



The Comparative Effects of Portfolio Assessment and Peer-Assessment on EFL Learners' Critical Thinking and Speaking Achievement

Haleh Askarzadeh

Islamic Azad University, Central Tehran Branch
Email: Hale.armeniakom1990@gmail.com

*Behdokht Mall-Amiri, Ph.D. **

Islamic Azad University, Central Tehran Branch
Email: Beh.malamiri@iauctb.ac.ir

Abstract

This study compared the effects of portfolio assessment and peer-assessment on EFL learners' critical thinking and speaking achievement. For this purpose, 32 EFL learners attending Diplomat Institute in Tehran were non-randomly selected based on their scores on PET. They were randomly assigned to two experimental groups of 16. The portfolio assessment group went through the procedure of creating portfolio based on Evaluation Portfolio Model recommended by Valencia and Calfee (1991). The peer-assessment group practiced peer-assessment according to Yamashiro and Johnson's (1997) Model. Finally, both experimental groups took a speaking test of PET and a critical thinking questionnaire as posttests. The data analysis using RM ANOVA revealed that both experimental groups had similarly a higher post-treatment level of critical thinking. The analysis of a Mann-Whitney U test on the gain scores revealed that the level of post-treatment speaking in the peer-assessment group was significantly higher compared to the portfolio assessment group.

Keywords: Critical Thinking; Peer Assessment; Portfolio; Speaking

1. Introduction

Assessment is crucially important in all educational contexts (Naeini, 2013) and has been subject to many studies (e.g., Davin, 2011; Garb & Kozulin, 2002; Ghahremani & Azarizad, 2013; Xiaoxiao & Yan, 2010). Marzano (2000) considers assessment as an effective tool which contributes to the enrichment of learning. It has been classified into different types. Two types of assessment, which were the focus of the current study, are portfolio and peer-assessment. As Learning (2007) asserts, portfolio is a “purposeful collections of student work which can serve as the basis for evaluation of student effort, progress, and achievements in English language arts” (p.14). Learning further maintains that portfolio refers to a collection concerning any aspects of students’ work that tells the story of their achievements, skills, efforts, abilities, and contributions to a particular skill.

Another kind of assessment is peer assessment. Pearce (2009, as stated in Liu & Carless, 2006) puts emphasis on the benefits resulting from peer-assessment, arguing that peer review motivates learners to assume an active part in exercising autonomy as well as managing their own learning . As pointed out by Crooks (2001), assessment is concerned with any process that yields information regarding the thinking, achievement or progress of learners. Since, as Crookes believes, assessment is concerned with any process that yields information regarding thinking, it is assumed that assessment might have contributions to critical thinking as well.

One of the essential cognitive abilities which has been emphasized by educational experts is critical thinking. According to educational experts (e.g. Moon, 2008; Wright, 2002), in the fast-paced and ever-changing world, critical thinking has been considered by many scholars as a basic survival skill. Philosophers of education (e.g. Ennis, 1996; Paul, 1988) agree that critical thinking is the fundamental goal of learning and particularly central to higher education. Educational psychologists, such as Thomas and Smoot (1994) and Huitt (1998) have asserted that critical thinking should be considered as a very important element in the educational systems of the 21st century.

Among the four language skills, speaking is probably of higher popularity. Speaking is a means through which learners can interact with each other to achieve certain goals and express their opinions (Miller, 2001). Those people who know a language are referred to as speakers of that language, as if speaking includes all other types of skills, and many, if not most foreign language learners are primarily interested in learning to speak (Miller, 2001). Hence, helping EFL learners to develop their speaking proficiency through some more effective techniques seems to be needed.

2. Literature Review

2.1. Assessment

According to Shohamy (1992) “Assessment is a superordinate term which includes all forms of assessment. It not only assigned scores to students but also diagnoses their problems and remedies them through employing specific methods and techniques” (p.54).

The term *portfolio* has been defined by many scholars. It is defined by Paulson (1991) as “a purposeful collection of student work that exhibits the students’ effort, progress achievement in one or more areas” (p.60). According to Genesee and Upshur (1996), a portfolio is considered as “a purposeful collection of student’s work that demonstrates their efforts, progress, and achievement in a given area” (p.99), and may strengthen students' learning in that they (a) capitalize on work that would normally be done in the classroom anyway; (b) focus learners’ attention on learning processes; (c) facilitate practice and revise processes; (d) help motivate students, if well-planned, because they present a series of meaningful and interesting activities; (e) increase students’ involvement in the learning processes; (f) foster student-teacher and student-student collaboration; (g) provide means for establishing minimum standard for classroom work and progress; (h) encourage students to learn the meta language necessary for students and teachers to talk about language growth (Brown & Hudson, 1998, p. 664). Portfolios are effective tools to foster learners’ reflection and they affect the students’ autonomy, take the teacher’s role and involve in assessment process (Yang, 2003).

