Assessment of Oral Presentations: Effectiveness of Self-, Peer-, and Teacher Assessments

Ali Mansoori Nejad
Dr., English Department, Ilam University, Ilam, Iran, *ali_mn1986@yahoo.com*

Omer Hassan Ali Mahfoodh
Dr., corresponding author, School of languages, Literacies and Translation, Universiti Sains Malaysia, Malaysia, *omer@usm.my*

This study investigated the effectiveness of self-, peer-, and teacher assessments in assessing EFL students' oral presentations. It also examined consistencies and differences among these three methods of assessment. Students' attitudes towards self-assessment and peer assessment were also investigated. This study is a mixed-methods sequential explanatory research in which both quantitative and qualitative data were collected. Using an adapted rating scale, 60 Iranian students in four advanced English courses were requested to assess their oral presentations and the oral presentations of their peers. Four teachers used the same rating scale to assess all students' oral presentations. To obtain students' attitudes towards self- and peer assessments, a questionnaire was administered before and after the employment of the rating scale. Students who revealed changes in their attitudes were also interviewed. Although the results showed no significant differences in the three assessment methods, the analysis of the mean scores revealed that teachers employed the strictest scoring criteria while peer assessors used the most lax ones. While the results of peer- and teacher assessments appeared to be consistent, no consistencies were discovered in self- and peer- assessments and in self- and teacher assessments. This study highlights that EFL students' involvement in assessing their own and peers' oral presentations can enhance their motivation to learn.

Keywords: oral presentations, methods of assessment, self-assessment, peer-assessment, teacher assessment

INTRODUCTION

This mixed-methods sequential explanatory study examined the effectiveness of three methods of assessment (i.e. self-, peer-, and teacher assessments) in assessing English as a Foreign Language (EFL) students' oral presentations. This topic is seen to be

important for some reasons. First, delving into students' understanding of instructional practices can help teachers to know more about their students' perceptions and beliefs and can pave the way for future studies (Xiao & Carless, 2013). Second, our study was motivated in part by the growing interest among researchers in examining the assessment of spoken communication (McNamara, 2001; Swain, 2001). For example, Sundrarajun and Kiely (2010) highlighted that "over recent decades the attention to speaking skills in the higher education of foreign language curriculum has increased substantially" (p. 101). Third, a recent concern has been directed towards the use of alternative assessments of students' oral proficiency skills (e.g. De Grez, Valcke, & Roozen, 2012). In a clear recommendation for research on comparing the effect of methods of assessing oral presentations, De Grez et al. (2012) have recently argued that "It might be interesting, in future research, to additionally compare self- and peer assessments" (p. 139). Fourth, although oral presentations have been found effective in improving students' competencies and preparing them for real life situations (Wisker, 1994), the assessment of oral presentations in Second Language (L2) has not been adequately examined in previous studies. Fifth, in EFL contexts, Sadeghi and Khonbi (2015) have called for more research on the effectiveness of peer- and self- assessments when compared with teacher assessment. Sixth, examining these three methods of assessment in one single study has not been adequately addressed in previous research. Thus, in our study, we argue that EFL students' involvement in assessing their own and peers' oral presentations can enhance their motivation to learn.

Traditional methods of assessment, such as standardised tests, teacher assessment, and examinations, have recently lost their popularity because they lack the focus on learners' responsibility (Jafarpur, 1991; McNamara, 2001). There is a general agreement among researchers that the multifaceted evaluation of students' learning appears to be more productive than sole evaluation that was dominant in traditional environment (Orsmond, Merry, & Reiling, 2000; Pope, 2005). Currently, teachers in universities are encouraged to "collaborate with their students in terms of the usage of peer and self-assessments" (Pantiwati & Husamah, 2017, p. 187) . Learners' active role in assessment encompasses two forms: self-assessment and peer-assessment. Self- and peer-assessments have lately received great attention because learners' independence and autonomy are more emphasized and these methods of assessment are of pedagogical significance (Patri, 2002). It has also been pointed out that self-assessment could be calibrated to increase and enhance students' active participation in the way they learn and can be viewed as an outlet for students to reflect on their own learning processes and performance (Dochy, Segers, & Sluijsmans, 1999). Self-assessment has been recently used in EFL teaching pedagogy in order "to give more opportunity to the students to reflect on their own learning and progress" (Ratminingsih, Marhaeni, & Vigayanti, 2018, p. 277).

