

Table 5

One way-ANOVA for the Products of Cooperative Reading Assessment

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.709	2	4.854	3.056	.060
Within Groups	281.174	177	1.589		
Total	290.882	179			

Testing the third null hypothesis. The numerical data for answering the third research question came from the participants' cooperative reading performance while contributing to the groups in class. The related null hypothesis was formulated as follows:

H_{03}: There is not any statistically significant relationship between the self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading process.

The relationship between self-, peer-, and instructor-assessment in rating the participants' cooperative reading performance was investigated using Pearson product-moment correlation coefficient (Table 6). The instructor's classroom observation of each group performance and field notes were used as the criteria reference. For each participant, the average scores of self-, peer-, and instructor-assessment were calculated.

Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. As Table 6 shows, there is a strong and positive correlation between the instructor and self-ratings, $r= 0.81$, $p < 0.01$. Similarly, there is a positive correlation between the peers' and the instructor' ratings, $r= 0.91$, $p < 0.01$. The correlation between self- and peer-assessment is also relatively strong and positive ($r= 0.72$, $p < 0.01$). As a result, the third research null hypothesis was rejected, suggesting significant relationships between the self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading process.

Table 6

Correlations Coefficients between the Instructor-, Self-, and Peer-assessment on the Process of Cooperative Reading

		Instructor	Self	Peer
Instructor	Pearson Correlation	1	.811**	.914**
	Sig. (2-tailed)		.000	.000
	N	60	60	60
Self	Pearson Correlation	.811**	1	.727**
	Sig. (2-tailed)	.000		.000
	N	60	60	60
Peer	Pearson Correlation	.914**	.727**	1
	Sig. (2-tailed)	.000	.000	
	N	60	60	60

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

Testing the fourth null hypothesis. One-way ANOVA was run to examine the participants' rating accuracy in self- and peer-assessment of cooperative reading process. The following null hypothesis was formulated:

H_{04}: there is not any statistically significant difference between the ratings of self-, peer-, and instructor-assessment of Iranian EFL learners' cooperative reading process.

The hypothesis was tested using One-way ANOVA because tests of normality indicated that the data was normally distributed. The Levene's test also confirmed the homogeneity of variances. The results of the descriptive statistics (Table 7) show that the participants overestimated their own cooperative reading performance (M=17.77). However, the peers' performance was underestimated (M=15.45). The results of one-way ANOVA (Table 8) also indicate that there is statistically significant difference between the mean scores of self- and peer-assessment and that of the instructor-assessment in assessing the process of cooperative reading, $F(2, 177) = 3.848$, $p < 0.1$. Despite reaching statistical significance, the actual difference in mean scores among the groups is relatively small because the effect size, calculated via eta squared, is 0.04.

Table 7

Descriptive Statistics for the Process of Cooperative Reading Assessment

	Minimum	Maximum	Mean	Std. Deviation
Instructor	8.90	19.40	16.1467	2.58670
Self	16.20	20.00	17.7750	1.06064
Peer	6.90	19.50	15.4500	2.70183
Valid N (listwise)				

Table 8

One-way ANOVA for the Process of Cooperative Reading Assessment

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	73.314	2	36.657	3.848	.023
Within Groups	1686.295	177	9.527		
Total	1759.610	179			

Table 9

Multiple Comparisons: Tukey HSD

(I) rater groups	(J) rater groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
IA	SA	-.88667	.56353	.260	-2.2186	.4453
	PA	.67167	.56353	.460	-.6603	2.0036
SA	IA	.88667	.56353	.260	-.4453	2.2186
	PA	1.55833*	.56353	.017	.2264	2.8903
PA	IA	-.67167	.56353	.460	-2.0036	.6603
	SA	-1.55833*	.56353	.017	-2.8903	-.2264

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

Post-hoc comparisons using the Tukey HDS test (Table 9) indicate that there are no statistically significant differences between the mean scores of the instructor-raters and either self- or peer-raters. However, the mean score for the group of self-raters (M= 17.77, SD= 1.06) is different from the group of peer-raters (M= 15.45, SD= 2.70). In other words, the

participants rated their own performance differently when peer-assessment was the yardstick. Therefore, the fourth null hypothesis was rejected.

