Iranian Journal of Applied Linguistics (*IJAL*)

Vol. 22, No. 2, September 2019, 1-35

The Comparative Effect of Self-assessment vs. Peer-assessment on Young EFL Learners' Performance on Selective and Productive Reading Tasks

Shohreh Esfandiari, Kobra Tavassoli*

Department of ELT, Karaj Branch, Islamic Azad University, Karaj, Iran

Abstract

This study aimed at investigating the comparative effect of using self-assessment vs. peerassessment on young EFL learners' performance on selective and productive reading tasks. To do so, 56 young learners from among 70 students in four intact classes were selected based on their performance on the A1 Movers Test. Then, the participants were randomly divided into two groups, self-assessment and peer-assessment. The reading section of a second A1 Movers Test was adapted into a reading test containing 20 selective and 20 productive items, and it was used as the pretest and posttest. This adapted test was piloted and its psychometric characteristics were checked. In the self-assessment group, the learners assessed their own performance after each reading task while in the peer-assessment group, the participants checked their friends' performance in pairs. The data were analyzed through repeated-measures two-way ANOVA and MANOVA. The findings indicated that selfassessment and peer-assessment are effective in improving young EFL learners' performance on both selective and productive reading tasks. Further, neither assessment method outdid the other in improving students' performance on either task. These findings have practical implications for EFL teachers and materials developers to use both assessment methods to encourage learners to have better performance on reading tasks.

Keywords: Peer-assessment, Productive tasks, Reading comprehension, Selective tasks, Selfassessment, Young learners

Article Information:

Received: 15 April 2019 **Revised:** 2 July 2019 **Accepted:** 13 July 2019

Corresponding author: Department of ELT, Faculty of Literature and Foreign Languages, Karaj Branch, Islamic Azad University, Karaj, Iran **Email:** kobra.tavassoli@kiau.ac.ir

1. Introduction

Reading is an important language skill that learners should master to make use of it not only in education but also in their daily life. One of the important uses of reading in foreign language classes is learning the foreign language itself. In other words, vocabulary, structures, or any language element can be learnt through reading in a foreign language. In the process of learning a new language, the ability to read and comprehend a text is very essential. However, reading a text and comprehending a text are not synonymous. Learners may be fluent readers but weak comprehenders since comprehending a text is beyond the ability to combine letters to read words, sentences, paragraphs, and finally texts. Comprehension basically refers to understanding the meaning of a text (Alderson, 2000), for which various types of tasks, such as receptive/selective and productive tasks (Brown, 1996), can be used.

Meanwhile, in the process of learning a foreign language, learners in general and young learners in particular usually want to know how much progress they make and what their mistakes are to overcome them. More specifically, dealing with young learners is more challenging for teachers because these students are not mentally mature enough to understand explicitly what goes on in their learning process. To make such learners more engaged in their learning process to identify their own progress and mistakes and to find solutions for their mistakes is possible by using new types of assessment. The various types of assessment which can be used in different EFL (English as a foreign language) classes include using checklists, portfolios, teacher observations, journals, logs, conferences, selfassessment, and peer-assessment (Brown & Abeywickrama, 2010), and the list is not exhaustive. Out of these techniques, self-assessment and peerassessment were selected to be studied in this research because of their better implementation in young EFL learners' classes (McKay, 2006). In fact, it was tried to see if young learners can benefit the use of self- vs. peerassessment in overcoming the mistakes they face in learning a new language.

The purpose of this study was therefore to investigate the comparative effect of self-assessment vs. peer-assessment on young EFL learners' performance on selective and productive reading comprehension tasks. Many

scholars believe that self- and peer-assessment help young learners in the process of learning a new language (McKay, 2006) since by exposing them to self- and peer-assessment, teachers actually make the students responsible for their own as well as their friends' learning which is very important for their future as learning is a lifelong process. This way, young learners will become familiar with the benefits of self- and peer-assessment and will be able to use them in their future education by becoming more autonomous and cooperative.

2. Review of the Literature

2.1. Reading comprehension

Reading comprehension is a complex language skill which involves perceiving the written language (Alderson, 2000). Hoover and Gough (1990) defined reading comprehension as the combination of decoding information and linguistic comprehension in which decoding refers to the ability to recognize letters and words and the ability to read a text while linguistic comprehension focuses on the ability to take lexical information and understand meaning to interpret the discourse. According to Hoover and Gough (1990), assessing the students' ability to answer questions and perform on different tasks about the content of a reading text is a manifestation of their comprehension of the passage. There are a variety of reading tasks which can be used for this purpose. Brown and Abeywickrama
(2010) categorized reading tasks into four groups of perceptive, selective, interactive, and extensive, whereas Alderson (2000) categorized reading tasks into selected response, limited production response, and extended production response tasks. On the other hand, Brown (1996) specified two major task formats as receptive/selective response and productive response tasks. Receptive/selective response tasks are those requiring students to select an answer from among the available options, including multiplechoice where students are asked to select the correct answer from available choices, matching in which test takers are required to find matches between two sets of information, true-false where students should decide if the statement is true or not, and picture-matching vocabulary items where test takers are given input in the form of lists of words and pictures to match. On the other hand, productive response tasks require students to produce an answer. Fill-in-

theblank and short-response tasks are some examples of productive response tasks. Fill-in-the-blank tasks are sentences or texts with some blanks in them where students should fill in the missing words or phrases while shortresponse tasks are those in which test takers are required to provide their own responses with one word up to a sentence according to the information in the reading text (Alderson, 2000). Since the focus of this research was on young language learners, Brown's (1996) classification of tasks into receptive/selective and productive tasks, which is fairly practical for EFL classes, was chosen to compare young learners' ability in two major areas of comprehending and producing the English language.