It has been hypothesized that peer-assessment helps students develop skills and abilities that are ignored in educational environment where teacher appraisal is the only method of learning evaluation (Cheng & Warren, 2005). It is also a fundamental step to establish student-centred learning or to promote a flexible and dynamic learning (Oldfield & Macalpine, 1995). Other studies have revealed that peer-assessment has a facilitative role in students' learning (Ballantyne, Hughes, & Mylonas, 2002). Although

peer-assessment has been described by students as challenging (Falchikov, 1986; Kearney, 2013), it is well-established that it "contributes to the development of student learning and promotes ownership of assessment processes" (Bryant & Carless, 2010, p. 3). Peer-assessment has been found to significantly contribute to students' learning process (Carless, 2005). In addition, research has shown that self-assessment is regarded as a viable activity because it makes students highly interested in learning, motivated to be responsible for their learning(Sadler & Good, 2006). Peer-assessment can also motivate students through the comments given by peers (Chen, 2010). Furthermore, Chang, Tseng, and Lou (2012) argued that when students assess their learning products, this can help them obtain self-reflection and improvement.

LITERATURE REVIEW

Although several studies have investigated self-assessment and peer-assessment from various aspects, there are very few studies comparing various methods of assessment. Cheng and Warren (2005), for example, examined the incorporation of peer-assessment in English classes among undergraduate students majoring in engineering in Hong Kong. The students were asked to assess their peers' language proficiency in oral representations, seminars, and written reports. They also studied students' attitudes towards peer-assessment and possible advantages of peer-assessment over teacher assessment. The results showed that there was a low positive attitude towards peer-assessment. Despite their attempt to highlight a noticeable difference between the effectiveness of peer-assessment and teacher assessment, no significant differences were found between marks given by the teachers and those given by the students in the assessment of oral representations. In another study, Saito and Fujita (2009) examined the correlation between teacher assessment and peer-assessment of EFL group presentation and showed a significant similarity between instructor assessment and peer-assessment. In a very recent study, Butler and Lee (2010) explored the effectiveness of self-assessment among 254 students at two public elementary schools in South Korea. They concluded that considering the context of education, students and teachers were of different views towards the effectiveness of self-assessment.

In the EFL Iranian context, only three studies have dealt with the effectiveness of assessment methods: Birjandi and Hadidi Tamjid (2012), Khonbi and Sadeghi (2013), and Sadeghi and Khonbi (2015). Employing a quasi-experimental study with one control group and four treatment groups, Birjandi and Hadidi Tamjid (2012) examined how effectively self/peer-assessments can improve writing skills of EFL Iranian students. They found that the alternative assessments can be highly conducive to students' writing. They also revealed that the employment of self, peer- and teacher assessments altogether improved students' writing. Khonbi and Sadeghi (2013) examined 63 Iranian EFL students' attitudes towards self-, peer-, and teacher assessments. They found that peer-assessment group outperformed self-assessment and teacher assessment groups. The students also showed high positive attitudes towards peer-assessment. In a recent study, Sadeghi and Khonbi (2015) examined the impact of alternative assessments (viz., self-, peer-, and teacher assessments) on the academic achievement among Iranian university students and students' attitudes towards these three types of assessment. Their

findings showed that peer-assessment had a positive influence on EFL Iranian students' academic achievement.

Our survey of literature has revealed that EFL contexts have received little attention in terms of studies on assessment, especially the alternative methods of assessment. Additionally, we found that few studies have examined methods of assessment in EFL contexts. These studies are Sasmaz Oren (2012), Cheng, Rogers, and Wang (2008), Birjandi and Hadidi Tamjid (2012), Wei (2015), and Ünaldı (2016). Furthermore, studies on assessment in EFL contexts have not adequately addressed the effectiveness of peer-, self-, and teacher assessments. Rather, these studies have dealt with the comparison of methods of assessment in EFL writing courses (e.g. Matsuno, 2009), classroom assessment practices in EFL/ESL contexts (e.g. Cheng et al., 2008), assessment of EFL writing performance (e.g. Birjandi & Hadidi Tamjid, 2012), formative and summative assessments of lecturers' teaching (e.g. Wei, 2015), self and teacher assessments in EFL Turkey context (e.g. Ünaldı, 2016), and instructors' views on their practices in teaching writing in EFL/ESL contexts (e.g. Cumming, 2001). Furthermore, few studies have examined the assessment of oral presentations in EFL contexts. For example, in Turkey Sasmaz Oren (2012) investigated the impact of two variables (experience and gender) on the alternative methods (self and peer-assessments) in the process of assessing oral presentations. They reported that female students were more successful than male students. Regarding peer-, self-, and teacher assessments in EFL contexts, studies have only drawn attention to the effect and merits of these three methods of assessment in the assessment of either EFL students' writing performance or general academic achievement. Specifically, the three studies that have focused on the use of alternative methods of assessment in the EFL Iranian context (Birjandi & Hadidi Tamjid, 2012; Khonbi & Sadeghi, 2013; Sadeghi & Khonbi, 2015) have not focused on oral presentations of EFL Iranian students.

