

Self- and Peer Assessment of Oral Presentations: Investigating Correlations and Attitudes

Mohammad Salehi¹, Behnam Daryabar²

¹ Assistant professor, Sharif University of Technology, M_Salehi@sharif.edu

² MA holder, Sharif University of Technology, Behnam_Daryabar@yahoo.com

Abstract. Despite the proven advantages of alternative assessment including self- and peer assessment, there is still uncertainty about the accuracy of these techniques especially in Iran since little research has conducted in this area. To this aim, a total of 2261 evaluation sheets were collected to investigate 118 peer assessments and 48 self-assessments of oral presentations in comparison with their teacher assessments. The data were collected during four semesters in general English classes held in Sharif University of Technology. Students were supposed to make an oral presentation for at least twenty minutes on a topic of their interest, and were then rated by their teacher, themselves, and their peers based on ten pre-established assessment criteria. The number of peer assessments for each presentation was within the range of 8-27. In addition, in order to delve into the students' attitudes toward the new techniques, 64 students expressed their opinions on five-point Likert scale surveys before and after the implementation of self- and peer assessment. It was indicated that the results of the peer assessment were more consistent with those of teacher assessment and the resemblance grew when the number of peers was within the range of 16-19. On the other hand, it was observed that the students' attitudes significantly changed in a positive way toward both techniques. All in all, the findings of the present study suggested that peer assessment closely resembled teacher assessment while the students' attitude changed more positively toward self-assessment in comparison with peer assessment. The findings of this study could have implications for language teachers, material designers, and educational policy makers.

Keywords: *Self-assessment; Peer assessment; Alternative Assessment; Attitude*

1. Introduction

The pivotal role of assessment in education cannot be neglected since it can help determine the learning success. For centuries, teachers decided on the criteria, standards, and the way students should be assessed (Leach, Neutze & Zepke, 2001). Nowadays, this phenomenon is changing and assessment partnership for adult learners is gaining momentum increasingly. In recent years various non-traditional forms of assessment, which are often referred to as alternative assessments (student portfolios, self-assessments, peer-assessments, and so forth), seem to be replacing the traditional achievement tests. However, the application of such alternative assessments appears to vary tremendously from teacher to teacher (Butler & Lee, 2010).

Many language researchers have been encouraged to scrutinize whether learners are able to make a meaningful contribution to their own evaluation. In language teaching, self-assessment (also termed self-rating, self-evaluation or self-appraisal) is often used to promote student-centered learning, to increase insight into the learning process, to encourage active learning, to save the teacher correction time and to support students in dealing with often very individual weaknesses (Dlaska & Krekeler, 2008). It would be ideal if the findings of all the empirical studies on the effectiveness of self-assessment were consistent; however, they have produced inconsistent results which have made test administrators and teachers skeptical about implementing self-assessment in the assessment of second language ability (Ross, 2006). Peer

assessment is a process wherein a group of individuals grade their peers and students either involved in the development of assessment criteria or not (Falchikov, 1995). The quality of the learning process can be enhanced by the use of peer assessment which can help develop self-direction and autonomy (Morris, 2001).

Despite the multiple advantages of new forms of assessments, students often make the assumption that evaluations made by teachers are more accurate, which may not always be true. As a matter of fact, teacher assessment cannot always be viewed as the most valid method; as a result, it can be triangulated with varied sources of assessments (Orsmond and Merry, 1996). On the other hand, there are still controversies about the accuracy of self- and peer assessments.

Although previous studies have addressed peer assessment exclusively, almost a few examined the self- and peer assessment together as well as the students' attitudes after experiencing these two techniques. The present study investigated students' self-assessments and peer assessments in comparison to traditional teacher assessment in oral presentations. It focuses on oral presentations in an EFL context in that speaking has been assigned a prominent role in recent types of proficiency tests (e.g. TOEFL iBT). Speaking has also formed backwash effect for motivating English learners to improve their oral skills (Peng, 2010). The correlation between teacher assessment and self-assessment as well as the correlation between teacher assessment and peer assessment was probed. Furthermore, the attitudes of learners toward the new forms of assessments were examined through pre-and post-survey questionnaires.