Qualitative data analysis. The purpose of last research question was to determine how the participants evaluated cooperative assessment. In order to answer this question, the data were collected via a semi-structured focus group interview. First, the gathered data were transcribed. Then, a content analysis was performed in order to identify some iterative ideas. The analyzed data were finally grouped into five general themes:

Individuals' contributions. The participants were asked to evaluate their own and their group members' contributions to the group. Almost all students (91%) were satisfied with their own performance within the group. They believed they did their best to perform cooperative roles and help the group members comprehend the passages. Similarly, most of them (82%) were pleased with their group members' contributions to the group. They positively rated the members' constructive comments and feedback. Some representative extractions are: "I used my friends' comments to correct my assignments"; "The members encouraged me to do my best when we were reading the passages in class"; "I had very nice and supportive friends in my group."

Process of cooperative reading. The participants' satisfaction with the assessment on the process of cooperative reading was also determined. Nearly all students (94%) agreed that cooperative assessment was highly effective. However, they preferred instructor- and peer-assessment since observing and rating others' performances could be easier than evaluating one's own cooperative performance. In addition, the participants reported that self-assessment was not the indispensable part of the cooperative assessment. One student said "self-assessment seemed stupid to me because I thought I did my best, and, of course, I should give high grades to myself."

Products of cooperative reading. The students were also asked whether they were satisfied with cooperative assessment on the assigned reading tasks. Almost all students (95%) believed that instructor-rating was the most effective approach to evaluate their reading task performance. They enjoyed receiving instructor's ratings and comments to modify their assigned tasks.

As regards peer-assessment, 19% of the participants remarked that it could be invaluable if the ratings were consistently reported to the peers. In their views, secret scores excited their curiosity to grasp what their friends reflected on their reading performance. However, majority of the respondents (81%) expressed a strong preference for unreported ratings. "It could affect me negatively if I received low grades from my peers," said one student. "I felt embarrassed to give real marks to my friends, so I could not be honest with them"; "It is meaningless because you only try to please them with good grades," said other students.

Although most students (71%) reported that self-assessment of reading tasks was difficult for them, 78% of the respondents referred to the utility of self- assessment as a tool for noticing their mistakes. It is evident in the following excerpts: "I used self-assessment as a checklist to find my errors"; "I realize how much I have learned"; "I found out that I should improve my reading skills."

Assessment criteria. Eighty-nine percent of those interviewed expressed that they were satisfied with the instructor's evaluation criteria, through which both their cooperative reading performance and individual accountability were assessed by the instructor, peers, and the students themselves. Only a small number of participants commented that more emphasis should be placed on the individual assignments, while a few others believed that the central focus should be on the in-class process of group work activities.

Group formation. Finally, none of the students showed negative perception of self-selection approach to form cooperative groups. When being interviewed, one student mentioned: "If you selected my group members, I would not enjoy working with the group. I prefer to work with my friends because we are responsible for each other's learning." In addition, most students thought that they did not like mixed ability groups because it could be a laborious task to work with low-achievers. They also did not prefer working with high achievers who may dominate group work. "I cannot accept the comments of a member who thinks she knows the best," said one male student.

DISCUSSIONS

The current study aimed at examining Iranian EFL learners' accuracy in assessing cooperative reading. The quantitative findings revealed that both the process and product of cooperative reading were accurately assessed by the participants, using self- and peer-assessment. A partial explanation for this may lie in the fact that self-assessment gave students practice in evaluating their own work. Possibly, peer-assessment of the reading tasks gave them the opportunity to review and learn from other students' work. That they found peer-assessment of the group contributions highly effective was probably owing to the fact that they considered the evaluative function of peer-assessment, and, so, they were sensitive to report accurate results. In return, receiving group members' informal comments, rather than numerical ratings allowed the students to learn about their own achievement as an individual learner and a group member.