Thus, the main purpose of this study is to examine the effectiveness of self-, peer-, and teacher assessments in assessing EFL students' oral presentations. The study also examined consistencies and differences among these three methods of assessment. Students' attitudes towards self- and peer assessments were also investigated in this study.

THE STUDY

This study addresses four research questions:

RQ1: Are there any statistically significant correlation and differences between self- and peer-assessments in assessing EFL students' oral presentations?

RQ2: Are there any statistically significant correlation and differences between self- and teacher assessments in assessing EFL students' oral presentations?

RQ3: Are there any statistically significant correlation and differences between teacher and peer-assessments in assessing EFL students' oral presentations?

RQ4: How do EFL students perceive the use of self- and peer-assessments before and after the application of the rating scale?

METHOD

Research Design

This study is a mixed-methods sequential explanatory research in which data were collected using quantitative and qualitative methods(Creswell & Plano-Clark, 2011). "Conducting mixed methods research involves collecting, analysing, and interpreting quantitative and qualitative data in a single study" (Onwuegbuzie & Leech, 2006, p. 474). In a mixed-methods research, a researcher integrates the findings and draws inferences using both quantitative and qualitative approaches. This study includes two phases: quantitative and qualitative. In the first phase, data were collected using an adapted rating scale used by students and teachers to assess oral presentations. To obtain students' attitudes towards self- and peer assessments, quantitative data were also collected in the first phase using questionnaires (Appendices B and C). In the second phase, interviews were carried out to collect qualitative data concerning changes in students' attitudes towards self-and peer-assessments.

We chose this research design for some reasons. First, the use of both types of data in one single study can help researchers to present strong findings. Regarding this, Cameron (2009) explained that the employment of both quantitative and qualitative methods of data collection in one study can strengthen the design of the study. Second, among the wide range of mixed-methods research designs, the mixed-methods sequential explanatory design is the most straightforward design (Creswell & Plano-Clark, 2011). Third, the use of interviews in the qualitative phase in this study helped the researchers to explain the quantitative results.

Participants

In this study, the participants were 60 Iranian EFL students (36 females and 24 males) and four teachers. The student participants were English students in a foreign language institute in Bushehr, Iran, studying English book, Interchange at the advanced level (C1 according to Common European Framework (CEFR)). Four assessment sessions were held; each session involved 15 students and four teachers. The four teachers (T1, T2, T3, and T4) were selected purposively based on their experience of teaching English in Iran and their familiarity with the context of the study. All the four teachers have been teaching English in EFL Iranian context for more than five years. The four teachers were assumed to attend all four sessions. The teachers themselves were English students for several years before they started teaching. Thus, they were very familiar with the environment of English teaching in Iran. For example, they are quite familiar with the fact that in Iran teachers are supposed to have an authoritarian role; they are responsible for almost any kinds of decisions made in class and students are usually followers of the given rules. Students do not usually play a significant role in making major decisions such as test design, scoring procedures, and pass/fail.

Training Sessions

The students were asked to attend the training sessions in order to practice assessing their own oral presentations and their peers'. The training sessions were held to set up

the rating scale and to remove any possible ambiguity and difficulties in the rating scale. The students were engaged in tasks to practice assessing their own and their peers' oral presentations for five sessions for each group of students; each session took one hour. At the end of each session, one of the teachers explained the categories of the rating scale in detail and guided the students on how to assess their own and their peers' oral presentations. The teacher explained what constituted language proficiency on the basis of different parts in the rubrics. The teacher tried to make the students aware of the fact that language proficiency is not a general unit entity but various discrete parts that include pronunciation, fluency, grammar, vocabulary, and communication strategies. The four teachers involved in the study also participated in the training sessions. Each teacher practiced independently marking students' oral presentations without knowing what grade the other teacher gave. Teachers also attended some meetings to discuss the training sessions and the rating scale. It could be assumed that teachers' marks could be used as reliable benchmark against the marks given by the students. In the post training session, questionnaires were administered to the students in each session.

In the training sessions, the students were made aware of applying all parts of the rubrics in marking oral presentations. Thus, not only did they considered the content of presentations but also the skills of presentation. Although there might have been some discussions among the students regarding the marks and marking procedures outside the class when the study finished, they were asked to mark their peers' oral presentations anonymously. The teachers were requested to keep the students unaware of their peers' marks in order to avoid any emotional involvement among the students.