2. Literature review

Palomba and Banta (1999, p. 4) defined assessment as the following: "Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development". It is the process of collecting information to monitor progress and draw educational decisions. Assessment is a broad term that includes testing. A test is a special form of assessment. Tests are assessments made under artificial circumstances especially for administration. In other words, an assessment may include a test, but also it can embrace methods such as observations, interviews, behavior monitoring, etc. (Overton, 2009).

Dochy, Segers, and Sluijsmans (1999) emphasized the idea of using assessment as a learning device in a way that one can switch from a teacher-centered classroom to a learner-centered classroom. In this context teachers not only monitor learning but also improve it. Orsmond and Merry (1996) stated that in order to encourage students to be more self-dependent in their own development we need to change the current tests. To this end, we need to empower students, two instances of which are self- and peer assessment.

Brown (1996 p. 291) defines peer assessment as "any items wherein students are asked to rate each other's knowledge, skills, or performance". The ratings can be performed either holistically or analytically. In the case of analytic peer assessment, students rate their peers based on several criteria. The assessment criteria are developed by the teacher before the class or by the students during the class, whereas in the holistic peer assessment students only give one score in order to rate their peers. Self-assessment, as can be inferred from its name, is a technique that required people to rate their own ability. Brown and Hudson (2002) categorized the self-assessment technique into three types in language learning contexts; (1) performance ability self-

assessment whereby learners are required to judge how well they or their peers would respond in a particular situation. (2) Comprehension self-assessment in which learners will judge their comprehension of a situation (3) observation self-assessment whereby learners might judge how well they have performed in a situation specially by responding to that performance for a later judgment. The same terminology is also applicable to peer assessment as well. Regarding the peer-assessment specifically, Satio and Fujita (2004) assigned two names for it on the basis of the purpose for which they are being used: peer review and peer-rating. The former is particularly used for review purposes (e.g., writing revision), and the latter is especially used for assessment purposes. The use of peer review as part of the assessment process can also exert a positive impact upon the quality of the learning process, and can help develop self-direction and autonomy (Morris, 2001).

Despite some demerits of self-and peer assessment, the merits of these techniques in learning cannot be neglected in high stakes decision-makings. In fact, a comprehensive review of the literature of self-and peer assessment shows that the advantages of self-and peer-assessment outnumber their disadvantages. For instance, they help to create a closer relationship among teachers and students. They also foster critical thinking, communication, feedback, responsibility, autonomy, and help students develop useful skills in academic and professional areas (Peng, 2010). Dłaska and Krekeler (2008) also pointed out that self- and peer assessment seem to be a viable option to enhance independent and autonomous learning. They stipulated that self-assessment is used to promote student-centered learning, to increase insight into the learning process, to encourage active learning, to save the teacher correction time and to support students in dealing with often very individual weaknesses.

Another point worth mentioning here is the role of social processes. Topping (2003) maintained that social processes play a dominant role which can affect the reliability and validity of peer assessment rather than self-assessment. Among these social processes which cannot be found in self-assessment contexts are friendship among students, group popularity among individuals, and power processes. Topping (2003) also asserts that peer-assessment is an interpersonal activity and came into conclusion that peer assessment has a far higher correlation with professional assessment than with self-assessment. In order to explain why peer-assessment yields a higher reliability than self-assessment, Falchikov and Goldfinch (2000) provide reasoning as follows:

Self-assessment is usually a private activity which may involve little or no knowledge of the work or performance of others. However, many of the peer-assessment studies which make up the present corpus involve assessment of oral presentations or professional practice in a group context. Thus, the act of assessment takes place within a public domain where comparison between performance become possible and ranking of peers becomes less difficult for students. (p.317)

Satio and Fujita (2004) reviewed previous studies for possible factors that may affect self-and peerassessment. These factors are acquiescence effect (a tendency to agree with items), task specificity, cultural norms, proficiency, and anxiety. It is worth of note that the above mentioned factors which are considered to affect self-and peer-assessment in practice are actually general factors. In other words, these factors can influence any rating regardless of the nature of rating (Fletcher & Baldry, 1999, as cited in Satio & Fujita, 2004). In this regard,

practicing self-and peer assessment necessitates a better strategy by setting a balance between the advantages and disadvantages.