The review of literature shows that portfolio assessment has been adopted in a lot of subject areas and adopted in various contexts, hence there are many descriptions of portfolio. Portfolio approaches taken to assessing literacy have been explained in a variety of publications (Flood & Lapp, 1989; Valencia, 1990; Hamps-Lyons, 1996). The investigations conducted on the use of portfolio assessment in L2 teaching (particularly foreign language) indicate that such type of assessment enhances writing skills. Nassirdoost and Mall-Amiri (2015) examined the impact of portfolio assessment on EFL learners’ vocabulary achievement and motivation. The findings of the study indicated that that the use of portfolio assessment had a significant effect on EFL learners’ vocabulary achievement but it did not affect EFL learners’ motivation level.

Another considerable form of alternative assessment is peer-assessment. The importance of this assessment highlighted in different educational learning and educational

research. Slavin (1997) refers to peer-assessment as one of the best and perfect success in educational history. Pedagogically, peer-assessment improves learning of student (Falchikov & GoldFinch, 2000) through “a sense of ownership and responsibility, motivation, and reflection of the students’ own learning” (Saito & Fujita 2009, p. 151).

Banerjee (2001) states that, “The increased interest in involving the learners in all phases of the learning process and in encouraging learners’ autonomy and decision making has led to the interest in self-assessment.” (p. 227). Self-assessment requires the students to rate their own language, whether through performance self-assessments, comprehension self-assessments, or observation self-assessments. It has been argued that peer-assessment can make the learners to perform actively (Orsmond & Merry, 1996) and increase their higher order thinking (Cheng & Warren, 2005).

2.2. Critical Thinking

Today, the presence of learners who are autonomous and critical thinker is a great necessity for the society, because of the many changes in academic requirements (Ming & Alias, 2003). Critical learners’ autonomy are both widely seen as desirable educational goals, and often understood as independent or even mutually indispensable attributes (Pemberton & Nix, 2012). Critical thinking is what seems to be essentially needed in higher education since as Elder and Paul (1994) believe, critical thinking is related to the thinkers’ ability in order to take cost of their own thinking and develop proper criteria and standards for evaluating and assessing their own thinking.

Learners are expected to be able to think critically to make decisions and solve their study problems. Halpern (1998) defines critical thinking as the use of cognitive skills or strategies that increase the probability of a desirable outcome. She says critical thinking is purposeful, reasoned, and goal directed. She concluded that critical thinking is the kind of thinking used in problem solving, setting suitable outcomes, expressing inferences, and making decisions.

According to Chaffee (1992), thinking plays an important role in life, and helps people in various issues. In the same way, Santrock (2008) believes that thinking has different functions such as reasoning, thinking critically, making decisions and solving problems. Chaffee (2009), states the most important purpose of CT is to make “more intelligent decisions”, and a person who is critical thinker can make intelligent judgments and think about “important ideas”, (p. 43).

2.3. Speaking

Developing speaking skill takes on importance, and merits attention, on two fronts. On the one hand, what portrays speaking more prominent than the other skills is the observation that many language students are keener to speak the second language as their primary goal. According to Lazaraton (2001)," the ability to speak a language is synonymous with knowing that language since speech is the most basic means of human communication" (p. 103). On the other, teaching English in EFL contexts may be much more demanding compared with ESL contexts in that in the former a dearth of access to speaking opportunities outside the classroom situation puts a heavier burden on the EFL teacher. With respect to the nature of speaking per se, Bialy and Savage (1994, as cited in Lazaraton, 2001, p.103) maintain that, "speaking in a second or foreign language has often been viewed as the most demanding of the four skills", the reasons for which could be what Brown (1994) has listed as the use of reduced forms in speech, such as reduced vowel reduction, contractions and elision, and use of slangs and idioms that learners need to be equipped with or to sound bookish. Furthermore, stress, rhythm and intonation patterns in English speaking are what need to be mastered by learners.

Hedge (2000) in his chapter for speaking holds that, " perhaps the first question to ask.. is what reasons we have for asking our students to practice speaking in the classroom. One is that, for many students, learning to speak competently in English is a priority. They may need this skill for a variety of reasons. For example to keep up rapport in relationships, influence people, and win or lose negotiations. It is a skill by which they are judged while first impressions are formed" (p.261).

Given the place and importance of critical thinking as well as speaking skill English language learning, the present researchers thought of the possible impact of alternative assessment as means of improving these abilities. Thus, the following research questions were addressed:

1. Does portfolio assessment have any significant effect on EFL learners' critical thinking ability?
2. Does peer assessment have any significant effect on EFL learners' critical thinking ability?
3. Is there any significant difference between the effects of portfolio assessment and peer assessment on EFL learners' critical thinking ability?

4. Is there any significant difference between the effects of portfolio assessment and peer assessment on EFL learners' speaking ability?

2. Methodology

2.1 Design of the Study

This study was quantitative and enjoyed a quasi-experimental, two equivalent groups pretest-posttest design. Alternative assessment was the independent variable with two modalities: portfolio assessment and peer-assessment which the researcher investigated their effect on EFL learners' critical thinking as well as speaking achievement of learners as the dependent variables.

