

Self-assessment impact on EFL learners' speaking fluency and accuracy: Does level of proficiency matter?

Goodarz Alibakhshi¹

Assistant Professor in Applied Linguistics, Allameh Tabataba'i University

Abdullah Sarani

Associate Professor in Applied Linguistics, University of Sistan & Baluchestan

Received on September 6, 2013

Accepted on November 15, 2014

Abstract

Since the 1970s, self-assessment in education has gained increasing currency and has been studied in a considerable number of quantitative studies. In line with this theoretical and experimental background, this study aimed at investigating the effect of self-assessment on intermediate and upper-intermediate language learners' speaking fluency and accuracy in Iran. Thirty pre-intermediate and thirty upper-intermediate students participated in the study in a pretest-posttest control/experimental group design. The data were analyzed through ANCOVA test. The results indicated that self-assessment positively affected participants' speaking accuracy and fluency. Also, self-assessment had a greater effect on upper-intermediate learners than pre-intermediate ones. Moreover, speaking fluency benefitted more from self-assessment than speaking accuracy did. The findings can be applied by both EFL learners and teachers.

¹ Corresponding author: alibakhshi@atu.ac.ir

Keywords: self-assessment, speaking accuracy, speaking fluency, Iranian EFL learners

1. Introduction

Assessment is an integral part of every educational system through which we evaluate learners' achievement on the basis of the materials taught (Alderson, 2005; Mousavi, 2012; Patri, 2002; Stefani, 1998). In a learner-centered pedagogy which considers learners as active participants in education and learning, the task of evaluation or assessment is given to the students. They take part in the process of evaluation or what is called self-assessment. Bailey (1998) defines self-assessment as "procedures by which the learners themselves evaluate their language skills and knowledge" (Bailey, 1998, p. 227).

What is clear in the self-assessment approach to evaluation is the learners' participation in the learning and evaluation process. This participation brings more motivation to the learners as they see themselves to be responsible for their own learning. It also brings autonomy which in long term makes learners life-long independent learners. They will be able to make judgment about their own learning, seek their own weakness and strength and be aware of their knowledge.

Oscarson (1997, 1989) believes in the importance of learners responsibility and argues that assessment is not the sole responsibility of the teacher but rather it is a mutual responsibility of both learners and teachers. This mutual responsibility will result in a democratic educational system and learning context. Not only the learners but also the teachers and the institute make benefits of self-assessment.

Many scholars (e.g., Alderson, 2005, Alibakhshi & Shaharaki-pour, 2014; Boud, 2000; Freeman & Lewis, 1998; Patri, 2002) believe that self-assessment helps learners in learning a language but what is disappointing is that self-assessment has not been attempted in many places yet (Abbaszadeh, 2012). In other words, to the researchers' best knowledge, it is not clear whether

self assessment has positive or negative impacts on EFL learners' speaking fluency and accuracy.

2. Literature Review

Many researches in the 1980s and 1990s were concerned with the development of self-assessment instruments and their validation (e.g., Lewkowicz & Moon, 1985; Oscarson, 1997). As a result, many approaches including pupil progress cards, learning diaries, log books, rating scales and questionnaires were developed (Boud, 1986; Dearing, 1997; Falchikov, 1997; Stefani, 1998; Taras, 2001, 2002). It helps learners to become autonomous and to be aware of their learning and reflect on their development (Boud, 2000; Freeman & Lewis, 1998). McDonald and Boud (2002, 2003) found that when learners assess their own learning, their learning will be promoted to a high extent. Higgins, Hartley, and Skelton (2001) and Ivanic, Clark, and Rimmershaw (2000) have mentioned that for the development of self-regulation in learners both teachers and learners' **feedback on the learning process are needed and that this ability to give feedback must be promoted in both.** Jewah et al. (2004) also confirm that learners must occupy an important and active role in the process of giving feedback.

Some scholars (Boud, 2000; Rowntree, 1987; Taras, 2001) have mentioned that the use of self-assessment in some area like England and Wales is very uncommon and it is probably because they see it as inconsistent with the conventional forms of assessments. In the same vein, Carton (1993) has discussed how self-assessment can become part of the learning process. He has described his use of questionnaires to encourage learners to reflect on their learning objectives and preferred modes of learning. He has also presented an approach to monitoring learning that involves the learners in devising their own criteria; an approach that he believes helps learners to become more aware of their own cognitive processes.