Ultimately, as noted by M. Osearson (1989), self-and peer assessments are not suitable to be used as single criteria for evaluation and they should be accompanied by other types of assessments. In other words, a very minor role of self-and peer assessments is assessment since these educational techniques are at the service of learning.

Research Questions

The present study was aimed at answering the following questions:

1. Is there any relationship between teacher assessment and peer assessment of oral presentations?
2. Is there any relationship between teacher assessment and self-assessment of oral presentations?
3. Do students change their perceptions after experiencing self- and peer-assessment techniques in the context of oral presentations?

3. Method

3.1. Participants

The study was conducted at Sharif University of Technology in four semesters. The sampling was convenient. The students were from 10 different intact classes which were structured with the same teacher and a textbook called 'Active 4'. The students' proficiency level ranged from intermediate to high-intermediate. They attended the class two sessions a week, while each session lasted one and a half hours. . They were in the age range of 18 to 20 coming from different colleges such as the colleges of engineering, or sciences such as chemistry, physics, mathematics, etc. The proportion of male and female was 65/35 in favor of male students.

a. Phase One

In the first two semesters, the study focused on the relationship between teacher assessment and peer assessment. Out of 95 students, 70 of them delivered oral presentations and were rated by their peers and the teacher. In this phase, 1199 evaluation sheets were collected which means every presenter was averagely evaluated by 16 peer assessors.

b. Phase Two

In this phase which occurred during the third and fourth semesters, not only students participated in peer-assessment but also they rated themselves as a self-assessment job. Their change of attitudes toward the new techniques was also examined in this phase. Out of 96 students, 48 self- and peer assessments were gathered from four classes. Each presenter was averagely evaluated by 16 peer assessors. A total of 1068 evaluation sheets were collected in this phase. Apart from the rating part, 64 students filled out both pre- and post-survey questionnaires of attitude measurement.

3.2. Instruments

a. *Pre-and Post-survey Questionnaire*

A five-point Likert scale survey was employed to investigate college students' perceptions toward self-and peer-assessment. The participants were asked to fill out the same survey at the beginning (pre-survey) and at the end of the semester (post-survey); so that the comparisons could be made. The survey was adopted from Wen, Tsai and Chang's (2006) and Peng's (2010) study. Wen et al. reported that Cronbach alpha reliability for the general peer assessment was .84. For the purpose of this study, a few items were modified. The items of the survey were classified under the headings of learning (2 items), social interaction (2 items), students' eligibility (2 items), motivation (2 items), and speaking (1 item).

The same nine items for peer assessment were used for self-assessment with the change of wordings. The reason was that based on the literature these two techniques are closely related. Then, the self- and peer assessment questions were combined, forming an 18-item questionnaire, in a way that odd numbers belonged to peer assessment questions and even numbers to self-assessment questions. The rationale behind this kind of sorting was that the respondents could better contrast these two techniques.

b. *Evaluation sheet*

A ten-item five-point Likert scale questionnaire from Peng (2010) was utilized as the main criterion for the teacher, self, and peer evaluation. The questionnaire was divided into two parts of content and delivery. The content was divided into five components; structure of presentation, evidence of rehearsal, pronunciation and clarity of expression, appropriate/accurate use of grammar and vocabulary, and quality of the content. The delivery part was also divided into: visual aids, interaction with the audience, confidence, timing and pacing, and eye contact/ voice/ gestures movements.

3.3. Procedures

The data collection procedure was conducted in four semesters with 10 different intact classes. During the first two semesters, the students only experienced peerassessment of oral presentations. Throughout the second two semesters, students not only practiced peer-assessment but also exercised self-assessment. Furthermore, their attitudes toward the new techniques were examined by pre- and post-survey questionnaires.