2.2. Assessment

Various factors are effective on improving EFL learners' reading comprehension, one of which is appropriate assessment. Regardless of the teaching methods used, assessment plays an important role in teaching and learning processes and provides feedback to students (Alias, Masek, & Salleh, 2015). In fact, assessment can be used as an effective technique to improve learning different aspects of a second/foreign language (Ashraf & Mahdinezhad, 2015) because of its focus on using authentic contexts, determining the strengths and weaknesses of learners, encouraging human judgment, and using open disclosure of standards and rating criteria (Brown & Abeywickrama, 2010; Brown & Hudson, 1998; Herman, Aschbacher, & Winters, 1992; Huerta-Macías, 1995). According to Cheng and Warren (2005), being involved in assessment methods, procedures, and outcomes is significant for both teachers and learners. One common way of assessment to make learners involved in assessment and responsible for their own learning is utilizing self-assessment. According to Boud (2000), learners must learn to evaluate their own performance to become effective learners during their lives; to do this, they need to learn how to evaluate themselves to become independent of their teachers. Similarly, Miller (2003) believed that selfassessment is a helpful technique for learners due to its significant role in keeping learners motivated, involved, interested, responsible, and autonomous. A significant issue in the self-assessment process is that learners should get familiar with the standard scoring similar to that of their teacher and the correct way of monitoring their own performance (Babaii, Taghaddomi, & Pashmforoosh, 2016; Gipps, 1994) by answering the

question "How well have I done?" (Oscarson, 1989) to achieve improvement in the process of self-regulated learning (Butler, 2018).
Promotion of the learners' knowledge about the goals of language learning is an additional rationale for self-assessment.

Another important technique to get students involved in the assessment process is to utilize peer-assessment. Peer-assessment can develop selfconfidence, sense of ownership, responsibility, social skills, negotiation, and group work among learners. It can also change passive learning to active learning in a way that it increases interest among the learners. Topping, Smith, Swanson, and Elliot (2000) mentioned that the social and communication skills of learners including negotiation skills, verbal communication skills, giving and accepting criticism, justifying one's position, and assessing suggestions objectively increase through the process of assessing peers. However, the implementation of peer-assessment is time consuming and it needs allocating time to teach learners how to assess their peers. Furthermore, learners may not accept their mistakes (Hung, 2018), or they may have the chance of cheating where they know the peer who is going to assess their product by asking them to give the same score as they will (McDowell, 1995).

2.3. Young learners

McKay (2006) defined the term young language learners as those learners learning a second or foreign language during their elementary or primary school years or between ages 5 to 12. Some of the major characteristics of young learners are their need to be supported in understanding the content of a message, their short attention span, responding positively to adult attention, and being spontaneous, eager, and interested in new materials (Brassard & Boehm, 2007).

Moreover, understanding the characteristics of assessment tasks which can be used for young learners is important. Providing a suitable setting to elicit information about young learners' language knowledge and to assess their language in a playful and non-threatening situation is crucial (Zangl, 2000). According to Alderson (2000), the input provided to the learners in an assessment situation should be chosen according to the young learners' interest to enhance their motivation and willingness to show their knowledge.

The response they are supposed to produce should also be in accordance to the difficulty level of the task for their age. Also, the teacher should provide appropriate support and feedback to the responses they produce (McKay, 2006). These features make young learners completely unique in terms of teaching and assessment practices.

2.4. Related empirical studies

Plenty of studies have been done to explore the influence of using different assessment techniques to increase different aspects of EFL learners' language ability. Similar to this study, Nikmard and Tavassoli (2020) as well as Tavassoli and Nikmard (2019) investigated the effect of different types of assessment on EFL learners' performance on selective and productive reading comprehension tasks. The results of these studies showed a significant improvement on the learners' performance on both types of tasks when implementing diagnostic assessment (Nikmard & Tavassoli, 2020) and dynamic assessment (Tavassoli & Nikmard, 2019). In addition, Zandi (2017) investigated the effectiveness of different types of assessment on EFL learners' performance on selective and productive listening comprehension tasks and similarly found a significant improvement on both types of listening tasks using both dynamic and diagnostic assessment. Other studies also examined the influence of dynamic vs. diagnostic assessment on EFL learners' performance on descriptive and narrative writing (Ardin, 2017) and their speaking ability (Kazemi & Tavassoli, 2020), and both reported significant positive influences of the used assessment types. Moreover, Zangoei, Zareian, Adel, and Amirian (2019) did a study on the impact of computerized dynamic assessment on EFL learners' interlanguage pragmatic development. The results showed that computerized dynamic assessment could improve test takers' pragmatic comprehension competence.

Furthermore, Shahrakipour (2014) investigated whether self-assessment can influence EFL learners' receptive skills and tried to observe whether it has the same effect on intermediate and beginner level language learners. The results showed that self-assessment significantly improved EFL learners' receptive skills, however, the effect of self-assessment on the listening skill was less than the reading skill. More recently, Liu and Brantmeier (2019) explored the influence of self-assessment on reading and writing abilities of

young Chinese learners of English; the findings indicated that young Chinese learners tended to self-assess their foreign language reading and writing abilities accurately. Ashraf and Mahdinezhad (2015) also investigated the effect of self- and peer-assessment on EFL learners' autonomy and speaking skills and concluded that peer-assessment had a more significant effect on EFL learners' autonomy and speaking skill than self-assessment. In another study, Alias, et al. (2015) studied the use of self, peer-, and teacherassessment in problem-based learning and found that there was an association between self- and peer-assessment scores but not between teachers' and students' assessment scores. They also found that students had a tendency to give similar scores to themselves and their peers, which were much higher than the scores given by their teacher.