Accordingly, Butler and Li (2001) have investigated the effectiveness of self-assessment among young EFL learners. They

Self-assessment impact on EFL learners' fluency and accuracy

found some positive effects of self-assessment on the students' English performance as well as their confidence in learning English, though the effect sizes were small. Their study also found that teachers' and students' perceptions of the effectiveness of the self-assessment are different depending on their teaching/learning contexts. A number of interesting insights were also discovered through interviews with teachers. The teachers were asked about the best way to utilize self-assessment as a part of foreign language instruction in the contexts wherein teacher-centered teaching has been traditionally valued.

In line with the above studies, Alderson (2005) has investigated the importance of self-evaluation in the second language class today, and stressed the advantages of having students keep a regular journal. By taking the methodological framework offered by the Communicative Approach to Language Teaching as a starting point, the dynamic inter-dependence of purpose, methodology and evaluation within the curriculum were studied. In this sense, formative or ongoing evaluation becomes one of the most practical assessment techniques for controlling our students' progress as well as the effectiveness of our teaching program. Self-evaluation has a number of additional advantages regarding both the affective implication of students in assessing their own learning processes and to their participation in class management.

The fact that training and feedback have some influences on the accuracy of self-assessment has been investigated by many researchers (Orsmond & Taras, 2001; Patri, 2002; Stefani, 1998). Having enough training before doing self-assessment is believed to be effective (Hanrahan & Isaacs, 2001; Li, 2001; Taras, 2002). Some researchers have given some recommendations for doing a better self-assessment. Lejk and Wyvill (2001), for example, have recommended a holistic approach rather than a category-based approach. Other studies, such as Blatchford (1997), have concluded that there is a significant association between self-assessments and attainments in both English/reading and mathematics. Taras (2001) has reported that the active participation of learners and the teachers' **experience will enhance the process of self**-assessment.

Fallows and Balasubramanyan (2001) have reported that compulsory training combined with multiple ratings offer many benefits. Motivation also plays a significant role in the accuracy of self-assessment. AlFallay (2001) has concluded that those learners who have integrative motivation do more accurately in assessing themselves than those with instrumental motivation. The former group was also seen to be less apt to reflect overestimation than the latter one. Moreover, AlFallay (2001) has acclaimed that language proficiency also influence on the accuracy of self-assessment inducing that those with higher proficiency were more accurate than those with lower proficiency. He found that high proficiency learners to some extent underestimate their performances while the lower proficiency level often overestimates their performances.

Similar results were reported by Davidson and Henning (1985), and Heilenmann (1990). Some researchers claim that motivation intensity is an important factor in the accuracy of self-assessment (Livesey, 1992; Morton et al., 1999). A positive relationship between self-esteem and some other personality traits has also been reported (Calderon, 1991; Collins, 1993; Lindholm-Leary & Borsato, 2002). Researchers (Heilenman, 1991; Wesche et al., 1990) have mentioned that learners are able to self-assess their achievements accurately. Stankowv (1998) reported that students are often overconfident on the tests of vocabulary and general knowledge.

Brantmeier (2005a, 2005b, 2006) has reported that levels of self-assessed abilities positively correlated with levels of enjoyment. The study also produced significant effects for both self-assessed ability and enjoyment on written recall, but no such effects were reported on multiple-choice questions. These studies lend support to the hypothesis that self-assessment can be accurate for placement. Oscarson (1997) has claimed that, "it seems to be fairly commonly agreed that the question of accuracy and appropriateness of self-estimates of proficiency depends, to a great extent, on the feature of context and on the intended purpose of the assessment in each individual case.

Harutyunyan and Gasparyan (2003) have investigated the possibility of integrating students self-assessment into the

Self-assessment impact on EFL learners' fluency and accuracy

evaluation process of the Intensive English Program (IEP) for students at the American University of Armenia (AUA) to raise the students' awareness of their strengths and weaknesses in different language learning areas and to prepare the students for autonomous English language learning.

Most studies have involved older subjects such as college students (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Topping, 1998) and in-service staff (e.g., Jones & Fletcher, 2002; Saavedra & Kwun, 1993). That is, little research has focused on the effects of self and peer-assessments in primary and middle schools. Significant differences between the characteristics of adolescents and adults suggest that studies should specifically investigate whether self- and peer-assessment are suitable for younger students.