a. *Rating Oral Presentations*

At the beginning of the first and second semester, the teacher explained the objectives and procedures of peer assessment and at the beginning of the third and fourth semester, the teacher explained the same purposes and procedures for peer assessment as well as self-assessment. In the next session, he discussed issues or concerns of students regarding self- and peer assessment. For example, students might be concerned that assigning grades to friends would jeopardize their friendship (Peng, 2010). Thus, he assured the students that their names will be kept confidential. In the third session, the instructor had to discuss various components of a good oral presentation. From session four to the end, in every session, one student was supposed to give oral presentation along with PowerPoint slides. While the presenters were presenting, the students were provided with the evaluation sheets. Then, they had a few minutes to discuss the performance they just saw and assign a grade from 1 (poor) to 5 (excellent) to the

presenter. After the presentation, the presenter, filled in the peer evaluation form as a self-assessment task along with the teacher as a teacher assessment job.

b. Pre-and Post-survey administration

In the first session of the third and fourth semester, the instructor introduced the syllabus and informed the class about implementing the new forms of assessment. Subsequently, he distributed the pre-survey questionnaire to the students. Students were supposed to write their names on the questionnaire and whether they had experienced self-and peer-assessment. The same questionnaire was filled out by the same students at the end of the two semesters. Via this method, the researchers discovered the changing attitudes of the students after they experienced self- and peer assessment techniques.

4. Results and discussion

In order to find the answer to the first and second research questions of the study, two Pearson correlations were run. The first correlation was run to compare the mean of students' scores with the teacher' scores and the second correlation was employed to compare the teacher' scores with the self-assessment scores (*Table1*).

Table 1. *The Pearson correlations between teacher assessment vs. self- and peer assessment*

Column1	Column2	N	Correlation
Teacher Assessment	Peer Assessments	118	.679**
Teacher Assessment	Self-Assessment	48	.462**

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

The results show that there is a significant relationship between teacher assessment and peerassessment, $r = .679$, p (two tailed) $< .01$. The findings are consistent with those of most previous researches such as the meta-analysis by Falchikov and Goldfinch (2000) which employed 48 quantitative peer assessment studies. In their finding, the correlation of peer and teacher marks varied from 0.14 to 0.99 and the overall mean correlation was reported $r = 0.69$.

Figure1 illustrates that the range of scores which are extracted from peer assessments are mostly located in the middle of the graph. This signifies that no matter how high or low the teacher assessment was, the students assessed their peers in a moderate fashion. For instance, in assessment numbers 9, 19, and 40, the teacher assessed the presenter very low, but students were not that much critical. The opposite also occurred when the teacher scored the students very high. In other words, when teacher was meticulous in giving scores, the students were not and in cases the teacher was generous with the scores, the students were excessively critical.