Overall, the findings mentioned above showed the positive impact of using various assessment types on EFL learners' performance on different aspects of the English language. However, none of these studies focused on the use of assessment types in young EFL learners' classes, even though young learners' language knowledge has been well studied. Chou (2014), for example, investigated the degree to which games, songs, and stories helped increase primary school pupils' English vocabulary size and found a significant effect of using games, songs, and stories on their vocabulary learning. In another study, Buckingham and Alpaslan (2017) found that the provision of out-of-class speaking practice to young learners of English contributed to improving their speaking proficiency and had a positive impact on their willingness to communicate. Blanch, Duran, Valdebenito, and Flores (2013) also studied the effect of a peer tutoring program on improving the reading comprehension ability of young learners by involving the students at school and their families at home, and they found that peer learning had a significant effect on primary students' reading comprehension ability whereas family involvement resulted in the development of academic skills in young learners. Further, Alsamadani (2017) and Tajareh and Oroji (2017) investigated the effectiveness of different techniques on young EFL learners' reading comprehension. The results showed a significant improvement on young EFL learners' reading comprehension ability through using talking story books (Alsamadani, 2017) and drama (Tajareh & Oroji, 2017).

In spite of the existence of the above-mentioned studies on various aspects of reading comprehension and using a variety of techniques to improve this

fundamental skill, such research on young EFL learners is scarce in the literature. Young learners are a very important group of learners especially when it comes to learning a new language, and if they do

not develop the necessary comprehension skills during reading, they will not be able to monitor their understanding of content, connect what they are reading to their own experience or prior knowledge, and may not be able to respond to tasks following each text (Muijselaar & de Jong, 2015). That is why expanding our knowledge about the impact of using different techniques including various assessment types on young learners' performance on reading tasks is vital. This was the main motive to do the present research.

2. Purpose of the Study

The purpose of this study was therefore to investigate the comparative effect of two assessment techniques, self-assessment and peer-assessment, on young EFL learners' performance on selective and productive reading comprehension tasks.

To achieve the goals of the present study, the following research questions were posed:

1. Does self-assessment have any significant impact on young EFL learners' performance on selective reading comprehension tasks?
2. Does self-assessment have any significant impact on young EFL learners' performance on productive reading comprehension tasks?
3. Does peer-assessment have any significant impact on young FFL learners' performance on selective reading comprehension tasks?
4. Does peer-assessment have any significant impact on young FFL learners' performance on productive reading comprehension tasks?
5. Is there any significant difference between using self-assessment vs. peerassessment on young EFL learners' performance on selective and productive reading comprehension tasks?

3. Method

3.1. Design

The present study was a pretest-posttest nonequivalent-groups study which is a kind of quasi-experimental design with intact classes (Best & Kahn, 2006). The reason for selecting such a design was the impossibility of having random selection of the participants for the researchers since they only had access to intact classes.

3.2. Participants

This study included 56 young EFL learners (31 males and 25 females) between 8-12 years of age who studied English in an institute. They were at the elementary proficiency level and they attended the class three times a week for 90 minutes. To ensure the homogeneity of the participants, they were selected from among 70 young learners in four intact classes on the basis of their performance on the A1 Movers test, which is a standard English language proficiency test for young learners. The students who scored between one standard deviation above and below the mean score on the A1 Movers test were selected. Then, the classes were randomly divided into two groups, self-assessment and peer-assessment, each group consisting of two classes to have enough number of students. Table 1 presents the pertinent demographic information of the participants.

Table 1. *Demographic information of the participants*

	Gender		Level	Age Range	Proficiency
Participants (N=56)	male	female		8 - 12	
	31	25			
			elementary		

3.3. Instrumentation

To collect the required data, the following instruments were used in this study:

- *Cambridge English Young Learners A1 Movers Test:* This is a series of fun and motivating English language tests for children in primary and lower secondary education, which was used as the proficiency test in this study. The A1 Movers test is the second level in Cambridge English Young Learners qualifications which are available at three assessment levels (Pre A1 Starters, A1 Movers, and A2 Flyers) that assess the English language knowledge of children (aged between 4 to 12). The A1 Movers test includes the four language skills of listening, reading and writing, and speaking. The listening part consists of 25 questions in which the test takers have to select the correct answers; there is also a section in which they must color a picture based on what they hear. The reading and writing parts are integrated in this test; there are 40 questions in the form of selective and productive reading comprehension tasks. The last part is speaking which includes four sections.

The first section is describing differences between two pictures; the second section is story telling; the third section is talking about odd-one-out pictures; and the fourth section is answering some personal questions.

- *Pretest and posttest:* An A1 Movers reading test containing 40 items in five tasks was selected and modified into a reading test containing 20 items in three selective reading tasks and 20 items in two productive reading tasks without changing the content of the test or the items. This test was piloted and then administered as the pretest and posttest in this study to check the students' performance in selective and productive reading comprehension tasks. The reliability of the pretest and posttest, which was calculated through Cronbach's alpha, is presented in Table 2. According to Muijs (2004), reliability values above .7 are high and the test is considered reliable.