The results are in line with a large body of research, reporting close agreement between self-assessment and the criterion reference (e.g., Bachman & Palmer, 1989; Falchikov & Boud, 1989; Ross, 1998; Shokri, 2015; Stefani, 1994; Williams, 1992). Regarding peer-assessment, the present research also corroborates the findings of several studies, suggesting high agreement between peer-assessment and that of the criterion measurement (e.g., Falchikov & Goldfinch,

2000; Freeman, 1995; Patri, 2002; Tepsuriwong & Bunsom, 2013). The results, however, do not support Paleczek, Seifert, Schwab, and Gasteiger-Klicpera's (2015) findings, which showed lower correlations between young learners' self-assessment scores and the objective reading test scores. This might be due to the differences between the two research samples regarding age and level of English proficiency.

Concerning the significant difference between self-rater and peer-raters, the results of the present study shared some similarities with Johnston and Miles's (2004) findings in that their research population rated their own contributions significantly higher than those of their peers. The findings also indicated that the participants were consistent with the instructor-raters in terms of assessing intra-group performance. Unlike the current research, in Johnston and Miles's study there was no significant correlation between self-ratings and peer-ratings.

Furthermore, the qualitative data analysis showed that the students evaluated the efficacy of cooperative assessment positively. Similar to the results reported by Ballantine and Mccourt Larres (2007), the evidence found in the current study also pointed to the participants' high level of satisfaction with group assessment.

The findings are in complete agreement with other studies, showing the positive perception of students toward self- and peer-assessment (e.g., Bachelor, 2017; Gatfield, 1999; Muñoz & Alvarez, 2007; Woolhouse, 1999). In general, however, the participants favored instructor-assessment of both the process and product of cooperative reading more positively comparing to self- and peer-assessment. This matches well with Ozogül and Sullivan's (2009) study in which the students' overall attitude was significantly more positive toward teacher-assessment than toward peer-assessment.

Unlike Praver, Rouault, and Eidswick's (2011) study in which students felt that peer-ratings with comments were more useful than only numerical assessment, in this study the participants reported that only receiving peer-comments were sufficient and efficient. These findings refute previous results reported by Cheng and Warren (2005) who found that while a majority of tertiary students in Hong Kong considered peer-assessment favorably, a minority perceived themselves or other students unable to provide fair evaluations.

In this study, although not rejecting the existence of bias in assessment, the participants believed that peer-assessment was a reliable approach to evaluate peers' contributions to the group. This aligns with Gatfield's (1999) study in which the students commented that peers can assess each other fairly.

The results also indicated that the participants favored the instructor's evaluation criteria which focus on both the process and product of cooperative reading. Like the participants of Li and Campbell's (2008) study, the participants of the present research held negative views about group assignments when an overall group grade was equally given to all members. The participants concluded that only when each individual's contrition to the group

was assessed, would the group assessment be reliable.

Finally, it is worth-mentioning that certain limitations were imposed on this study, which constrained the researchers from drawing robust conclusions and limited the generalizability of the results. Firstly, the small sample size provided only a limited amount of data, and, therefore, the findings cannot be generalizable. Secondly, the participants could not be randomly selected. The researcher/writer, as a result, had to resort to the use of an intact group. Finally, both classes included the students majoring in English translation, making the generalizability of the findings more restricted.

CONCLUSION

The finding of the study indicated that the students accurately assessed their own, as well as their peers', reading performance. Hence, taking consensus as an indicator of accuracy, the researchers can conclude that cooperative assessment was a valid tool for assessing EFL learners' cooperative reading performance. The participants' favorable attitude toward cooperative assessment also confirms the findings.