Data Collection

This study was conducted in four sessions in a foreign language institute in Bushehr, Iran. Data were collected using an analytical rating scale (used to assess students' oral presentations), a questionnaire, and interviews. Oral presentations were done in four sessions; 15 students and one teacher in each session. The students were asked to prepare oral presentations on topics that were similar to those used in the training sessions. An example of these topics was *Describe a journey you made which was special to you*. They were given four days to prepare their oral presentations. On the day of the presentation, each student was given 10 minutes to do the oral presentation. Each presentation was assessed by the teacher and the presenter's peers. Furthermore, each presenter was asked to assess his/her oral presentation after he/she finished the presentation. The analytical rating scale developed and validated by Bonk and Ockey (2003) was adopted because it can yield more consistent and reliable results than holistic ones (Hamp-Lyons, 1991; Perkins, 1983). Using Rasch Model, Bonk and Ockey (2003) supported the validity and reliability of this analytical rating scale (Appendix A). The analytical rating scale consists of five categories: pronunciation, fluency, grammar, vocabulary/content, and communication skills/strategies (refer to Bonk & Ockey, 2003).

After the oral presentations, students' attitudes towards self- and peer-assessments before and after assessments were obtained through a four-item questionnaire which was modified based on Burnett and Cavaye (1980) (see Appendices B and C). Although Burnett and Cavaye (1980) designed the questionnaire to monitor students' attitudes

towards self-assessment, it was partially modified to be applicable to obtain students' attitudes towards peer-assessment as well as self-assessment. The modified of the questionnaire was minor for the purpose of making it suitable for getting data concerning students' attitudes towards peer-assessment. Each questionnaire is comprised of four questions that focus on students' feeling, attitudes, responsibility, and fairness of assessing themselves as well their peers. The questionnaire used before and after oral presentations had the same questions and were used to identify any changes in the students' ideas concerning the use of alternative methods of assessment. Using identical questions in both questionnaires helped the researchers to trace back the underpinning reasons for a particular attitude and belief concerning the use of alternative methods of assessment. Students who showed changes in their attitudes were interviewed using semi-structured interviews which focused on the changes in students' attitudes towards self- and peer- assessments. The interviews had a semi-structured design and continued for around 10-15 minutes. The aim of the interviews was to explore reasons for a change in students' attitudes towards self- and peer- assessments and whether they had any previous experience in self- and peer- assessments before the study.

Data Analysis

On each category of the analytical rating scale, a raw score is assigned based on a scale of 0-4 which includes half-points, with a total of 20 points. Similar to what Bonk and Ockey (2003) did, raw values were doubled to remove half points. Zero was also eliminated by adding one point to each category. These modifications in the rating scale produced a 1-9-point scale. This analytical rating scale is a set of rubrics which the students used to mark their oral presentations and their peers'. The quantitative data obtained from the questionnaires were analysed using both descriptive and inferential statistics. The correlations among teacher, self- and peer-assessments were calculated by Pearson Product Moment correlation coefficient (r) and t-test. The level of significance for the analysis of data was set at 0.05. The qualitative data obtained from the interviews were analysed thematically.

FINDINGS

Research Question 1: Self- and Peer-assessments

Table 1 displays the results of t-test and the coefficients of Pearson's correlation of self- and peer- assessments. The correlation between these two methods was found to be statistically insignificant. This reflects a low amount of consistency between these two methods of assessment. As for calculating the differences between alternative methods of assessment, there were no significant differences between the marks obtained through self- and peer-assessments. Referring to the estimation of mean scores, it was found that peer-assessment was lax compared with the marks of self-assessment. Overall, the mean score of self-assessment was lower than that of peer- assessment. In Table 1, the p value is given in parenthesis and the t-value outside the parenthesis.

Table 1
Pearson's Correlation and T-test of Self- and Peer-Assessments

Correlation coefficient (sig.)	Assessment method	Mean	Standard deviation	t (sig.)
0.18 (0.5)	Self	10.5	3.02	0.76(0.3)
	Peer	13.4	2.01	

Note: *p<0.05

Research Question 2: Teacher and Self-assessments

The results of t-test and coefficients of Pearson's correlation concerning teacher and self-assessments are summarised in Table 2. There is a significant correlation between teacher and self-assessments. This implies that there is a consistency between the results of these two methods. Provided that teacher assessment could be considered as an outside valid benchmark, self-assessment could acceptably appear to be valid. However, the estimation of the difference between these two methods was insignificant, suggesting high degree of no-difference. Regarding the mean score results, the analysis revealed that the means score of self-assessment was higher than teacher assessment. This implies that self-assessment is generally lax when compared with teacher assessment. In Table 2, the p value is given in parenthesis and the t-value outside the parenthesis.

Table 2
Pearson's Correlation and T-test of Self- and Teacher-Assessments

Correlation coefficient (sig.)	Assessment method	Mean	Standard deviation	t (sig.)
0.80 (0.02*)	Self	10.5	3.02	0.67 (0.2)
	Teacher	9.09	1.4	

Note: *p<0.05

Research Question 3: Teacher and Peer-assessments

Table 3 illustrates the estimations of Pearson's correlation and t-test of teacher and peer-assessments. As shown in Table 3, these two assessment methods (peer- and teacher assessments) are significantly correlated. The results indicated a high level of consistency between peer- and teacher assessments. However, the results of t-test did not show a significant difference between these two methods. Considering the mean scores of each method, it can be recognized that peer-assessment gained a quite higher mean score than teacher assessment. This shows that peer-assessors intended to be more lax in using scoring maxims than teacher assessors. In Table 3, the p value is given in parenthesis and the t-value outside the parenthesis.