Table 2. *Reliability of the pretest and posttest of the self-assessment and peer-assessment groups*

	Pretest, self-assessment group	Posttest, self-assessment group	Pretest, peer-assessment group	Posttest, peer-assessment group
N of Items	40	40	40	40
Cronbach's Alpha	.74	.62	.75	.72

Based on the information in Table 2, the pretest of the self-assessment group as well as the pretest and posttest of the peer-assessment group had acceptable reliability whereas the posttest of the self-assessment group had moderate reliability. Consequently, due to the not very high reliability values of pretest and posttest, the findings of this research should be interpreted cautiously.

3.4. Procedure

At the beginning of the study, the A1 Movers proficiency test was administered to 70 young learners in four intact classes to determine their proficiency level. Based on the results of this test, 56 young EFL learners at the elementary proficiency level whose scores were between one standard deviation above and below the mean score on the A1 Movers test were selected. Then, the four classes were randomly assigned to two experimental groups: self-assessment and peer-assessment, each group consisting two classes. Next, an A1 Movers reading test with five different reading tasks was selected and modified to include an equal number of selective and productive items. The modified test included 20 items in three selective tasks and also 20 items in two productive tasks. After piloting the test with a group of 30 other young EFL learners and checking the reliability and validity of the modified test, it was administered to the two groups of participants as the pretest. After the pretest, the treatment started. The treatment was carried out during 10 sessions in which the teacher, who was one of the researchers, and the students met three sessions per week, each session lasting for 90 minutes. During each session, both experimental groups were exposed to both selective and productive reading tasks after the reading texts they studied in their course book. In both groups, some techniques and strategies to answer selective and productive tasks were taught. However, in the self-assessment group, the learners assessed their own performance after each reading task with the teacher's help and talked about their feeling about the degree of their progress regularly, whereas, in the peer-assessment group, the participants talked about the same items in peers and checked their friends' performance in pairs with the teacher's guidance. After 10 sessions of treatment, the posttest which was

the same as the pretest with 20 selective and 20 productive items was administered in both groups to check the comparative effect of using self-

assessment vs. peer-assessment on the students' performance on selective and productive reading comprehension tasks.

3.5. Data analysis

Before running any statistical analyses, it was necessary to check whether the data gathered in this study were normally distributed or not. To do so, Onesample KS tests were used to check the normality of the data on the proficiency test, pretest, and posttest, which are reported in the results section. All the KS tests showed normal distribution of the scores. Therefore, parametric data analysis techniques could be safely used in this study. In order to respond to the first four research questions of the study, the pretest and posttest scores of the two experimental groups on the selective and productive tasks were compared through two repeatedmeasures two-way ANOVAs. This statistical technique is usually used to compare the performance of two groups (i.e., self-assessment and peerassessment groups) on one dependent variable (i.e., selective or productive tasks) on repeated measures (i.e., pretest and posttest) (Hinton, Brownlow, McMurray, & Cozens, 2004; Page, Braver, & Mackinnon, 2003). To respond to the last research question, multi-variate analysis of variance (MANOVA) was carried out. The selection of MANOVA was due to finding the relationship between the dependent variables in combination and with respect to the independent variables, and deciding whether the mean differences between the groups on the combination of dependent variables were significant or not (Hinton, et al., 2004; Pallant, 2011).

4. Results and Discussion

4.1. Preliminary analyses

Before running any statistical analyses, it was necessary to check whether the different pieces of data gathered in this study were normally distributed or not. One-sample KS test was used to check the normality of the scores by the two groups on the proficiency test, pretest, and posttest (Tables 3 and 4).

Table 3. *One-sample Kolmogorov-Smirnov test of the proficiency test of the self-assessment and peer-assessment groups*

		Proficiency test of the self-assessment group	Proficiency test of the peer-assessment group
N		29	27
Normal Parameter	Mean	81.62	82.19
	SD	2.92	2.80
Asymp. Sig. (2tailed)		.44	.30

Table 4. *One-sample Kolmogorov-Smirnov test of the pretest and posttest of the self-assessment and peer-assessment groups*

		Pretest, selfassessment group	Posttest, self-assessment group	Pretest, peer-assessment group	Posttest, peer-assessment group
N		29	29	27	27
Normal Parameters	Mean	25.14	31.55	28.15	33.56
	SD	3.72	2.47	3.35	3.51
Asymp. Sig. (2tailed)		.65	.11	.13	.92

According to the outcomes shown in Tables 3 and 4, it can be said that the data from the proficiency test, pretest, and posttest of both selfassessment and peer-assessment groups were normal as their asymptotic two-tailed levels of significance are all higher than the critical .05 level of significance (= 05; *p* >). Therefore, the data in this study were normally distributed and could be safely assigned to parametric data analyses which are suitable for normal distributions.

Next, it was necessary to check if the students in the two groups had similar levels of English knowledge at the outset of the study. Tables 5 and 6 are devoted to the related analyses. First, Table 5 provides the descriptive statistics of the proficiency test of the two groups, then, Table 6 reports the results of an independent-samples t-test on the proficiency test of the two groups.