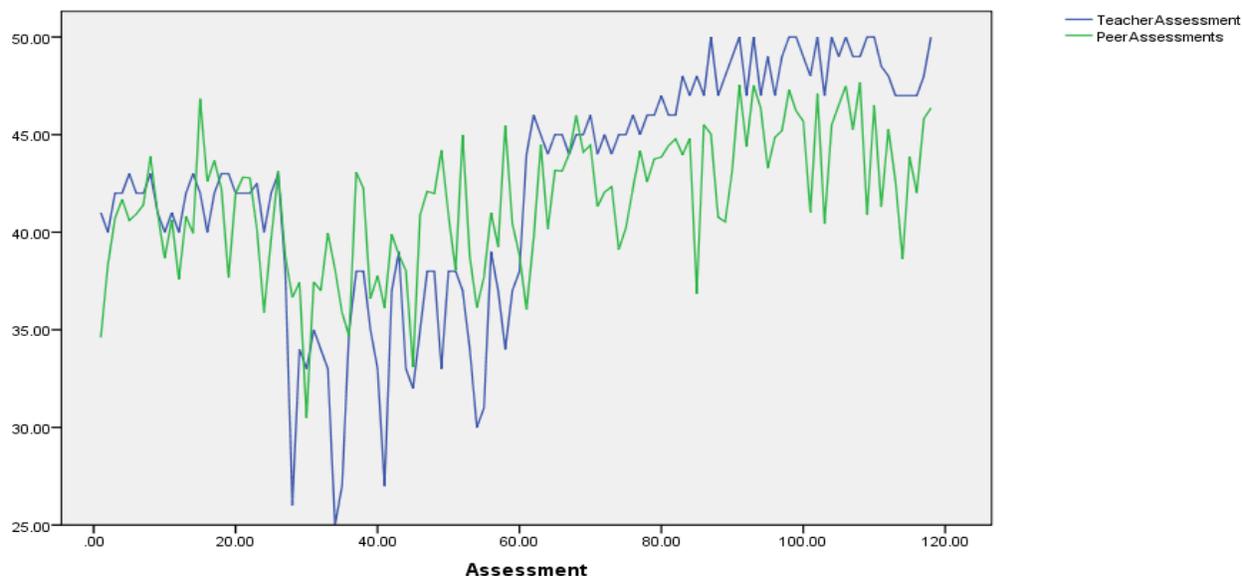


Figure 1. The graph of teacher scores and the mean of students’ scores in peer assessments

This finding is in concordance with that of Alfalay (2004) who claims that students usually overrate the performance of their classmates. Patri (2002) also reported that the peers were overrating low-ability students; a similar finding is mentioned as well by Falchikov (1995). On the other hand, cases where the students were very critical toward their peers might be interpreted as lack of competence or even as a sign of jealousy toward the performance of their peers, which needs further research.

In calculation of the Pearson moment calculation between teacher and peer marks, the number of peers varied from 8 to 27 with an average of 17 peers per each presentation. Since the number of peers was changing dramatically from one assessment to another, the authors decided to divide all 118 peer-assessments into three equal groups. That is 8-15 peers, 16-19 peers, and 20-27 peers (Table 2)

Table 2. Number of peers involved in peer-assessments

Range of peers	8-15	16-19	20-27
Correlation	.614 **	.772 **	.608 **
Number of peers	40	39	39
Percent of peers	33.89%	33.05%	33.05%

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

As displayed in Table 2 , the highest Pearson moment $r = .772$, p (two tailed) $< .01$ was extracted from the teacher and peer assessments where the number of peers ranged from 16 to 19. Falchikov and Goldfinch (2000) employed a meta-analysis using 48 quantitative peer assessment studies. The studies were categorized into four group sizes (1 peer, 2-7 peers, 8-19 peers, and 20+ peers per assessment) and the mean of correlations for each group was calculated. The results revealed that as the number of peers increased, the correlations decreased. The group of 20+ peers per assessment which contained 15 studies had the lowest mean of correlations among other groups $r=0.59$. On the other hand, the highest mean of correlations belonged to the group of 8-19 peers per assessment with 12 studies $r=0.77$. The findings of this study are in line with the Falchikov and Goldfinch’s (2000) meta-analysis study about decreasing the correlation

when the number of peers become 20 and more. However, the highest correlation in this study belonged to the group of 16-19 peers while in their meta-analysis it was reported to belong to the group of 8-19 peers per assessment.

Regarding the second research question, the results revealed that there is a significant relationship between teacher assessment and self-assessment, $r = .462$, p (two tailed) $< .01$. Ross (1998) conducted a meta-analysis on 60 reported self-assessment studies in four different language skills. The average correlation was reported $r = 0.63$, although the correlation for speaking skill ($r = .55$, $N = 29$) revealed that students are less competent in assessing their own speaking performance. In Ross's meta-analysis, the results of speaking criteria were reported to be more homogeneous than other skills, usually in the form of teacher ratings. The correlation statistics of the current study was a bit lower than those of other studies which had been conducted to probe the relationship between teacher's scores and students' self-scores. The followings might be considered as the cause of low correlation result: (1) students didn't have proper training regarding different criteria of the assessment. (2) Most of the students who volunteered for oral presentation could be categorized as advanced students. Matsuno (2009) concluded that high-achieving students assessed themselves lower than predicted. The same phenomenon occurred in our study (Table 3).

Table 3. Students' estimation from self-assessment in relation to teacher assessment

Estimation	Frequency	Percent
Students' underestimation	29	60.4
Students' overestimation	15	31.2
the same score	4	8.3

As can be observed in Table 3, most of the students underestimated their scores compared with that of the teacher (60.4%). In other words, the majority of the students were critical about their own performance in oral presentations. On the other hand, the higher correlation of peer assessment in contrast to self-assessment is in accordance with the Topping's (2003) assertion that peer-assessment yields higher correlation than self-assessment.