Table 3
Pearson's Correlation and T-test of Teacher- and Peer-Assessments

Correlation coefficient (sig.)	Assessment method	Mean	Standard deviation	t (sig.)
0.78 (0.00*)	Teacher	9.09	1.4	30.5 (0.0)
	Peer	13.4	2.01	

Note: *p<0.05

Research Question 4: Attitudes towards Methods of Assessment

In this study, one third of the participants was interviewed after the employment of peer- and self-assessments. The remaining students were not interviewed because they did not show a clear and distinct difference in their attitudes towards self- and peer-assessments. Generally, the majority of the interviewed students showed positive attitudes towards alternative methods in assessing oral presentations. The interviewees revealed that self- and peer- assessments were suitable alternatives over the conventional teacher assessment. In Excerpt One, one of the interviewees revealed that both self- and peer-assessments were effective because they enabled the students to evaluate their performance and their peers' performance in oral presentations in a critical manner.

Excerpt One

Effective self- and peer- assessments support students to evaluate their work and their peers' critically. It will support to identify their strengths and weaknesses, which is an essential learning-skill in the classrooms.

However, some of interviewees cast doubt on the frankness of peer-assessment. Although peer-assessment was done anonymously in this study, some students still believed that peer-assessment could not be reliable if it were not conducted anonymously. They argued that friendship could negatively affect the assessment procedures. As shown in Excerpt Two, one of interviewees held a thought-provoking view of peer-assessment by revealing that friendship can influence peer-assessment. In Excerpt Three, one of the interviewees revealed his concerns regarding friendship when he was giving marks on his peer's oral presentations.

Excerpt Two

Although my classmates do not have a clue who had given the poor or good marks, I still think that the reliability of grades in the process of peer-assessment is open to question as friendship influences the assessment. I do not like to threaten my friendship with low score so I give high score to make sure our friendship with other students is safe and sound. I think that this friendship evokes some emotional reaction subconsciously; that is above students' control.

Excerpt Three

I wanted to write a score for my classmate, I was very anxious about our friendship and afraid of ending that with a low score and I completely forgot anonymity of marking; I just wanted to continue my friendship.

The analysis of the interviews has also revealed that some students believed that the marks could not be considered fair and responsible as their sympathetic considerations might have influenced the assessment process. Thus, assessment that does not hurt students' feelings and does not embarrass them in class would cause assessors to give higher scores. This is clearly reflected in Excerpt Four.

Excerpt Four

Evaluating students' performances, I am very careful not to hurt their feelings

because some of my classmates are very squeamish. I do not like to see them depressed because of some low marks.

Furthermore, in our examination of students' attitudes towards self- and peer-assessments of oral presentations, one of the participants claimed that alternative assessment methods were more practical in higher levels with older students. This is shown in Excerpt Five.

Excerpt Five

Older students have a better control over their feelings in giving marks to their peers as well as themselves. Logic has been fully shaped in them. That is why, they are relatively less in danger of being biased in scoring. I think the factor of age should be considered in using these methods of scoring.

When asked why they felt less comfortable in assessing themselves and their peers, one third of the interviewees responded that some students' high reputation in language use in class has overwhelmed the assessment. Furthermore, some of them said that since they felt they had a poor English proficiency, compared to their peers, they were more confused to give a reliable score. On the other hand, they wondered how a student who has been considered a week and less competent student could possibly give himself or herself a good mark. These points of view are reflected in Excerpts Six and Seven.

Excerpt Six

I gave high scores to some peers' performances because the whole class considered them to be good at English in class. In addition, some students were extroverted, this quality constrained the accuracy of measurement. For example, one of the peers was talkative and active in his speech and this influences my marking.

Excerpt Seven

Because I received poor marks for my language use and have been considered less competent by teachers and other students, I think my self-assessment is more consistent with a priori proficiency mark than my present performance. My prior marks have negatively influenced my own marking.

The interviews revealed that in self- and peer-assessments, students did not consider all relevant elements of oral presentations as much as teachers did. For example, the interviewee in Excerpt Eight shows that the student found it difficult to assess the oral presentation of another student whose language proficiency was better than the assessor. This reflects that peer-assessment is highly influenced by the language proficiency of the assessors. Furthermore, Excerpt Nine suggests that peer-assessment can be influenced by misconception: students who were fluent in Persian (mother tongue) were also fluent in English.