2.2. Participants

The participants of the current study were 32 female intermediate EFL learners within the age range of 18-26 studying English at intermediate level in Diplomat Foreign Language Institute located in Tehran. Initially, 52 learners determined by the institute's placement criterion to be at intermediate level of proficiency, were non-randomly selected to participate in this study. Based on their scores on PET (2008), 32 students whose scores fell within one standard deviation below and above the mean were selected. Then they were randomly assigned to two experimental groups of 16. An experienced teacher of English along with one of the researchers, rated the speaking performances of the learners as well as their writings, and inter-rater reliability was estimated.

2.3. Instruments

A number of instruments were applied in this study in order to conduct the research and collect the required data.

2.3.1. Critical Thinking Questionnaire (CTQ)

Honey's (2000) CT, adopted from Naieni (2005), was used to measure the learners' critical thinking. It contains 30 items exploring what a person might or might not do when critically thinking about a subject. It was administered to the participants to evaluate the three macro-skills of comprehension: the extent to which one ensures that s/he has a good understanding of an issue (10 items), analysis: the extent to which one breaks a subject down

into its component parts and scrutinizes each part (10 items), and evaluation: the extent to which one considers or assesses a topic in order to judge its value, quality, quantity, importance, condition, reliability, validity and logic (10 items) (Honey, 2000, as cited in Naieni, 2005).

The Likert-type CTQ, as it is stated by Naieni (2005), is reliable (.86 on Cronbach's Alpha). Also, it is a valid (highlighted by the literature) and practical (easy to administer, score, and interpret) measure of critical thinking ability.

2.3.2. Portfolio Assessment Model

In this study the Valencia and Calfee's (1991) Evaluation Portfolio Model was used to evaluate the oral communications of the students. Evaluation in this model includes: Grammar and vocabulary, Discourse management, Pronunciation, and Interactive Communication, scoring between 0-5 marks. The maximum possible score was 20 and the minimum was 0.

2.3.3. Peer-assessment Checklist

For peer-assessment group, the researcher used 'pre-flight checklist' strategy for assessing the peers. The checklist was adapted from Yamashiro and Johnson (1997) and it included 14 items. Four items for voice control, 3 items for body language, 3 items for contents of presentation and 4 items for effectiveness. Every section 4-5 students had to present the previous topic, orally and then their peers had to listen to them and to assess them by scoring the items from peer-assessment checklist, the scale ranges from 5(very good) to 1(poor).

2.3.4. Materials

The main course book applied to both experimental groups during the instruction was NORTHSTAR 1, by Merdinger and Barton (2015). This book consists of 9 units; each unit includes four parts as A, B, C, and D. The purpose of the units is integration of speaking skill. In this study, during 10 sessions of treatments, students worked on 3 units, 4-5-6, which were about Creativity in Business, Understanding Fears and Phobias, and Risks and Challenges.

2.3.5. Speaking Posttest

To inspect the comparative effect of the two treatments on speaking achievement of the learners, the PET speaking section, another version adapted from the book Cambridge English

Preliminary for Schools, Official Past Papers, 2009), was administered as a posttest. This section comprised 4 parts and 10 questions totally. The allotted time for this test was approximately 10-12 minutes. It was scored by the researcher and another qualified teacher based on the General Mark Schemes for speaking developed by Cambridge ESOL for PET, and finally the inter-rater reliability of the two sets of scores was estimated.

2.4. Data Collection Procedure

The present study was an attempt to investigate the comparative effect of portfolio and peer –assessments on EFL learners' critical thinking and speaking achievement of the learners. To accomplish this study, the following procedure was followed:

Prior to the treatment, piloting the PET test was the first phase for implementing the study. A version of PET adapted from PET Practice (Quintana, 2008) was administered to 30 non-participating candidates. The Cronbach's alpha formula was employed for calculating the reliability of the test scores gained by the participants.

The speaking part was rated according to General Mark Scheme by two raters, one of the researchers and another rater. Later on, the inter-rater reliability was calculated using the Pearson's rank correlation formula. So, the final score of each student was calculated by the average scores of the two raters. The participants whose score fell within the range of one standard deviation above and one standard deviation below the mean were selected as the participants in the study. Then, the selected participants were randomly divided in to 2 experimental groups. Their speaking ability was checked and compared primarily to see if the two groups were homogenous in this respect. The result showed that the two groups were not homogenous regarding their speaking ability prior to the treatment ($t= 3.27$, $p=.003<.05$). Furthermore, the comparison of the pre-treatment CT scores of the two groups, using t test, revealed that the two groups were homogeneous in this regard ($t=1.03$, $p=.31>.05$).

One group was assigned to be taught through portfolio assessment and the other group through peer assessment. All the participants were taught the same amount of instruction according to the syllabus of the language school including units 4, 5 and 6 of their course book. Both groups were taught by the same teacher (the researcher). The course consisted of 12 sessions, ninety minutes each, spanning over a period of approximately six weeks.

Experimental Group I: Portfolio Assessment Group

Since the study required the completion of portfolio, the learners were instructed for the first 2 sessions of the class on how to make and keep their portfolios for each conversation occurring between them and their partners while participating in the speaking and discussion activities and they were provided with the explanation of the nature, purpose and design of the portfolio.