Table 5. *Descriptive statistics of the proficiency test of the self-assessment and peer-assessment groups*

Group	N	Mean	SD
Self-Assessment Group	29	81.62	.54
Peer-Assessment Group	27	82.19	.53

Table 6. *Independent-samples t-test on the proficiency test of the selfassessment and peer-assessment groups*

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Proficiency test	Equal variances assumed	.10	.75	-.73	54	.46

The mean scores of the participants' performance on the proficiency test which were 81.62 and 82.19 for the self-assessment and peerassessment groups (Table 5) were so close to each other that it could be claimed the learners were almost at the same level of proficiency at the beginning of the study. This assumption was statistically checked through an independentsamples t-test (Table 6), which showed that there was not a significant difference between the two groups' variances as the sig. value reported for Levene's test for equality of variances was .75 and higher than the critical .05 level of significance. Also, the sig. value reported for the ttest for equality of

means was .46 and above the critical .05 level of significance (t = -.73; p = .46; □ .05; p >). In□other words, there was not a significant difference between the participants' performance on the proficiency test at the beginning of the study. Therefore, the participants were initially homogeneous regarding their English language ability and the results of the two groups could be compared safely.

4.2. Investigation of the research questions

As it was mentioned in the data analysis section, to answer the research questions of the study, the researchers used two repeated-measures two-way ANOVAs and a MANOVA to check the effect of the independent variables on the dependent variables.

The results of these analyses are presented in Tables 7-10. First, Table 7 presents the descriptive statistics of the scores on selective and productive reading comprehension tasks in the pretest and posttest of the selfassessment and peer-assessment groups.

Table 7. *Descriptive statistics of the selective and productive tasks in the pretest and posttest of the self-assessment and peer-assessment groups*

		Selective Tasks in Pretest	Selective Tasks in Posttest	Productive Tasks in Pretest	Productive Tasks in Posttest
Self-Assessment Group (N=29)	Mean	13.10	16.14	12.03	15.41
	SD	1.52	1.62	3.14	2.26
PeerAssessment Group (N=27)	Mean	14.63	17.00	13.15	16.56
	SD	2.84	2.03	1.76	1.90

Checking the results of Table 7, it becomes clear that the performance of the participants of the self-assessment group improved in both selective and productive tasks since their mean score in selective tasks changed from 13.10 to 16.14 from pretest to posttest and their mean score in productive tasks had a change from 12.03 to 15.41, both of which show a fair amount of progress. Similarly, the performance of the participants of the peerassessment group showed a good amount of improvement in both kinds of tasks too, from 14.63 in pretest to 17.00 in posttest of selective tasks and from 13.15 in pretest to 16.56 in posttest of productive tasks.

On the other hand, checking the mean scores of selective tasks in pretests of the two groups, i.e. self-assessment and peer-assessment groups, which are 13.10 and 14.63, and the mean scores of selective tasks in posttests of the groups, which are 16.14 and 17.00, makes it clear that the participants of the self-assessment group performed better on the posttest. Moreover, checking the mean scores of productive tasks in pretests of the self-assessment and

peer-assessment groups, which are 12.03 and 13.15, and comparing them with the related posttest mean scores, which are 15.41 and 16.56 respectively, show that the two groups' performance improved in the posttest, and the degree of the improvement of the two groups in productive tasks was almost the same.

However, to check whether the improvements in selective and productive tasks of the two groups were significant or not, two repeatedmeasures twoway ANOVAs were run, the outcomes of which are presented in Tables 8 and 9.

Table 8. *Repeated-measures two-way ANOVA of selective tasks in the pretest and posttest of the self-assessment and peer-assessment groups*

Effect		Value	F	Sig.	Partial Eta Squared
Time	Pillai's Trace	.67	110.69	.00*	.67
Group				.02*	5 .09
Time * Group	Pillai's Trace	.03	1.67	.20	.03

According to the sig. value reported for time (within-subjects effects) in Table 8, which is .00, it could be concluded that there was a significant difference within each group from pretest to posttest in selective tasks. That is to say, the participants of the two groups had a considerable amount of improvement in selective tasks from pretest to posttest (F = 110.69; p = .00; □ = .05; p <□). In addition, the partial eta squared reported is .67 which shows the large effect size of this factor, according to Cohen's (1988) categorization

of partial eta squared values cited in Pallant (2011) in which .01=small effect, .06=medium effect, and .138=large effect.

On the other hand, the sig. value reported for group (between-subjects effects) in the second row is .02 which is also smaller than the critical .05 level of significance (F = 5.48; p = .02; α = .05; $p < \alpha$). It means that there was a significant difference between the performance of the two groups. The effect size of this factor is medium as the partial eta squared is .09.

Furthermore, the significance level for the interaction of time and group in the last row of Table 8 is .20 which is higher than the critical .05 level of significance (F = 1.67; p = .20; α = .05; $p >$). Hence, the conclusion was that the progress made in the two groups from pretest to posttest was similar to each other on selective reading comprehension tasks. However, as it is shown by the partial eta squared, .03, the size of this effect is small.

Next, it was the time to check the amount of the influence of the treatments given to the participants in the two groups on productive tasks (Table 9).

Table 9. *Repeated-measures two-way ANOVA of productive tasks in the pretest and posttest of the self-assessment and peer-assessment groups*

Effect		Value	F	Sig.	Partial Eta Squared
Time	Pillai's Trace	.78	198.33	.00*	.78
Group			.05	3	.06
Time * Group	Pillai's Trace	.00	.00	.95	.00

According to the information reported in the first row of Table 9, named time, there was a significant difference within each group from pretest to

posttest on productive tasks as the sig. value reported is .00 which is smaller than the critical .05 level (F = 198.33; p = .00; α = .05; $p < \alpha$). The partial eta squared reported in this case, which is .78, shows the large effect size of time.