Excerpt Eight

I think awarding score to a fluent speaker might make me ignore other elements like vocabulary and grammar. I should say that his or her fluency can overshadow my scoring.

Excerpt Nine

I think the way students marked fluency has been influenced by Persian language. I want to say some student mistakenly think that those students who speak Persian fluently could speak English fluently too. That is why I think our marks, regarding fluency in specific, are representing Persian fluency.

DISCUSSION

This study sought to examine the effectiveness of three methods of assessment that were used in the assessment of oral presentations in Iranian EFL context. The results showed a lack of consistency between self- and peer- assessments. Overall, the consistencies were only found between peer- and teacher assessments and self- and teacher assessments. Regarding this, Magin and Helmore (2001) hold that combining teacher and peer-assessments could produce a reliable measure of oral skills. Unlike the pervious results, the present results showed a large amount of consistency between peer- and teacher assessors.

The second and third research questions were answered through some statistical techniques which showed a positive correlation between teacher and peer and self-assessments which were used for assessing oral presentations. Our findings are in agreement with what has been reported by Sadler and Good (2006) and Sung, Chang, Chiou, and Hou (2005) who found a large amount of consistency between peer/self-assessments and teacher assessment. This similarity can be explained through considering students' education levels, cultural background, assessment rubrics, and assessment training.

Comparing self-assessment with peer-assessment, our study has revealed that self- and peer- assessment scores were not different because the assessment methods involve the students who use specific scales and take advantage of certain rules to draw on different strategies to examine not only their own work and but also their peers'. With reference to the comparison between self-assessment and teacher assessment, unlike De Grez et al. (2012), our study has revealed an aspect of consistency. This can be attributed to the quality and quantity of experience of the teachers who have been in the context for long time and they have been involved in various tasks of assessing students' oral presentations. Likewise, Chang et al. (2012) made a compatible conclusion to the present findings by showing consistency between self- and teacher assessments. Other researchers also discovered a high correlation between self and teacher assessments (Cho, Schunn, & Wilson, 2006; Tsai & Liang, 2009; Tseng & Tsai, 2007). The results showed a consistency between peer/self- and teacher assessments, which was not in line with those found by De Grez et al. (2012). However, our findings re-echoed the findings of Sadler and Good (2006) and Sung et al. (2005). The divergence in the findings in different studies could be attributed to different assessment environment and assessment procedures employed by students. On the other hand, the no consistency was found between self- and peer- assessments. This finding was supported by the findings of Chang et al. (2012).

Discrepancy among results could be compensated through a sufficient amount of

assessment practice. It is based on the assumption that assessment practices would help students sharpen their assessment abilities, which in turn can give rise to more reliable scores. The outcome of the assessment could be negatively influenced by the assessment practices because these practices demand a great amount of effort and time (Ballantyne et al., 2002; Davies, 2000; Miller, 2003; Topping, Smith, Swanson, & Elliot, 2000; Tsai, Lin, & Yuan, 2002; Wen & Tsai, 2008; Yang & Tsai, 2010). The analysis of mean scores illustrated that peer assessors gave the most lax while teachers were the strictest. These findings are parallel to those of Chang et al. (2012): teacher assessment is the strictest followed by self- and peer-assessments. Thus, teachers should pay adequate attention to the assessment practices to avoid adverse impact on the scoring process. According to our results, no meaningful differences were found among the three methods of assessment.

Concerning peer-assessment, Woolhouse (1999) indicated that peer assessors may face difficulty in giving fair grades. The analysis of interviews in our study revealed findings similar to the observations done by Oldfield and Macalpine (1995). They found out that students feel emotionally biased when they give low scores to their peers. Some concepts such as fluency are very subjective in the students' views, which raise variations in the self- and peer-assessments.

Finally, our study has revealed that Iranian EFL students have positive attitudes towards self- and peer- assessments in the assessment of oral presentations. This finding is similar to other studies in the Iranian context (refer to Birjandi & Hadidi Tamjid, 2012; Khonbi & Sadeghi, 2013; Sadeghi & Khonbi, 2015). Yet, Khonbi and Sadeghi (2013) and Sadeghi and Khonbi (2015) showed that Iranian EFL students showed more positive views of using peer-assessment.

CONCLUSION AND IMPLICATIONS

Our study has shown that there was no significant consistency between self- and peer-assessments. Yet, the study revealed that there was a statistically significant consistency in case of peer- and teacher assessments on one hand and between self- and teacher assessments where a high level of correlation was observed. Concerning the differences among the alternative assessment methods, no significant difference was reported. Furthermore, teachers are more oriented to assess students' oral presentations in a narrower range than self- and peer- assessments. This is because the standard deviation of teacher assessment was relatively lower than self- and peer- assessments. Furthermore, our study found that students in an advanced English course appear to feel less comfortable, confident, and qualified to grade their own oral presentations and their peers'. Such uncertainty and unqualified sense show that EFL students have an ambiguous image of what the oral language proficiency is. Being assessed solely by teachers for a long time in the Iranian educational culture has convinced students that language assessment is only a teacher's responsibility. Thus, they feel insufficiently competent to be involved in applying alternative assessments. It is thought that the provision of more assessment opportunities for students can resolve this misunderstanding that teachers are the dominant language assessors. Nonetheless, both teachers and students confirm the educational benefits of students' engagement in the

assessment process of oral presentations.