To evaluate the oral communication of the students in the Portfolio assessment group, the researchers used the Evaluation Portfolio Model proposed by Valencia and Calfee (1991). In this model, the quality of the portfolio is evaluated based on the criteria and scoring schemes rating scale including 4 parts namely Grammar and vocabulary, Discourse management, Pronunciation, and Interactive Communication, scoring between 0-5 marks. Therefore, the maximum possible score was 20 and the minimum was 0.

According to the course book, there were 3 different parts in each chapter named; 1. Focus on topic, 2. Focus on listening, 3. Focus on speaking. In the first part the learners thought about the title of the chapter and also the following questions and tried to talk about them. The next part they listened to the text carefully to answer some questions and did some different listening activities, and in the last part which concentrated on speaking, they were supposed to work in pairs to read a story on that page, answer some questions about the text and interview each other using some questions given by the book. Also, in the last session of each chapter, all the students were supposed to bring their recorders to the class and before starting the conversation, they must turn on their devices to record oral communications based on the previous topic and after the class they uploaded their records to the instructor's email address. The instructor downloaded their files, scored them based on scoring system and provided feedback to the learners. So, every student created a portfolio containing their voices and feedback given by the instructor.

Experimental Group II: Peer-assessment Group

For this part of the study, the peer-assessment concept was taught to the learners in the first 2 sessions and the students were provided by the nature and the process of peer-assessment.

Every session, some students were supposed to deliver an oral presentation about the previous session's course book topics (Creativity in Business, Understanding Fears and

Phobias, and Risks and Challenges), while the peers graded them according to peer-assessment checklist in which 14 different items for oral presentations comprised voice control, body language, content and effectiveness which were adapted from Yamashiro and Johnson (1997) Peer-assessment Model. After each session, the instructor collected all the assessment checklists and gave them to the presenter.

At the end of the study, the participants of both groups sat for the posttest that was the speaking section of PET test which took 10-12 minutes. There were 4 different parts to assess their speaking abilities. The first part was some personal information, second part was a simulated situation, followed by section 4 which was responding to a photograph and the last one was a general conversation based on the photograph. It should be mentioned that the result of the test was evaluated by two qualified raters (the instructor and one of her colleagues) based on the PET rating scale.

Also, Honey's Critical Thinking Questionnaire was administered again in order to compare critical thinking abilities of the participants in the two groups after the treatment.

3. Data Analysis Procedure

Initially, two independent sample *t*-tests were run to compare the critical thinking and speaking abilities of the two groups. Finally, after treatment the researcher ran Repeated Measures ANOVA to measure the improvement of learners' critical thinking in each group from pretest to posttest and at the same time to compare the critical thinking posttest mean scores of the two groups.

To compare the speaking posttest scores of the two groups but remove the effect of pre-treatment speaking difference from the posttest, the researcher had to use an ANCOVA, but as the assumptions were not completely met, the procedure of comparing the gain scores was used as an alternative for ANCOVA, and as the normality of distribution of the gain scores was not met, Mann-Whitney U test was utilized to compare the gain scores.

4. Results

4.1. Selection of the Participants

In order to select homogeneous participants at intermediate level of the study, the researchers used a PET test. Moreover, prior to the selection phase, the PET test was piloted to

make sure that it could be used confidently for this screening. The following sections describe the details of consecutive processes of piloting and administration plus the further measures the researchers took to ensure as much homogeneity as possible.

The PET test was administered to a group of 30 EFL learners having almost the same characteristics as the target sample. All items went through an item analysis procedure, including item discrimination, item facility, and choice distribution. The results showed that all the items exhibited acceptable IF, ID, and CD indices. Accordingly, no items were discarded from the test.

The internal consistency of the PET scores gained from the participants in the piloting phase was estimated through using Cronbach's alpha coefficient which turned out to be .96. furthermore, the inter-rater reliability of the speaking scores given by the two raters came out to be as large as .87, and that for the writing scores was estimated as .83 using Pearson correlation formula.

The piloted PET test was administered among 52 EFL learners in order to enable the researcher to choose the homogenous participants of the study. Thirty two learners who scored one standard deviation above and below the mean score were selected as the main participants of this study. They were then randomly divided into two groups to receive the two treatments.

4.2. Answers to the First and Second and Third Questions

To test the first three hypotheses related to the effect of the two treatments on critical thinking of the learners separately as well as the comparison of the effectiveness of the treatments, a Repeated Measures ANOVA was used. Firstly, it was checked that the two groups were homogeneous regarding their critical thinking before the start of the treatments. The descriptive statistics of the pre and post-treatment critical thinking scores of the two groups is provided below:

Table 1.
Descriptive Statistics of Pre and Post Treatment CT Scores Across Groups

	N	Mean	Std. Deviation	Skewness		Ratios
	Statistic	Statistic	Statistic	Statistic	Std. Error	
portPreCT	16	98.0625	17.96466	-.231	.564	-.409
portPostCT	16	107.6250	21.53718	-.133	.564	-.23
peerPreCT	16	92.1875	13.93422	.260	.564	.46
peerPostCT	16	96.6250	11.56936	.173	.564	.30
Valid N (listwise)	16					

As shown in Table 1, all the skewness ratios were less than 1.96, which ensures that the four sets of data were normally distributed. The pre-treatment critical thinking scores of the two groups were compared and it was revealed that the two groups were not significantly different ($t=1.03$, $p=.31>.05$).