The results of this study may have some pedagogical implications for language teachers. Managed correctly, group-assessment can contribute to formal classroom assessment if students are provided with clear measurement criteria, adequate training and practice, and constructive feedback. Teachers can also benefit from the findings of this study to consider learners' attitude toward self- and peer-assessment using opinionnaires or conducting interviews. According to Patri (2002), individual interviews after self- or peer-assessment can be helpful for both teachers and students to understand the psychological factors involved in learners' tendencies to over- or under-estimate their performance.

The results also foreground the importance of considering teacher-assessment as a criterion reference for assessing the process of group work, an uncommon case in the realm of group assessment. To this end, teachers should carefully observe and monitor group interactions and dynamics using their field notes and suitable checklists. Teachers should also consider group members' ratings and comments given to one another.

In the group assessment literature, several statistical approaches have been proposed to calculate individuals' contributions to group projects (Lejk et al., 1996). As a theoretical implication, the following simple formula is proposed for summative purposes to estimate a cooperative performance (CP) score for an individual in an EFL classroom:

$$CP = \frac{\overset{\text{Product}}{\left(\frac{2\overline{TA}+\overline{SA}+\overline{PA}}{}\right)} + \overset{\text{Process}}{\left(\overline{TA} \text{ of } \frac{IC}{\sum GC}(\overline{SA}+\overline{PA})\right)}}{2}$$

where TA stands for the mean of teacher-assessment, SA is the mean score of self-assessment, and PA denotes the mean score of peer-assessment. In addition, IC stands for individual contribution to the group, while ∑GC is sum of the group members' contributions to the group. It is the instructor's choice to consider either equal or weighted proportions for the instructor-ratings and those of the student-ratings. In this study, the researcher doubled the average score of the instructor-assessment since it was the criterion measurement and assumed to be less biased than student-assessment.

Concerning the results, further research can be undertaken in some areas. For instance, correlation coefficients were calculated in the current study as pieces of evidence to support the criterion validity of cooperative assessment. The present researchers suggest using Generalizability Theory framework to evaluate the reliability of cooperative assessment as well. In addition, the same research can be replicated to examine whether there is a difference between male and female students' rating accuracy in assessing cooperative reading. Finally, in the current study the groups were formed via the self-selection approach. Further studies are needed to examine if random and/or teacher selection approaches can affect EFL students' performance on cooperative activities. It is hoped that the current research can serve as a base for future studies on group assessment in EFL contexts.

References

Azizinezhad, M., Hashemi, M., & Darvishi, S. (2013). Application of cooperative learning in EFL classes to enhance the students' language learning. *Procedia-Social and Behavioral Sciences, 93,* 138-141. doi:10.1016/j.sbspro.2013.09.166

Bachelor, R. B. (2017). Alternative assessments and student perceptions in the world language classroom. *The TFLTA Journal, 6,* 30-44.Retrieved from https://files.eric.ed.gov/

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

Ballantine, J., & Mccourt Larres, P. (2007). Final year accounting undergraduates' attitudes to group assessment and the role of learning logs. *Accounting Education: An International Journal, 16*(2), 163-183. doi: 10.1080/09639280701234419

Brown, A. L., & Day, J. D. (1983). Macr orules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1-14. Retrieved from https://www.ideals.illinois.edu/bitstream/

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education, 5*(2), 243-255. doi: 10.1080/135625100114885

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93-121. doi: 10.1191/0265532205lt298oa

Davies, W. M. (2009). Groupwork as a form of assessment: Common problems and recommended solutions. *Higher Education, 58*(4), 563-584. doi: 10.1007/s10734-009-9216-y

Dörnyei, Z. (1997). Psychological processes in cooperative language learning: Group dynamics and motivation. *The Modern Language Journal, 81*(4), 482-493. doi: 10.1111/j.1540-4781. 1997. tb055 15.x/