Our study provides teachers and students with some important procedural issues of employing various assessment methods. The inclusion and exclusion of some assessment methods in educational programs should not be only determined according to the degree of consistency between self-, peer-, and teacher ratings (Cheng & Warren, 2005). Rather, it should be based on the considerations of the effective influence of self- and peer- assessments. Our study has also addressed the assessment of oral presentations and students' attitudes towards the practice of the two alternatives (self- and peer assessments). Yet, more research is needed to examine the effectiveness of self- and peer- assessments in other skills. It is also worth investigating how to promote positive attitudes towards self-, peer-, and teacher assessments. Similar studies in the EFL contexts with a larger sample size should be done to observe whether providing an inspiring learning atmosphere could result in a positive attitude towards these assessment methods.

REFERENCES

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education, 27*(5), 427-441.

Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education, 37*(5), 513-533.

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20*(1), 89-110.

Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice, 9*(1), 3-15.

Burnett, W., & Cavaye, G. (1980). Peer assessment by fifth year students of surgery. *Assessment in Higher Education, 5*(3), 273-278.

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Cameron, R. (2009). A sequential mixed model research design: Design, analytical and display issues. *International Journal of Multiple Research Approaches, 3*(2), 140-152.

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice, 12*(1), 39-54.

Chang, C.-C., Tseng, K.-H., & Lou, S.-J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. *Computers & Education, 58*(1), 303-320.

Chen, C.-h. (2010). The implementation and evaluation of a mobile self-and peer-assessment system. *Computers & Education, 55*(1), 229-236.

Cheng, L., Rogers, W. T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education, 33*(1), 9-32.

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

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98*(4), 891.

Creswell, J. W., & Plano-Clark, V. L. (2011). Choosing a mixed methods design. In J. W. Creswell (Ed.), *Designing and conducting mixed methods research* (pp. 53-106). London: Sage.

Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? *Language Testing, 18*(2), 207-224.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International, 37*(4), 346-355.

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129-142.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education, 24*(3), 331-350.

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

Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-276). Norwood, NJ: Ablex.

Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation & Research in Education, 5*(3), 145-157.

Kearney, S. (2013). Improving engagement: the use of 'Authentic self-and peer-assessment for learning'to enhance the student learning experience. *Assessment & Evaluation in Higher Education, 38*(7), 875-891.

Khonbi, Z. A., & Sadeghi, K. (2013). Self-, peer-, and teacher-assessment: An investigation into Iranian EFL students' attitudes. *Studies in Second Language Learning and Teaching, 3*(1), 87-107.

Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? *Studies in Higher Education, 26*(3), 287-298.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1), 075-100.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333-349.

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

Oldfield, K. A., & Macalpine, J. M. K. (1995). Peer and self-assessment at tertiary level: An experiential report. *Assessment in Higher Education, 20*(1), 125-132.

Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report, 11*(3), 474-498.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 25*(1), 23-38.

Pantiwati, Y., & Husamah. (2017). Self and peer assessments in active learning model to increase metacognitive awareness and cognitive abilities. *International Journal of Instruction, 10*(4), 185-202.

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

Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly, 17*(4), 651-671.

Pope, N. K. L. (2005). The impact of stress in self-and peer assessment. *Assessment & Evaluation in Higher Education, 30*(1), 51-63.

Ratminingsih, N., Marhaeni, A., & Vigayanti, L. (2018). Self-assessment: The effect on students' independence and writing competence. *International Journal of Instruction, 11*(3), 277-290.

Sadeghi, K., & Khonbi, Z. A. (2015). Iranian university students' experiences of and attitudes towards alternatives in assessment. *Assessment & Evaluation in Higher Education, 40*(5), 641-665.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment, 11*(1), 1-31.

Saito, H., & Fujita, T. (2009). Peer-assessing peers' contribution to EFL group presentations. *RELC Journal, 40*(2), 149-171.

Sasmaz Oren, F. (2012). The effects of gender and previous experience on the approach of self and peer assessment: a case from Turkey. *Innovations in Education and Teaching International, 49*(2), 123-133.

Sundrarajun, C., & Kiely, R. (2010). The oral presentation as a context for learning and assessment. *Innovation in Language Learning and Teaching, 4*(2), 101-117.

Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a web-based self-and peer-assessment system. *Computers & Education, 45*(2), 187-202.

Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing, 18*(3), 275-302.

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

Tsai, C.-C., & Liang, J.-C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science, 37*(3), 293-310.

Tsai, C.-C., Lin, S. S., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education, 38*(1-3), 241-252.

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*(4), 1161-1174.

Ünaldı, İ. (2016). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education, 41*(1), 67-80.

Wei, W. (2015). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment & Evaluation in Higher Education, 40*(4), 611-623.

Wen, M. L., & Tsai, C.-C. (2008). Online peer assessment in an inservice science and mathematics teacher education course. *Teaching in Higher Education, 13*(1), 55-67.

Wisker, G. (1994). Innovative assessment: Peer group and oral assessment. *Programmed Learning and Educational Technology, 31*(2), 104-114.

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

Xiao, Y., & Carless, D. R. (2013). Illustrating students' perceptions of English language assessment: Voices from China. *RELC Journal, 44*(3), 319-340.

Yang, Y.-F., & Tsai, C.-C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*(1), 72-83.

APPENDIX A

Analytical rating scale (Bonk & Ockey, 2003)

	Pronunciation	Fluency	Grammar	Vocabulary/content	Communicative skills/strategies
4.0	Rarely mispronounces, able to speak with near native like pronunciation	Near-native like fluency, effortless, smooth, naturally rhythm	Uses high level discourse structure, with near native-like accuracy	Wide range of vocabulary with near native-like use, vocabulary is clearly appropriate to express opinions	Confident and natural, asks others to expand on views, shows ability to negotiate meaning, shows how own and others' ideas are related
3.5	Pronunciation is clear, occasionally mispronounces some words, but has mastered all sounds, accent may sound foreign but does	Speak with confidence, but has some unnatural pauses, some errors in speech rhythm, rarely gropes for words	Shows ability to use full range of grammatical structures but makes some errors, errors do not impede the meaning of the utterances.	Lexis sufficient for task although not always precisely used	Generally confident, responds appropriately to others; opinions, shows ability to negotiate meaning
3.0	not interfere with meaning				
2.5	Pronunciation is not native-like but can be understood, mispronounces unfamiliar words, may not have	Speech is hesitant, some unnatural rephrasing and groping for words	Relies mostly on simple (but generally accurate) sentences, has enough grammar to express meaning, complex	Lexis generally adequate for expressing opinions but often used inaccurately	Responds to others, shows agreement and disagreement to others' opinions
2.0	Mastered some sounds.		sentences are used but often inaccurately		
1.5	Frequently mispronounces, accent often impedes the meaning, difficult to understand even with concentrated listening	Slow strained speech, constant groping for words, and long unnatural pauses	Uses simple inaccurate sentences and fragmented phrases, doesn't have enough	Lexis not adequate for task, cannot express opinion	Does not initiate interaction, produces monologue only, shows some turn taking,
1.0	Frequently mispronounces, heavy accent, may use	(except for routine phrases) fragments of speech that are so halting that	grammar to express opinions clearly, only says a few words, cannot make	Little lexis, inadequate for Simple communication	may say " I agree with you" but does not relate ideas in explanations, may require prompting, shows no awareness of other speakers
0.5	Persian like-speech which is virtually not comprehensible	conversation is virtually impossible	reasonable judgments of student's grammatical ability		
0.0	Does not discuss	Does not discuss	Does not discuss	Does not discuss	Does not discuss

APPENDIX B

Peer-assessment: (pre-questionnaire)

Name:⎯⎯⎯⎯⎯Class: ⎯⎯⎯⎯⎯

How do you feel about peer assessment? Answer the following questions by circling your answer to each question.

1. Do you think students should take part in assessing their peers?
   A. Yes. B. No. C. Not sure.
2. Do you believe a student should be able to assign grades to peers in a responsible manner?
   A. Yes. B. No. C. Not sure.
3. Do you think you will feel comfortable in making peer assessments?
   A. Yes. B. No. C. Not sure.
4. Do you think you will make a fair and responsible assessment of your peers?
   A. Yes. B. No. C. Not sure.
   Thank you.

APPENDIX C

Peer-assessment: (post-questionnaire)

Name:_____ Class: _____

You have assessed your peers' performance in an oral task.

What are your feelings about peer assessment when you think back on it? Answer the following questions by circling your answer to each question.

1. Do you think students should take part in assessing their peers?

    A. Yes.   B. No.   C. Not sure.

2. Do you believe a student should be able to assign grades to peers in a responsible manner?

    A. Yes.   B. No.   C. Not sure.

3. Did you feel comfortable when you made peer assessments?

    A. Yes.   B. No.   C. Not sure.

4. Do you think you have made a fair and responsible assessment of your peers?

    A. Yes.   B. No.   C. Not sure.

Thank you.