According to Matsuno (2009), many researchers have reported high correlations between students and teacher-assessment, while other studies have reported low correlation between them. The work of Pierce, Swain and Hart (1993) is based on school aged learners in a French immersion program in Canada. Learners assessed themselves against two criteria: by comparing themselves with a native speaker and by reflecting on the difficulty they experienced with everyday tasks in French. Results were compared against learners

researchers concluded that self-assessment is not a reliable indicator of proficiency. However, as they mention, many of the subjects have little or no access to the target language or native speakers outside the classroom. In effect, it would be difficult for them to imagine how they would perform. In a comparison of a test of Dutch as a second language for adult learners with a self-assessed version of the same test, Janssen-van-Dieten (1989) has found the self-assessed version less reliable although earlier studies and her pilot studies had been more encouraging. For her, the value of self-assessment is "its positive influence on the learning process" (Janssen-van-Dieten, 1989: 44). Thomson (1996), in studying learners of Japanese as a foreign language, also felt very positive about using self-assessment despite finding considerable diversity in the accuracy of self-making.

results on

Other studies have reported that self-assessment is reliable. Bachman and Palmer (1989), for example, found that members of a multilingual, multicultural group of adult learners of English as a foreign language in the US were able to reliably self-rate themselves for their communicative language abilities. Another example of success with self-assessment has been reported in Blanche's (1990) study wherein the ability of a group of adult learners of French as a foreign language to estimate their own speaking ability (incomplete). He concludes that "the overall accuracy of the self-evaluation... is impressive" (Blanche 1990: 226). Variability in sample size, age of subjects, cultural and educational backgrounds, target language, the test format, the education context and the criteria against which self-assessment is compared all affect reliability. What is comforting is that even when the results were tested against reliability, researchers maintain the value of self-assessment. One way to validate the individual self-assessments is to have teacher randomly check some of the results. This would encourage learners to be honest and realistic in their self-making and would contribute to accreditation. In addition, regular random checking would provide a clearer understanding of the reliability issue (Gardner & Miller, 1999). Xiao and Lucking (2008) examined the validity and reliability of student generated assessment scores. The findings indicated that the validity and reliability of student generated rating scores were extremely high. Alfallay (2004) investigates the role of some selected psychological and personality traits of learners of English as a foreign language for accuracy in their self- and peer-assessments. The study also shows that long periods of practice and sufficient feedback have a positive effect on the accuracy of self-assessment. He also maintains that students with low self-esteem are the most accurate in assessing their performance, whereas learners with instrumental motivation are the least accurate (Alfallay, 2004). Sung et al. (2005) show that significant consistency is found between the results of student self- and peer-assessments and the results of teacher assessment.

Blaska and Krekeler (2008) investigated the reliability of self-assessments of pronunciation skills and attempted to understand the

Self-assessment impact on EFL learners' fluency and accuracy

causes of difficulties. In this study, 46 advanced learners of German assessed their own articulation of different speech sounds in comparison with the sounds produced by a native speaker. In 85% of all cases the assessments of the raters and the self-assessments were identical. However, the learners only identified half of the number of speech sounds which the raters believed to be inaccurate. The study therefore concluded that even experienced L2 learners seem to find it difficult to self-assess correctly their pronunciation skills.

Oscarson (1997) sums up progress in the area of self-assessment by reminding us that research in self-assessments is fairly new. He concludes that there are still many problems remained. For instance, learner goals and interpretations need to be adjusted with external necessities. Also self-assessment is not self-explanatory; it must be introduced slowly and learners need to be guided and supported in their use of the instruments.

3. Purpose of the Study

The main objective of the present study was to find out whether self-assessment has significant impacts on Iranian EFL learners' speaking fluency and accuracy. The next objective was to investigate whether self-assessment has the same impact on intermediate and upper-intermediate language learners' speaking accuracy and fluency. More specifically the following research questions were raised:

1. Does self-assessment significantly improve EFL learners' speaking fluency and accuracy?
2. Does self-assessment have the same impact on intermediate and upper-intermediate EFL learners?

4. Method

4.1 Participants

The present study was conducted in an English language institute called Zabansara. The books taught to the EFL learners in this institute were New Interchange II and III. These books contained language skills such as reading, writing, speaking as well as language components such as grammar and vocabulary. The classes were held three times a week and each session took two hours. Learners were from different age range and they had different goals for language learning.

The participants of this study were 30 pre-intermediate and 30 intermediate EFL learners. The participants' level of proficiency was determined on the basis of the result of a pretest. The researcher used the Solutions placement test designed by Edwards (2007) at Oxford University. The test contains: 50 multiple choice questions which assesses students' knowledge of grammar and vocabulary from elementary to intermediate level. The participants' with scores of 26 to 36 are labeled pre-intermediate and the students with scores above 39 are labeled intermediate.