Meanwhile, the sig. value of the second row, i.e. group, is .05, which is equal to the critical .05 level, meaning that the performance of the participants of the two groups on productive tasks was not significantly different from each other (F = 3.77; p = .05; α = .05; $p = \alpha$). That is, the two groups performed almost the same on productive tasks. The effect size of this factor is considered medium according to the partial eta squared reported which is .06.

Furthermore, the level of significance calculated for the interaction of time and group is reported as .95 which is a value much higher than the critical .05 level, meaning that there was a similar progress in the participants' performance on productive reading comprehension tasks from pretest to posttest in both groups (F = .00; p = .95; α = .05; $p > \alpha$). However, as it is shown by the partial eta squared, .00, the size of this effect is zero.

Finally, as the researchers were keen on investigating the effect of both self-assessment and peer-assessment on both selective and productive tasks simultaneously, a MANOVA was used (Hinton, et al. 2004). The outcomes of this analysis are presented in Table 10.

Table 10. *MANOVA on the selective and productive tasks in the pretest and posttest of the self-assessment and peer-assessment groups*

Source	Measure	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Selective	39.87	1	39.87	8.74	.00*	.07
	Productive	35.56	1	35.56	6.44	.01*	.05
Time	Selective	204.22	1	204.22	44.79	.00*	.29
	Productive	322.00	1	322.00	58.34	.00*	.35
Time * Group	Selective	3.08	1	3.08	.67	.41	.00
	Productive	.00	1	.00	.00	.97	.00

The sig. value of the selective tasks of the two groups reported in Table 10 (.00) makes it clear that there was a significant difference between the performance of the two groups on selective reading tasks (F = 8.74; p = .00; □ = .05; p <□). The partial eta squared reported which is .07 shows a moderate effect size (Cohen, 1988 as cited in Pallant, 2011). Similarly, the sig. value of the productive reading tasks of the two groups which is .01 is also smaller than the critical .05 level, which means there was a significant difference between the performance of the two groups on productive reading tasks as well (F = 6.44; p = .01; □ = .05; p < □). The size of this effect is low as the partial eta squared reported is .05.

On the other hand, the information in the second row called time (from pretest to posttest) shows that there was a significant difference within each group from pretest to posttest in both selective and productive types of tasks as the sig. values reported for both are .00 (F = 44.79; p = .00; □ .05; p <

□ for selective tasks, and F = 58.34; p = .00; □ = .05; p < □ for productive tasks). It means the two groups had a considerable amount of improvement from pretest to posttest in both kinds of selective and productive tasks, which can be considered as the positive effect of the two types of assessments, i.e., self-assessment and peer-assessment. Moreover, the size of this effect in both selective and productive tasks is large due to their partial eta squared reported which are .29 and .35 respectively.

Once more, the information in the last row is more important as it reports the results of the interaction of time and group. Here, the conclusion is that the participants did not progress considerably differently from each other on either selective or productive reading tasks from pretest to posttest as the sig. values reported for selective and productive tasks are .41 and .97 respectively which are both higher than the critical .05 level (F = .67; p = .41; □ .05; p > for selective tasks, and F = .00; p = .97; □ = .05; p > □ for productive tasks). According to the partial eta squared values which are .00 for both types of tasks, the effect size of the interaction of time and group is zero for both selective and productive tasks.

The outcomes of data analyses in this study reported in Tables 7-9 showed that the null hypotheses corresponding to research questions 1-4 could be rejected. However, the outcomes of Tables 7 and 10 showed that the null hypothesis corresponding to research question 5 about the interaction of all the variables could not be rejected. In other words, it was found that:

- Self-assessment has a significant impact on young EFL learners' performance on selective reading comprehension tasks.
- Self-assessment has a significant impact on young EFL learners' performance on productive reading comprehension tasks.

- Peer-assessment has a significant impact on young FFL learners' performance on selective reading comprehension tasks.
- Peer-assessment has a significant impact on young FFL learners' performance on productive reading comprehension tasks.
- There is no significant difference between using self-assessment or peerassessment on young EFL learners' performance on selective and productive reading comprehension tasks.

5.3. Discussion

Congruent with the results of this study regarding enhancing EFL learners' performance in different task types through various assessment techniques, Nikmard and Tavassoli (2020), Tavassoli and Nikmard (2019), and Zandi (2017) investigated the effectiveness of diagnostic assessment and dynamic assessment on EFL learners' performance on selective and productive reading as well as listening comprehension tasks and found that to improve the learners' performance on selective and productive tasks, both diagnostic and dynamic assessment are two influential techniques. These similar results highlight the significance of using assessment techniques in EFL classes to improve the learners' language knowledge.