In order to find the answer to the third research question of the study, 64 students from third and fourth semesters answered both pre- and post-survey questionnaires. Two paired sample t-tests were run to compare the mean scores of the pre-survey questionnaire with the mean scores of the post-survey questionnaire (Table 4).

Table 4. Descriptive information and scale-score differences between the self- and peer-assessment surveys

Alternative assessment	Pre-survey		Post-survey		t	eta ²	sig
	Mean	SD	Mean	SD			
Peer-assessment (PA)	3.59	.58	3.78	.56	2.41	.04	.019 *
Self-assessment (SA)	3.23	.55	3.57	.55	4.56	.14	.000 **

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

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

The results of the paired sample t-tests (Table 4) reveal that the mean scores of both pre- and post-surveys for both techniques were statistically higher than the value of 3 which is the neutral

attitude. Hence, in both pre and post-surveys students held positive attitudes toward self- and peer assessments in general. Regarding peer assessment, students scored better in post-test ($M=3.74$, $SE=.56$) than in pre-test ($M=3.59$, $SE=.58$), $t(63) = 2.41$, $p < .05$. In order to find the effect size, eta squared statistic was calculated by value interpretation as follows: .01=small effect, .06=moderate effect, .14=large effect (Cohen, 1988). Subsequently, the eta squared statistic (.044) indicated a small effect size for peer assessment survey. On the other hand, the result of the self-assessment survey indicated that students scored better in the post-test ($M=3.57$, $SE=.55$) than in the pre-test ($M=3.23$, $SE=.55$), $t(63) = 4.56$, $p < .05$. The eta squared statistic (.141) indicated large effect size based on Cohen (1988).

Table 5 illustrates the detailed comparisons on score differences for all 9 items between pre- and post-surveys. The items in both self- and peer assessments` pre- and post-surveys can be grouped into several broad categories. Items 1 and 2 are concerned with learning in general while items 3 and 4 pertain to social interactions. Items 5 and 6 relate to students` eligibility to assess their peers. The students could respond to motivation in items 7 and 8 and finally item 9 refers to speaking.

Table 5. The Paired *t*-values of Individual Items for the Students` Responses

Item Category	(N=64)	<i>t</i> -value (sig 2-tailed)	
		Peer assessment	Self-assessment
Learning,			
1. AA is helpful to your learning.		2.20 (.031) *	2.83 (.006) **
2. AA makes you understand more about teacher`s requirement		.86 (.391)	2.03 (.046) *
Social Interaction			
3. AA activities increase the interaction between the T & Ss		2.10 (.040) *	3.47 (.001) **
4. AA helps you develop a sense of participation.		-.64 (.519)	3.42 (.001) **
Students` Eligibility			
5. I think students are eligible to assess their own & their peers.		2.16 (.034) *	2.48 (.016) *
6. I appreciate to be graded by my own/by my peers		4.09 (.000) **	2.14 (.036) *
Motivation			
7. AA activities motivate you to learn.		.00 (1)	1.15 (.252)
8. Being graded by AA motivates you to participate more in the class.		.814 (.419)	3.39 (.001) **
Speaking			
9. AA betters your oral presentation skills.		-.15 (.877)	3.41 (.001) **

Note. AA = alternative assessment (self- and peer-assessments); * $p < .05$, ** $p < .01$

The detailed comparisons illustrated that not all items were statistically significant. This is the case especially for peer-assessment survey in which only four items were positively significant. Peer assessment is helpful for students` learning (item 1); its activities enhance the interaction between the teacher and students (item 3); students are eligible to assess their peers (item 5); and students appreciate being graded by their peers (item 6). Nonetheless, it is noteworthy that in peer assessment survey, the students` attitude toward two items was negatively changed after

experiencing the technique but the change was not significant. Peer assessment helps students to develop a sense of participation (item 4), and better students' oral presentation skills (item 9).

In the self-assessment survey, all items were significant except one item: self-assessment activities motivate students to learn (item 7). Needless to say, more items were significant $p < .01$ in self-assessment survey (5 items) compared with one significant $p < .01$ item in peer-assessment survey.

All in all, in both self- and peer-assessments survey, four items were statistically significant in a positive way: the self- and peer-assessments were seen helpful for learning (item 1), they increase the interaction between the teacher and students (item 3), students think they are eligible to be evaluated by their own and their peers (item 5), and they appreciate to be graded by these techniques.