The participants at each level of proficiency were selected through convenience sampling in different stages. At first from language learners attending language institute in Zahedan 30 students who were labeled as pre-intermediate were selected and divided into two groups. Then 30 intermediate students were selected in the same way. The experimental groups of each level received treatment on self-assessment and they were evaluated through self-assessment technique. However, the participants of the control groups were evaluated by teachers.

4.2 Design of the Study

The design of the present study was experimental. The dependent variables were speaking fluency and accuracy. The independent variable was self-assessment and moderate variable was learners' proficiency level. The control group received a pretest, placebo, and posttest. However, the experimental group received pretest, treatment, and posttest.

4.3 Instruments

Different instruments were used in the present study: a placement test, speaking pretest, speaking achievement test, and semantic differential scale. Each is explained in details in the following sections.

Placement test: This placement test was intended to help teachers decide which level of *Solutions*, Elementary, Pre-Intermediate or Intermediate is the most suitable for their students. It should be given at the beginning of the school year or semesters. *Solutions* placement test has been developed by Edvards (2007) at Oxford University after consultation with teachers and it is designed to assess students' knowledge of language as well as their receptive and productive skills. The test contains: 50 multiple choice questions which assess students' knowledge of key grammar and vocabulary from elementary to intermediate levels, and a reading text with 10 graded comprehension questions. The grammar items in this placement test were scored separately as grammar pretest. The reliability index of the test which was estimated through KR-21 approach was .82 which was an acceptable reliability index. This test had also functioned as the pretest to see whether the participants were homogenous or not. The scores of this test were also used for estimating the criterion related validity of the teacher made test.

Speaking Pretest: This test was taken at the onset of the treatment to see whether the participants of each level of proficiency were homogenous in terms of speaking fluency and accuracy or not. It consisted of some productive tasks such as storytelling, a picture description task, some essay-type questions about the speakers' personal information and general knowledge about the internet, computer, football, etc. Each participant's responses to the questions were recorded and evaluated in terms of fluency and accuracy using the semantic differential scale. The reliability of this test was estimated through inter-rater approach. The reliability index was .78 which was acceptable.

Speaking posttest: Posttest is a test taken after the implementation of the treatment to determine the effect of treatment. The speaking posttest used in the present study consisted

of the same productive tasks as used in the pretest. Each participant's responses to the questions were recorded and evaluated in terms of fluency and accuracy using the semantic differential scale. The reliability of this test was estimated through inter-rater approach. The reliability index was .80 which was acceptable.

Semantic differential scale: This instrument measured speaking fluency and accuracy of the participants. The semantic differential scale is used when two opposite adjectives are used to measure one variable. In this study, speaking fluency of the test takers was evaluated using the following semantic differential scale, None-Fluent 1,2,3,4,5,6,7,8,9,10 Fluent. The raters were required to evaluate the test takers 'performance using one of the numbers (1-2-3-4-5-6-7-8-9). The accuracy of the speakers' grammar, pronunciation, stress, and intonation were also evaluated using the following semantic differential scale: none-accurate 1,2,3,4,5,6,7,8,9,10 accurate.

4.4 Data Collection Procedure

After the participants were selected, the researcher labelled them as either control or experimental groups. For speaking courses taken by both intermediate and upper intermediate participants, the teachers used interchange books I & II for speaking topics. For speaking courses, the teachers used different tasks. The main focus was on teaching speaking skills to language learners. The only difference between the control and experimental groups was the way they were assessed. Experimental groups received treatment on self-assessment and they were given instructions to self-assess their speaking using self assessment report sheets. During the treatment period, different techniques of self-assessment were introduced by the teachers. The teachers defined each technique in details and asked learners to practice the technique for the next coming section. Whenever necessary, the teacher provided the learners with enough information regarding self-assessment. At the beginning of the training the teacher gave supports to each step taken by the learners and as the learners became more and more proficient in using the self-assessment techniques, the teachers' supports became less to

make the learners more autonomous and independent. After a 15-week treatment all groups received the speaking posttest.

4.5 Data analysis

Due to the design of the study, (two pretests, two dependent variables, two variables and two groups), the best statistical test was the analysis of covariance because it permitted researchers to statistically control for differences on the pretest so that posttest differences would not be due to initial differences before training.