To answer the first three questions and test the corresponding hypotheses, Repeated Measures ANOVA was used to compare the pretest-posttest scores of the two groups at the one hand and compare the posttest scores of the two groups at the other simultaneously after verifying all the assumptions of the test, namely normality of distributions, and equality of covariance variances. The following tables are the outcome of the RM ANOVA:

Table 2.

Descriptive Statistics of Pre and Post CT Scores Across Groups

	Grouping	Mean	Std. Deviation	N
preCT	Portfolio	98.0625	17.96466	16
	Peer	92.1875	13.93422	16
	Total	95.1250	16.09398	32
postCT	portfolio	107.6250	21.53718	16
	Peer	96.6250	11.56936	16
	Total	102.1250	17.90071	32

Table 2 shows that the inter-correlations among the levels of the within-subjects' variables were the same ($F=2.54$, $p=.05>.001$) for both groups. So, the assumption of equality of covariance matrices is met.

Table 3.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Time	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Time	Linear	784.000	1	784.000	5.539	.025	.156
time * grouping	Linear	105.063	1	105.063	.742	.396	.024
Error(time)	Linear	4245.937	30	141.531			

As depicted in Table 3, the interaction effect turned out to be non-significant ($F=.742$, $p=.396>.05$) which means that the time difference does not depend on the treatments. Furthermore, the effect of the time came out to be significant ($F=5.54$, $p=.025<.05$) indicating that the increase in the CT scores from pretest to posttest for both groups was statistically significant: for the portfolio group (98.06 to 107.62) and for the peer group (92.18 to 96.62). Therefore, the first and second null hypotheses are rejected, with the conclusion that both treatments were significantly effective on critical thinking ability of learners. The effect size, as reported in the above table was as large as .156, which implies that 15.6 percent of the variation in the scores was due to time (the treatment that intervened in the two times of testing) which is a very large effect size according to Cohen (1988).

To test the third hypothesis, the between subject effects part of the RM ANOVA was utilized:

Table 4.

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	622521.000	1	622521.000	1497.292	.000	.980
grouping	1139.062	1	1139.062	2.740	.108	.084
Error	12472.937	30	415.765			

As illustrated in Table 4, the effect pertaining to the grouping variable turned out to be non-significant ($F=2.74$, $p=.108>.05$), which means that there was no significant difference between the effect of the two treatments on the critical thinking of the learners, and both treatments were equally effective. Thus, the third null hypothesis is maintained.

The following graph represents the pretest to posttest development of critical thinking of both groups.

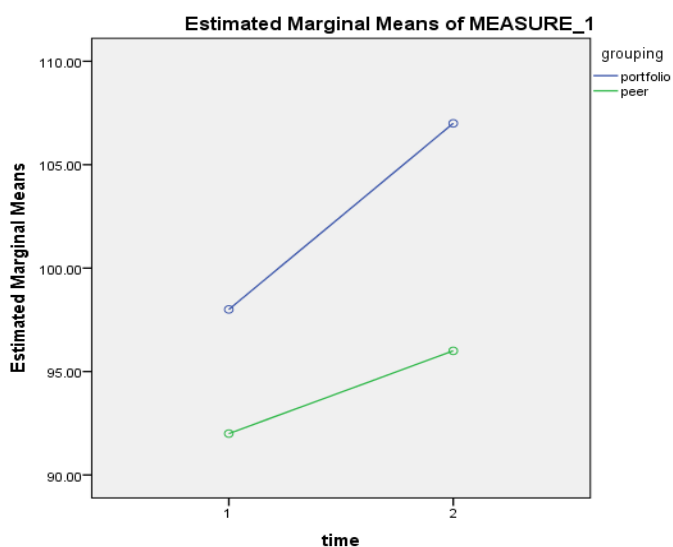


Figure 1. Scatter plot of the CT scores across time for both groups.

As figure 1 demonstrates visually, the two groups developed their critical thinking after the treatment shown by the rising lines from bottom left to the top right; and as the lines do not cross each other at any point, the lack of interaction effect is visually ensured.

4.3. Answer to the Fourth Question

To test the fourth hypothesis, related to the comparative effect of the two treatments on the learners' speaking ability, firstly, the researchers had to make sure that the two groups were homogeneous regarding their speaking ability at the outset. To legitimize the use of a t test to compare their pre-treatment speaking scores, the normality condition was checked. The following table shows the result of the independent t test on the pre-treatment speaking mean scores:

Table 5.
Independent Samples Test on the Speaking Pretest

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Pretreatment speaking	Equal variances assumed	.980	.330	3.273	30	.003	2.37500	.72565	.89303	3.85697
	Equal variances not assumed			3.273	29.497	.003	2.37500	.72565	.89197	3.85803

Table 5 displays that the difference between the pre-treatment speaking scores of the two groups was significant ($t=3.27, p=.003<.05$). As such, ANCOVA had to be used to remove the effect of this initial difference from the posttest scores.