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education, 11*(2), 146-166. doi: 10.1080/0260293860110206

Falchikov, N. (1993). Group process analysis: Self and peer assessment of working together in a group. *Educational and Training Technology, 30*(3), 275-284. doi: 10.1080/0954730930300308

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

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322. doi: 10.3102/00346543070003287

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education, 20*(3), 289-300. doi: 10. 108 0/0 26029 3950 200305

Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evaluation in Higher Education, 24*(4), 365-377. doi: 10.1080/0260293990 240401

George, J. M. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal, 35*(1), 191-202. doi: 10.2307/256478

Gibbs, G. (2009). The assessment of group work: lessons from the literature. *Assessment Standards Knowledge Exchange*, 1-17.Retrieved from http://www.brookes.ac.uk/aske

Goldfinch, J. (1994). Further developments in peer assessment of group projects. *Assessment & Evaluation in Higher Education, 19*(1), 29-35. doi: 10.1080/0260293940190103

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: : CUP.

Houldsworth, C., & Mathews, B. P. (2000). Group composition, performance, and educational attainment. *Education+ Training, 42*(1), 40-53. doi: 10.1108/00400910010317086

Jalilifar, A. (2010). The effect of cooperative learning techniques on college students

reading comprehension. *System, 38*, 96-108. doi:10.1016/ 2009.12.009

Johnson, D. W., & Johnson, R. (1990). What is cooperative learning? . In M. Brubacher, R. Payne, & K. Rickett (Eds.), *Perspectives on small group learning: Theory and practice* (pp. 68-79). Ontario, Canada: Rubicon.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into Practice, 38*(2), 67-73. doi: 10.1080/00405849909543834

Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education, 29*(6), 751-768. doi: 10.1080/0260293042000227272

Klingner, J. K., & Vaughn, S. (1999). Promoting reading comprehension, content learning, and English acquisition though collaborative strategic reading (CSR). *The Reading Teacher, 52*(7), 738-747. Retrieved fromhttp://www.jstor.org/stable/20204676

Latané, B. (1981). The psychology of social impact. *American psychologist, 36*(4), 343-356. doi: 10.1037/0003-066X.36.4.343

Lejk, M., Wyvill, M., & Farrow, S. (1996). A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education, 21*(3), 267-280. doi: 10.1080/0260293960210306

Li, M., & Campbell, J. (2008). Asian students' perceptions of group work and group assignments in a New Zealand tertiary institution. *Intercultural Education, 19*(3), 203-216. doi: 10. 1080 /026 0293032000066191

Marzban, A., & Alinejad, F. (2014). The effect of cooperative learning on reading comprehension of Iranian EFL learners. *Procedia-Social and Behavioral Sciences, 116*, 3744-3748.doi:10.1016/j.sbspro.2014.01.834

McWhaw, K., Schnackenberg, H., Sclater, J., & Abrami, P. C. (2003). From co-operation to collaboration: Helping students become collaborative learners. In R. M. Gillies & A. F. Ashman (Eds.), *Co-operative learning: The social and intellectual outcomes of learning in groups* (pp. 54-68). New York, NY: Routledge.

Muñoz, A., & Alvarez, M. E. (2007). Students' objectivity and perception of self-assessment in an EFL classroom. *The Journal of Asia TEFL, 4*(2), 1-25.Retrieved from http://www .academia.edu

Nuttall, C. (1996). *Teaching reading skills in a foreign language* (2nd ed.). Oxford: Macmillan Heinemann.

Ozogul, G., & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self and peer evaluators. *Educational Technology Research and Development, 57*(3), 393-410. doi: 10.1007/s11423-007-9052-7

Paleczek, L., Seifert, S., Schwab, S., & Gasteiger-Klicpera, B. (2015). Assessing reading and spelling abilities from three different angles– correlations between test scores, teachers' assessment and children's self-assessments in L1 and L2 children. *Procedia-Social and Behavioral Sciences, 174*, 2200-2210. doi: 1016/j.sbspro.2015.01.876

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175. doi:10.1207/s1532690xci0102

Pallant, J. (2013). *SPSS survival manual* (5th ed.). New York, NY: : McGraw-Hill Education.