5. Results

As there was a two-group pretest/posttest design for each level, the best statistical procedure for analyzing the data was multivariate ANCOVA. The scores on each pretest were treated as a covariate to 'control' for pre-existing differences between the groups. As ANCOVA has some assumptions and the researchers should make sure that the assumptions are not violated, at first we checked for the assumptions such as reliability, liberality, etc. In this study, we tried to check for at least two assumptions of ACNOVA. The results for each level of proficiency are presented in the following parts.

5.1 Reliability Assumption for Pretest

In order to calculate the reliability of both pretest and posttest, we tried to estimate KR-21. The results showed that reliability indices of speaking accuracy and fluency of intermediate language learners which were estimated through inter-rater reliability technique were .75 and .8, respectively. Moreover, reliability indices of speaking accuracy and fluency of upper-intermediate language learners were 0.79 and 0.75. Therefore, the reliability assumption of the tests was not violated.

5.2 Assumptions for Homogeneity of Regression Slopes

Alibakhshi and Sarani

The results of assumptions for homogeneity of regression slopes for intermediate and language learners are shown in Tables 1 and 2.

As shown in Table 1, the interaction between intermediate groups and fluency pretest is not significant ($p=0.78>0.05$). Moreover, the interaction between the groups and speaking accuracy is not significant ($p=0.78>0.05$). Therefore, the assumption for homogeneity of regression slopes was not violated.

Table 1: Assumption for homogeneity of regression slopes of speaking fluency and accuracy pretest and posttest (intermediate groups)

	Effect	Value	F	Sig.
Groups * speaking fluency pretest	Pillai's Trace	.32	.314 ^a	.73
	Wilks' Lambda	.97	.314 ^a	.73
	Hotelling's Trace	.07	.314 ^a	.73
	Roy's Largest	.09	.314 ^a	.73
	Root			
Groups * speaking accuracy	Pillai's Trace	.6	.253 ^a	.78
	Wilks' Lambda	.99	.253 ^a	.78
	Hotelling's Trace	.7	.253 ^a	.78
	Roy's Largest	.05	.253 ^a	.78
	Root			

As shown in Table 2, the interaction between upper-intermediate groups and fluency in pretest is not significant ($p=0.73>0.05$). Moreover, the interaction between the groups and speaking accuracy is not significant ($p=0.77>0.05$). Therefore, the assumption for homogeneity of regression slopes was not violated.

Table 2: Assumption for homogeneity of regression slopes of speaking fluency and accuracy pretest and posttest (upper intermediate groups)

	Effect	Value	F	Sig.
Groups * speaking fluency pretest	Pillai's Trace	.012	.412 ^a	.73
	Wilks' Lambda	.988	.412 ^a	.73

Self-assessment impact on EFL learners' fluency and accuracy

	Hotelling's Trace	.012	.412 ^a	.73
	Roy's Largest Root	.012	.412 ^a	.73
Groups * speaking accuracy	Pillai's Trace	.009	.273 ^a	.77
	Wilks' Lambda	.991	.273 ^a	.77
	Hotelling's Trace	.010	.273 ^a	.77
	Roy's Largest Root	.010	.273 ^a	.77

5.3 Descriptive Statistics

Descriptive statistics of the study including mean and standard deviation (SD) of vocabulary and grammar posttests are presented in the following table.

As shown in Table 3, the means of control and experimental group on the fluency posttest were 4.2(SD= 0.56) and 6.6 (SD= 0.48), respectively. The results also indicate that the means of control and experimental group on the accuracy posttest were 3.6 (SD= 0.48) and 6.3 (SD= 0.89), respectively. As the descriptive statistics cannot tell the difference between the mean score, we had to run an inferential statistics procedure (ACNOVA). The results are shown in the following table.

Table 3: Intermediate groups' descriptive statistics

	Groups	Mean	SD
Fluency posttest	Control (intermediate)	4.20	0.56
	Experimental (intermediate)	6.1	0.48
	Control (upper intermediate)	5.3	1.1
	Experimental (upper intermediate)	7	0.55
Accuracy posttest	Control (intermediate)	3.6	0.48
	Experimental (intermediate)	5.3	0.89
	Control (upper-intermediate)	4.5	1.23
	Experimental (upper-intermediate)	6.7	0.7

5.4 The Results of ACNOVA for Posttest (Upper-intermediate)

Levene's Test of Equality of Error Variances table tells if the researcher violated the assumption of equality of variance. The sig. value *greater* than .05 is needed. If this value is smaller than .05 (and therefore significant), this means that variances are not equal and that we have violated the assumption. The results of Leven's test are shown in the following table.