To make sure that the use of ANCOVA is legitimate, firstly the assumption of Homogeneity of regression slopes was checked both graphically and statistically. The following graph was generated for the visual check:

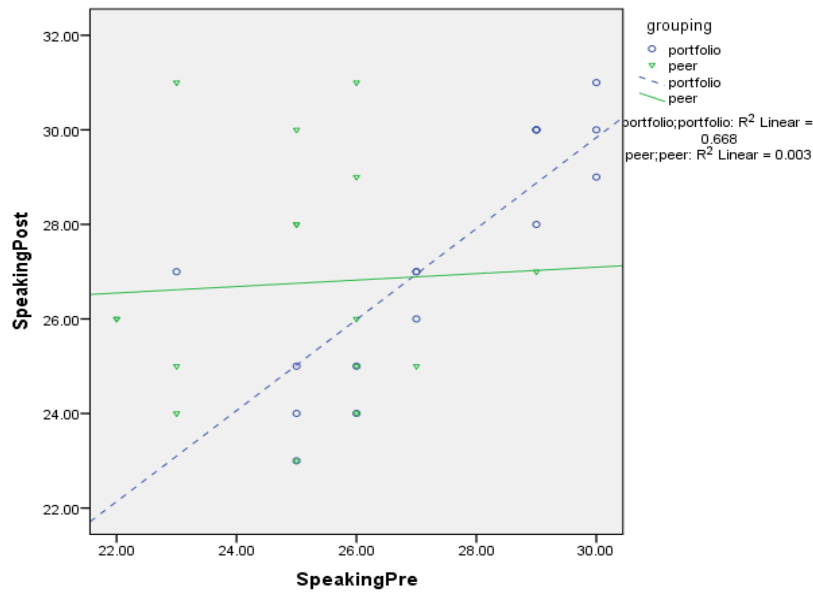


Figure 2. Scatter plot of the interaction between the covariate and the treatment for both groups.

Firstly, what graph 2 shows is the linear relationship for both groups as there are straight lines. Secondly, the two lines are very different in their slopes indicating the violation of the assumption. The following table was also used to verify this assumption statistically:

Table 6.

Tests of Between-Subjects Effects

Dependent Variable: Speaking Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	68.420 ^a	3	22.807	5.005	.007
Intercept	30.916	1	30.916	6.785	.015
Grouping	26.499	1	26.499	5.816	.023
SpeakingPre	32.987	1	32.987	7.240	.012
grouping * SpeakingPre	24.827	1	24.827	5.449	.027

As the sig value for the grouping* speaking pre- test turned out to be less than .05 (.027), as shown in table 6, the conclusion is that the interaction was significant which ensures the violation of the assumption. Therefore, the researchers resorted to the statistical procedure of comparing the *gain scores* of the two groups as an alternative method when ANCOVA is not in place. The following table shows the descriptive statistics of the speaking gain scores of the two groups:

Table 7.

Descriptive Statistics of the Gain Scores Across Groups

	N	Mean	Std. Deviation		Skewness		Ratios
	Statistic	Statistic	Statistic	Statistic	Std. Error		
PortGain	16	-.0625	1.48183	1.245	.564	2.20	
PeerGain	16	2.0625	2.90904	.115	.564	.20	
Valid N (listwise)	16						

As Table 7 exhibits, the peer group outperformed the portfolio group by virtue of the higher gain score (2.06 vs. -.06). Also, the distribution of the gain scores belonging to the portfolio group turned out to be skewed as the ratio exceeded 1.96. Therefore, to compare their gain mean scores, the non-parametric Mann-Whitney U test was employed:

Table 8.

Ranks of Gain Scores

	grouping	N	Mean Rank	Sum of Ranks
GainScores	portfolio	16	13.03	208.50
	peer	16	19.97	319.50
	Total	32		

Table 8 shows that the peer group obtained a higher mean rank (319.5) than the portfolio group (208.5). The following table shows if the difference was statistically significant:

Table 9.

Mann-Whitney U test

	GainScores
Mann-Whitney U	72.500
Wilcoxon W	208.500
Z	-2.112
Asymp. Sig. (2-tailed)	.035
Exact Sig. [2*(1-tailed Sig.)]	.035 ^b

As Table 9 depicts, the asymptotic sig value turned out to be .035, and as this value is less than .05, the conclusion is that the difference between the mean ranks was significant and the peer group improved their speaking significantly more than the portfolio group. Thus the null hypothesis is rejected. The effect size was estimated as 0.37 which is medium size according to Cohen (1988). The following formula suggested by Pallant (2007) was used to calculate the effect size: $r = \frac{Z}{\sqrt{N}}$

5. Discussion

This study employed both between and within group comparisons to find the answers to the research questions dealing with the effect of two alternative assessment types on critical thinking ability of EFL learners and the comparison of the effects on their speaking proficiency.