Patri, M. (2002). The influence of peer feedback on self-and peer-assessment of oral skills. *Language Testing, 19*(2), 109-131. doi: 10.1191/0265532202 lt224oa

Praver, M., Rouault, G., & Eidswick, J. (2011). Attitudes and affect toward peer evaluation in EFL reading circles. *The Reading Matrix, 11*(2), 89-101.Retrieved from http://www.readingmatrix.com/

Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL Quarterly, 21*(1), 87-110. doi: 10.2307/3586356/full

Roger, T., & Johnson, D. W. (1994). An overview of cooperative learning. In J. Thousand, A. Villa, & A. Nevin (Eds.), *Creativity and collaborative learning* (pp. 1-21). Baltimore, Maryland: Brookes Press.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15*(1), 1-20. doi: 10.1177/026553229801500101

Saito, H., & Fujita, T. (2009). Peer-assessing peers' contribution to EFL group presentations. *RELC Journal, 40*(2), 149-171. doi: 10. 1177/0033688209 105868

Shokri, M. (2015). On the accuracy of Iranian EFL students' reading self-assessment and their level of reading proficiency. *International Journal of English Language & Translation Studies, 3*(3), 159-168.

Slavin, R.E. (1995). *Cooperative learning*: *Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Spatar, C., Penna, N., Mills, H., Kutija, V., & Cooke, M. (2015). A robust approach for mapping group marks to individual marks using peer assessment. *Assessment & Evaluation in Higher Education, 40*(3), 371-389. doi: 10.1080/02602938.2014.917270

Stefani, L. A. (1994). Peer, self, and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75. doi: 10. 1080/030750794123 31382153

Strauss, P. (2001). 'I'd rather vomit up a live hedgehog'-L2 students and group assessment in mainstream university

programs. *Prospect, 16*(2), 55-66.Retrieved from http://www.researchonline.mq.edu.au

Tepsuriwong, S., & Bunsom, T. (2013). Introducing peer assessment to a reading classroom: Expanding Thai University students' learning boundaries beyond alphabetical symbols. *Mediterranean Journal of Social Sciences, 4*(14), 279-286.Retrieved from http://www.mcser.org

Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education, 17*(1), 45-58. doi: 10.1080/0260293920170105

Woolhouse, M. (1999). Peer assessment: The participants' perception of two activities on a further education teacher education course. *Journal of Further and Higher Education, 23*(2), 211-219. doi: 10.1080/0309877990230205

Zhang, B., Johnston, L., & Kilic, G. B. (2008). Assessing the reliability of self-and peer rating in student group work. *Assessment & Evaluation in Higher Education, 33*(3), 329-340. doi: 10.1080/02602930701293181.

Biodata

Mina Saghaieh Bolghari is a PhD candidate in TEFL at Islamic Azad University, Science and Research Branch, Tehran. Over the past few years, she has been teaching several academic courses at IAU. Her primary ELT research interests are classroom-based assessment, activity theory, and cooperative learning.
Email: m.s.bolghari@gmail.com

Professor Parviz Birjandi is a full professor of applied linguistics at Islamic Azad University, Science and Research Branch. His primary research interests include language testing and assessment, research methodology, and first language acquisition. He has published numerous articles and authored many English textbooks for high school and university levels.
Email: birjandi@srbiau.ac.ir

Dr Parviz Maftoon is an associate professor of applied linguistics at Islamic Azad University, Science and Research Branch, Tehran, Iran. His primary research interests concern second language acquisition, SL/FL language teaching methodologies, and language curriculum development. He has published nationally and internationally and written and edited numerous English books.
Email: pmaftoon@gmail.com