Furthermore, various studies showed the positive impact of selfassessment and/or peer-assessment on the learners' language achievement. For example, Vangah (2013) found that self-assessment had a

significant effect on students' reading ability and Shahrakipour (2014) detected that self-assessment significantly improved EFL learners' receptive skills. These findings in addition to the results of this study on the usefulness of selfassessment in EFL classes are in line with what is stated in the literature about self-assessment as a useful technique to make the learners interested,

involved, responsible, and autonomous in their process of learning (Miller, 2003). Topping, et al. (2000) also did a study on peer-assessment whose results showed that although students believed peer-assessment was time consuming, intellectually challenging, and socially uncomfortable, it improved the quality of their own subsequent written work and helped them develop other transferable skills. Further, Ashraf and Mahdinezhad (2015) reported the significant effect of peer-assessment on EFL learners' autonomy and speaking. These results in addition to what is found in this study also support the literature on peer-assessment which focuses on enhancing responsibility, self-confidence, social and communication skills, verbal skills, and group work among the learners (Topping, et al., 2000). Self- and peerassessment were also the focus of recent studies. Samaie, Mansouri Nejad, and Qaracholloo (2018) used self- and peer-assessment to explore the oral language proficiency of the learners online. The results showed that employing these assessment types even in online form can lead to the learners' language improvement. In addition, Seifert and Feliks (2019) employed both self- and peer-assessment to ensure students would take more responsibility for their learning, and the findings indicated that the students significantly benefited from these two techniques to improve their assessment skills and learning processes. These results are in fact in accordance with what was found in the present study regarding the positive impact of both self- and peerassessment on improving EFL learners' language ability. Finally, in a rare study on young learners, Hung, Samuelson, and Chen (2016) investigated how self- and peer-assessment can be implemented to evaluate young EFL

learners' oral presentation. Similar to the results of this study, the learners' improvement at the end of the study was considerable following both selfassessment and peerassessment. These findings all show the importance of eliciting young learners' language knowledge in a non-threatening and friendly atmosphere in which they are involved in assessment themselves (Zangl, 2000).

Overall, the results of various studies mentioned above as well as the results of the present study showed that self-assessment and peer-assessment can enhance students' achievement to a great extent, most probably because when self-scoring and peer-scoring, students' metacognitive strategies and self-consciousness are enhanced (Anderson & Krathwohl, 2001; Bouzidi & Jaillet, 2009). That is why, self-assessment and peer-assessment are regarded as favorable and valuable learning activities, especially in collaboration with each other, because the multi-dimensional nature of assessment practice is beneficial for students (Bouzidi & Jaillet, 2009). In fact, self- and peerassessment can lead to successful acquisition of any set of skills by improving the learners' level of responsibility (Topping & Ehly, 1998), autonomy, motivation (Brown & Abeywickrama, 2010), and shared experience. Furthermore, the use of self- and peer-assessment among young learners results in a cooperative situation in which the learners are willing to help and assess their peers and take responsibility for their own achievement which can lead to more in-depth study, better evaluation, better social skills, and more efficient assessment (Blanche & Merino, 1989; Oscarson, 1989; Ross, 1998).

5. Conclusion

As it was mentioned earlier, EFL learners' reading comprehension can be improved through a variety of techniques including suitable assessment types. Learners take an active role in the assessment process with self- and peerassessment, both of which play an important role in increasing their reading comprehension ability. Furthermore, from the findings of this study, it could be concluded that both self- and peer-assessment can significantly improve young EFL learners' performance on selective and productive reading comprehension tasks in comparison to the common reading techniques which

might not give any opportunity to these learners to assess their own or their peer's performance.

The findings of this study have a number of implications for different involved parties. Language teachers are highly recommended to include more educational practices such as self-and peer-assessment in their teaching, which might guarantee the students' learning and increase their motivation and autonomy which are by themselves important factors in the process of learning. Teachers also do not need to worry about the reliability of the students' self- and peer-ratings since learners can get more and more accurate in rating after enough training and practice is offered. Furthermore, students can benefit from assessing themselves or their peers as both these techniques encourage sophistication in learner awareness and help learners to make accurate judgments on their own abilities, to acquire how to do an evaluation that covers the whole learning process, and to see errors as something helpful.

Finally, some recommendations are presented that may aid other researchers to do further similar studies. This study was limited to two types of alternative assessment and two types of reading tasks; it is desirable to use other types of assessment to examine their effects on the learners' performance on different

reading comprehension tasks. The effect of self- and peer-assessment as well as other types of assessment on other skills and subskills of language can also be explored in further studies. In addition, because of the widespread use of technology in recent years, investigating the effectiveness of various alternative assessments on different skills in online settings can also be studied. Lastly, qualitative investigations on using different assessment techniques and their impact on teachers, students, and parents can be done by interviewing them and observing what happens in classroom contexts.

Acknowledgements: The authors wish to thank the IJAL editor-in-chief and the two anonymous reviewers for their insightful comments to improve the quality of the article.

6. References

Alderson, J. C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Alias, M., Masek, A., & Salleh, H. H. M. (2015). Self, peer, and teacher assessments in problem-based learning: Are they in agreements? *Procedia-Social and Behavioral Sciences*, *204*, 309-317.

Alsamadani, H. A. (2017). The effect of talking story books on Saudi young EFL learners' reading comprehension. *English Language Teaching*, *10*(5), 204-213.

Anderson, L., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: London.

Ardin, M. (2017). *The effect of diagnostic assessment vs. dynamic assessment on EFL learners' performance on descriptive and narrative writing.* Unpublished MA Thesis, Islamic Azad University, Karaj, Iran.

Ashraf, H., & Mahdinezhad, M. (2015). The role of peer-assessment versus self-assessment in prompting autonomy in language use: A case of EFL learners. *Iranian Journal of Language Testing, 5*(2), 110-120.

Babaii, E., Taghaddomi, S., Pashmforoosh, R. (2016). Speaking selfassessment: Mismatches between learners' and teachers' criteria. *Language Testing, 33*(3), 411-437.

Best, J. W. & Kahn, J. V. (2006). *Research in education (3rd ed.).* Boston: Pearson Education Inc.