The fact that both portfolio assessment and peer assessment contributed to critical thinking of the language learners, as shown in this study, suggests that the nature of assessment per se might play a role in enhancing the critical thinking of learners. In the definition of assessment there are certain elements that are shared with critical thinking definition which might explain why portfolio assessment and peer assessment significantly and similarly affected the critical thinking of the language learners. Critical thinking can be defined as the ability to solve complex problems in different forms by asking questions, gathering information and communicative effectively (Paul & Elder, 2006). Hence, the elements of collecting information and problem solving are embedded in this very definition. This means that through the treatment period students learned how to gather information about problems

and how to solve them which might have consequently raised their critical thinking.

Elder and Paul (1994) assert that critical thinking is related to the thinkers' ability in order to take cost of their own thinking and develop proper criteria and standards for evaluating and assessing their own thinking. On the other hand, portfolio and peer assessment can prepare the learners to take the control of their learning more personally and act more autonomously. In the current study, either in peer assessment or portfolio assessment, it was learners who were instructed to assess their performances. In other words, since learners either assessed themselves or their peers they could learn how to be more independent of their teachers and learn about ways they could rely on themselves with regard to their learning and their problems.

One further finding was that peer assessment contributed more to the speaking ability of the language learners than portfolio assessment. The better effect of peer assessment on speaking ability of the language learners can be attributed to the interactive nature of peer assessment. One of the benefits of the peer assessment is that more interaction occurs between the peer (Turner & Purpura, 2015) which partly explains the better contribution of peer-assessment to speaking ability of the participants. Furthermore, peer-assessment has one big advantage that can affect the speaking of the language learners. This advantage is that peer assessment cause less anxiety in language learners (Joo, 2016) and there is no question that anxiety is detrimental to the speaking development of the language learners (Yalçın & İnceçay, 2014). Regarding the 'working together' nature of the treatment, this is congruent with Marandi and Jahanbazian's (2015) finding that team-based learning improves oral performance of Iranian intermediate EFL learners.

6. Conclusion

This study was conducted to investigate and compare the effect of peer and portfolio assessment on EFL learners' critical thinking ability as well as speaking proficiency. The data analyses led to the conclusion that portfolio assessment and peer assessment improved the learners' critical thinking, and there was no significant difference between their effects in this respect. Thus, the first and second null hypotheses were rejected showing large effect size, and the third hypothesis was confirmed. It was further revealed that peer assessment improved the learners' speaking proficiency significantly more than portfolio assessment, hence the fourth hypothesis was rejected showing a medium effect size.

Based on the findings of this study, and on the grounds that critical thinking plays a tremendous role in education success, teachers are specifically advised to shift to alternative assessment to increase the learners' critical thinking. Furthermore, EFL teachers may focus more on peer assessment, when activities pivot on speaking, to increase their learners' speaking ability. In fact, this study showed that speaking performances of students may be assessed by peers and portfolio technique, the former leading to higher speaking achievement. Therefore, for an optimum speaking enhancement on the side of the learners, EFL teachers may choose to invite their learners to be actively involved in the assessing process more specifically through assessing their peers' oral performances.

This study can be replicated with participants at higher levels of language proficiency, as at higher levels learners are expected to show a higher ability to assess and see to their abilities retrospectively. Therefore, the comparison between the two alternative assessment types may be made from their perspective too.

References

- Brown, H. D. (1994). *Teaching by principle: An interactive approach to language pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653-675.
- Chaffee, J. (2009). Thinking critically. *Journal of Marketing Education*, 27(3), 264-276.
- Chaffee, J. (1992). Teaching critical thinking across the curriculum. *New Directions for Community Colleges*, 77, 25-35.
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency, *Language Testing*, 22(3), 93-121.
- Crooks, T. (2001). *The validity of formative assessment*. Paper presented to the British educational research association annual conference, University of Leeds, 13-15.
- Davin, K. J. (2011). *Group dynamic assessment in an early foreign language learning program: Tracking movement through the zone of proximal development*. Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, Pennsylvania.
- Elder, L. & Paul, R. (1994). Critical thinking: Why we must transform our teaching. *Journal of Developmental Education*, 18(1), 34-35.
- Ennis, R. (1996). *Critical thinking*. New Jersey: Prentice-Hall.

- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.
- Fahim, M., & Sheikhy, R. (2011). Critical thinking ability and autonomy of Iranian EFL learners. *American Journal of Scientific Research*, 29(1), 59-72.
- 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.
- Flood, J., & Lapp, D. (1989). Reporting Reading Progress: A Comprehension Portfolio for Parents. *The Reading Teacher*, vol. 42, 508-514.
- Garb, E. & Kozulin, A. (2002). Dynamic assessment of EFL text comprehension of at-risk learners. *School Psychology International*, 23, 112–127.
- Gardiner, L. (1995). *Redesigning higher education: Producing dramatic gains in student learning*. New York: Jossey Bass.
- Genesee, F. & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press. pp.98-100.
- Ghahremani, M., & Azarizad, P. (2013). *The Effect of Dynamic Assessment on EFL Process Writing: Content and Organization*, *International Research Journal of Applied and Basic Sciences*.
- Halpern, D. A. (1998). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, 80, 69-74.
- Hamp-Lyons, L. (1996) Applying Ethical Standards to Portfolio Assessment of Writing in English as a Second Language. In M. Milanovic and N. Saville (eds.) *Performance Testing, Cognition and Assessment*. Cambridge: Cambridge University Press.151–162.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: OUP.
- Honey, P. (2000). *Critical thinking questionnaire*. Retrieved from [http://www.Peter Honey Publication. Com](http://www.PeterHoneyPublication.Com).
- Huitt, W. (1998). *Critical thinking: An overview*. Educational Psychology Interactive Retrieved from <http://chiron.valdosta.edu/whuitt/col/cogsys/critthnk.html>. Revision of paper presented at the Critical Thinking Conference sponsored by Gordon College, Barnesville, GA, March, 1993.