5. Conclusions

Self- and peer assessment are to some extent similar to teacher assessment but they will not ever be as accurate as teacher assessment. It must be noted that reaching high reliability and validity is not the primary objective of self- and peer assessment. As mentioned by Devenney (1989), the goals and functions of self, peer, and teacher assessment are different; teacher assessment is a summative assessment technique which is mostly used for evaluation at the end of the courses while self- and peer assessment are formative assessment techniques which aimed at ongoing learning processes. One of the main advantages of having several input samples of students is to help teachers understand learning processes and their outcomes. In other words, self- and peer assessment are accompanying tools for students' engagement and empowerment which should be used along with teacher assessment. Shohamy (2001) also stipulated that self- and peer assessment should be used as tools for gathering samples of language from learners. All in all, with careful training, monitoring and utilization, self- and peer assessment can be beneficial as good as teacher assessment in “cognitive, social, affective, transferable skill, and systemic domains” (Topping, 1998, p. 269).

6. Pedagogical Implications

The most genuine implications of the findings of this study are for language teaching and learning. There has always been an attempt to find more efficient techniques and practices for language learning. Given the observations made in this study, self- and peer assessment can turn a class from a teacher-centered classroom to a learner-centered one. As LeBlanc & Painchaud (1985) argues self- and peer assessment could provide a fair means of assessment away from biased domination in comparison to traditional testing. However, while peer assessment in this study resembled more similarity to teacher assessment than that of self-assessment, the student's attitudes were more positive for self-assessment. It must be noted, that in order to fully experience the advantages of self- and peer assessment, the students should have positive attitudes toward using these new techniques.

This study can also have central importance for syllabus designers, and course developers. Material designers can use the findings of the present study to design tasks, activities, and exercises which encourage both teacher and students to conduct self- and peer assessment. Teachers, syllabus designers, and materials developers would not maintain the use of self- and peer assessment if they do not receive the support of educational policy makers. For this aim, localized empirical studies (e.g., the present study) should be considered by policy

makers to encourage syllabus designers and materials developers to include these techniques in their course books.

REFERENCES

- Alfallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System* 32, 407–425.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5–31.
- Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System* 36, 506-516.
- 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. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, 32(2), 175-187.
- 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.
- Leach, L., Neutze, G., & Zepke, N. (2001). Assessment and empowerment: Some critical questions. *Assessment and Evaluation in Higher Education*, 26(4), 293-305.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26 (1), 75-100.
- Morris, J. (2001). Peer assessment: a missing link between teaching and learning? A review of the literature. *Nurse Education Today*, 21, 507-515.
- Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education*, 21(3), 239-250.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rational and applications. *Language Testing*, 6(1), 1-13.
- Patri, M. (2002). The influence of peer feedback on self-and peer assessment of oral skills. *Language Testing*, 19(2), 109-131.

- Peng, J.C. (2010). Peer Assessment in an EFL Context: Attitudes and Correlations. In T. P. Matthew, Y. Watanabe, and S. K. Lee. (Eds.), *Selected proceedings of the 2008 second language research forum* (pp. 89-107). Somerville, MA: Cascadilla Proceedings Project.
- Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. San Francisco: Jossey-Bass,
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research and Evaluation*, 11(10), 1-13.
- Satio, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classroom. *Language Teaching Research*, 8(1), 31-54.
- Topping, K. (2003). Self and peer assessment in school and university: reliability, validity and utility. In M. Segers, F. Dochy, and E. Cascaller (Eds.), *Optimizing new modes of assessment: in search of qualities and standards*, pp. 55-87. Dordrecht: Kluwer Academic Publishers.
- Overton, T. (2009). *Assessing learners with special needs. An Applied Approach. (6th Edition)*. Upper Saddle River, NJ: Pearson, Merrill, Prentice Hall.
- Wen, M. L., Tsai, C. C., & Chang, C. Y. (2006). Attitudes towards peer assessment: A comparison of the perspectives of pre-service and in-service teachers. *Innovations in Education and Teaching International*, 43(1), 83-92.