Blanch, S., Duran, D., Valdebenito, V., & Flores, M. (2013). The effects and characteristics of family involvement on a peer tutoring programme to

improve the reading comprehension competence. *European Journal of Psychology of Education, 28*(1),101-119.

Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning, 39*(3), 313-340.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-16.

Bouzidi, L., & Jaillet, A. (2009). Can online peer-assessment be trusted? *Educational Technology & Society, 12*(4), 257-268.

Brassard, M. R., & Boehm, A. (2007). *Preschool assessment: Principles and practices.* New York: Guilford press.

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

Brown, J. D. (1996). *Testing in language programs.* NY: McGraw hill.

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

Buckingham, L., & Alpaslan, R. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System, 65*, 25-37. Butler, Y. G.

(2018). The role of context in young learners' processes for responding to self-assessment items. *The Modern Language Journal, 102*(1), 242-261.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93-121.

Chou, M. H. (2014). Assessing English vocabulary and enhancing young English as a foreign language (EFL) learners' motivation through games, songs, and stories. *Education, 42*(3), 284-297.

Gipps, C. (1994). *Beyond testing: Towards the theory of educational assessment.* London: The farmer press.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004*). SPSS explained*. London and New York: Routledge.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*(2), 127-160.

Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5*(1), 8-11.

Hung, Y. J. (2018). Group peer assessment of oral English performance in a Taiwanese elementary school. *Studies in Educational Evaluation, 59*, 19-28.

Hung, Y. J., Samuelson, B. L., & Chen, S. C. (2016). Relationships between peer- and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 317-338). New York: Springer.

Kazemi, N., & Tavassoli, K. (2020). The comparative effect of dynamic vs. diagnostic assessment on EFL learner's speaking ability. *Research in English Language Pedagogy, 8*(2), 223-241.

Liu, H., & Brantmeier, C. (2019). I know English: Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. *System, 80*, 60-72.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International, 32*, 302313.

McKay, P. (2006). *Assessing young language learners.* New York: Cambridge university press.

Miller, P. J. (2003). The effect of scoring criteria specificity on peer and selfassessment. *Assessment and Evaluation in Higher Education, 28*(4), 383-394.

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. London: Sage Publications Ltd.

Muijselaar, M. M., & de Jong, P. F. (2015). The effects of updating ability and knowledge of reading strategies on reading comprehension. *Learning and Individual Differences, 43*, 111-117.

Nikmard, F., & Tavassoli, K. (2020). The effect of diagnostic assessment on EFL learners' performance on selective and productive reading tasks. *Journal of Modern Research in English Language Studies, 7*(1), 79104.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(1), 1-13.

Page, M. C., Braver, S. L., & Mackinnon, D. P. (2003). *Levine's guide to SPSS for analysis of variance (2nd ed.).* London: Lawrence Erlbaum Associates.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows (4th ed.)*. Crows Nest NSW 2065 Australia.

Ross, S. (1998). Self-assessment in second language testing: A metaanalysis and analysis of experimental factors. *Language Testing, 15*(1), 1-19.

Samaie, M., Mansouri Nejad, A., & Qaracholloo, M. (2018). An inquiry into the efficiency of WhatsApp for self- and peer-assessments of oral

language proficiency. *British Journal of Educational Technology, 49*(1), 111-126.

Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teacher assessment skills. *Assessment & Evaluation in Higher Education, 44*(2), 169-185.

Shahrakipour, H. (2014). On the impact of self-assessment on EFL learners' receptive skills performance. *International Research Journal of Arts & Humanities (IRJAH), 6*(1), 1-13.

Tajareh, M. J., & Oroji, M. R. (2017). The impact of drama on Iranian EFL young learners' reading comprehension performance. *Bulletin de la Société Royale des Sciences de Liège, 86,* 67-80.

Tavassoli, K., & Nikmard, F. (2019). The effect of dynamic assessment on EFL learners' performance on selective and productive reading comprehension tasks. *Foreign Language Research Journal, 9*(2), 445478.

Topping, K. J., & Ehly, S. (1998). *Peer-assisted learning.* Lawrence Erlbaum, Mahwah.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education, 25*(2), 149-169.

Vangah, F. P. (2013). Effect of self-assessment on Iranian EFL learners' reading skill and vocabulary knowledge. *International Research Journal of Applied and Basic Sciences, 4*(3), 676-680.

Zandi, G. (2017). *The effect of diagnostic assessment vs. dynamic assessment on EFL learners' performance on selective and productive listening tasks*. Unpublished MA Thesis, Islamic Azad University, Karaj, Iran.

Zangl, R. (2000). Monitoring language skills in Austrian primary (elementary) schools: A case study. *Language Testing, 17*(2), 250-60.

Zangoei, A., Zareian, G., Adel, S. M. R., & Amirian, S. M. R. (2019). The impact of computerized dynamic assessment on Iranian EFL learners' interlanguage pragmatic development. *Journal of Modern Research in English Language Studies, 6*(4), 139-165.

Notes on Contributors:

Shohreh Esfandiari is an MA graduate of TEFL from Islamic Azad University, Karaj, Iran. She has been a successful teacher of young EFL learners for more than 7 years. Her major area of interest is teaching, assessing, and developing materials for young learners.

Kobra Tavassoli is an assistant professor of TEFL at Islamic Azad University, Karaj, Iran. She is teaching language assessment, research methodology, and language teaching methodology courses at BA, MA, and PhD levels. Her areas of interest are language assessment and teacher education. She has published in different journals and presented in national and international conferences on these topics.