- Joo, S., H. (2016). Self-and peer-assessment of speaking. *Teachers College, Columbia University Working Papers in Applied Linguistics & TESOL*, 16 (2).
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16-46.
- Lazaraton, A. (2001). Teaching oral skills. In M. Celce-Murcia (Ed.), *teaching English as a second or foreign language* (3rd ed.) (pp.103-115). USA: Heinle & Heinle.
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290.
- Marandi, M., & Jahanbazian, T. (2015). A study into the effect of competitive team-based learning and 'learning together' on the oral performance of intermediate EFL learners. *Research in English Language Pedagogy*, 3(1), 60-73.
- Marzano, R. J. (2000). Twentieth century advances in instruction. In R. Brandt (Ed.), *ASCD Yearbook, 2000* (pp. 67–90). Alexandria, VA: Association for Supervision and Curriculum Development.
- Miller, L. (2001). A speaking lesson. How to make the course book more interesting? *MET*. 10(2), 25-29.
- Ming, S., & Alias, A. (2010). Investigating Readiness for Autonomy. A Comparison of Malaysian ESL Undergraduates of Three Public Universities. *Reflections on English Language Teaching*, 6(1), 1-18.
- Moon, J. (2008). *Critical thinking: An exploration of theory and practice*. New York: Routledge.
- Naeini, J. (2013). *Graduated prompts and mediated learning experience: A comparative study of the effects of two approaches of dynamic assessment on the reading comprehension of Iranian EFL learners* (Unpublished doctoral dissertation). Islamic Azad University, Science and Research Branch, Tehran, Iran.
- Nassirdoost, P., Mall-Amiri, B. (2015). The Impact of portfolio assessment on EFL learners' vocabulary achievement and motivation. *Journal for the Study of English Linguistics* 3 (1), 38-50.
- Orsmond, P., Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education* 21(2), 239-250.
- Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS*. (Version 12): Allan and Unwin: Sydney.

- Paul, R. W. (1988). Critical thinking in the classroom. *Teaching K-8* (18), 127-148.
- Paul, R., & Elder, L. W. (2006). *The miniature guide to critical thinking concepts and tools*. Retrieved November 10, 2016 From <http://www.criticalthinking.org/files/Concepts-Tools.pdf>
- Paulson, F. L. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.
- Pearce, C. (2009). *From closed books to open doors – West Africa's literacy challenge*. Oxfam International.
- Pemberton, R., & Nix, M. (2012). Practices of critical thinking, criticality, and learner autonomy. *Special Issue of Learning*, 19(2), 79-95.
- Quintana, J. (2011). *PET Practice Tests*. Oxford. Oxford University Press.
- Saito, H., & Fujita, T. (2009). Peer-assessing peers' contribution to EFL group presentations. *RELC Journal*, 40(2), 149–171.
- Santrock, J. W. (2008). *Educational psychology* (3rd ed.). New York: Mac Graw Hill.
- Shohamy, E. (1992). *The power of tests: a critical perspective on the uses of language tests*. London: Longman.
- Slavin, R. E. (1997). *Educational psychology: theory and practice* (5th ed.). Allyn & Bacon, Needham Heights, MA.
- Thomas, G., & Smoot, G. (1994). Critical thinking: A vital work skill. *Trust for Educational Leadership*, 23(1), 34-38.
- Turner, C. E. & Purpura, J. E. (2015). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Baneerjee (Eds.), *Handbook of Second Language Assessment*. Boston, MA: De Gruyter, Inc.
- Valencia, S. W., & Calfee, R. (1991) The development and use of literacy portfolios for students, classes, and teachers. *Applied Measurement in Education*, 4(4), 120-131.
- Valencia, S. W. (1990). A Portfolio Approach to Classroom Reading Assessment: The Whys, Whats, and Hows. *The Reading Teacher*, vol.43, 338-340.
- Wright, I. (2002). *Is that right? Critical thinking and the social world of the young learner*. Toronto: Pipping Publishing.

- Xiaoxiao, L., & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33(1), 24-40.
- Yalçın, Ö & İnceçay, V. (2014). Foreign language speaking anxiety: The case of spontaneous speaking activities. *Procedia - Social and Behavioral Sciences*, pp. 2620 – 2624.
- Yamashiro, A. D., & Johnson, J. (1997). Public speaking in EFL: Elements for course design. *The Language Teacher*, 21, 13–17.
- Yang, N. (2003). Integrating portfolios into learning strategy-based instruction for EFL college students. *IRAL*, 41,293-317.