Table 4: Levene's test of equality of error variances

	F	df1	df2	Sig.
Fluency (intermediate groups)	.233	1	28	0.63
Accuracy (intermediate groups)	0.2	1	28	0.82
Fluency (upper-intermediate groups)	.233	1	28	0.70
Accuracy (upper-intermediate groups)	0.3	1	28	0.90

The results in Table 4 indicate that we have not violated the assumption ($p > .05$).

Table 5: The results of ACNOVA for upper-intermediate groups

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta ²
Corrected Model	Fluency posttest	69	3	23.249	92.250	.001	.914
	Accuracy posttest	43	3	14.659	22.658	.001	.723
Intercept	Fluency posttest	41	1	41.855	166.073	.001	.865
	Accuracy posttest	25	1	25.949	40.107	.001	.607
Accuracy pretest	Fluency posttest	1.8	1	1.813	7.192	.12	.027

Self-assessment impact on EFL learners' fluency and accuracy

	Accuracy posttest	.76	1	.761	1.176	.288	.043
Fluency pretest	Fluency posttest	.44	1	.442	1.755	.197	.063
	Accuracy posttest	.37	1	.373	.577	.454	.022
groups	Fluency posttest	58	1	58.423	231.814	.001	.899
	Accuracy posttest	36	1	36.952	57.113	.001	.687

a. R Squared = .914 (Adjusted R Squared = 904)

The main ANCOVA results presented in Table 5 show that there is a significant difference between the groups' means on fluency and accuracy tests ($p=0.001 < 0.05$). Therefore, it could be argued that self-assessment significantly improved the upper-intermediate language learners' speaking fluency. The results also indicate that the partial Eta squared of speaking fluency was shown to be .89 but that of speaking accuracy test was .68 which is smaller than the effect of speaking test. Therefore, it could be strongly argued that self-assessment impact on speaking fluency is greater than its impact on speaking accuracy.

Table 6: The results of ACNOVA for intermediate groups

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta ²
Corrected Model	Fluency posttest	69	3	24.25	94	.001	.814
	Accuracy posttest	43	3	15.65	24	.001	.8
Intercept	Fluency posttest	41	1	42.85	180	.001	.75
	Accuracy posttest	25	1	26.94	42.1	.001	.56
Accuracy pretest	Fluency posttest	1.8	1	2.813	7.192	.16	.027

Alibakhshi and Sarani

	Accuracy posttest	.76	1	1.76	1.176	.38	.043
Fluency pretest	Fluency posttest	.44	1	.47	2.755	.197	.08
	Accuracy posttest	.37	1	.42	.73	.454	.04
groups	Fluency posttest	58	1	60.41	271.8	.001	.70
	Accuracy posttest	36	1	37	68.1	.001	.63

a. R Squared = .914 (Adjusted R Squared = .904)

The main ANCOVA results presented in Table 6 show that there is a significant difference between the intermediate groups' means on fluency and accuracy tests ($p = 0.001 > 0.05$). Therefore, it could be argued that self-assessment significantly improved the upper-intermediate language learners' speaking fluency and accuracy. However, the results showed that the partial Eta squared of speaking fluency was shown to be .70 but that of writing test was .63 which is smaller than the effect of speaking accuracy test. Therefore, it could be strongly argued that self-assessment impact on speaking fluency is greater than its impact on speaking accuracy.

6. Discussion

The present study tested the hypotheses that self assessment does not significantly improve Iranian upper-intermediate and intermediate EFL learners' speaking fluency and accuracy. In doing so, participants in experimental groups in both upper-intermediate and intermediate levels received a 15-session treatment in which they got familiar with self-assessment and its techniques. In addition, they learned how to apply self-assessment in the process of their language learning particularly in speaking skill.

The data of the study for both upper-intermediate and intermediate were analyzed through ACNOVA. From the two-way ACNOVA, several interesting findings were revealed. First, the

Self-assessment impact on EFL learners' fluency and accuracy

results of the study showed that there was a significant difference between the mean scores of the upper-intermediate and intermediate on both speaking fluency and accuracy posttests. The mean of the experimental groups were significantly higher than the mean scores of the control groups. Therefore, it could be strongly argued that self-assessment had significant impact on upper-intermediate and intermediate learners' speaking performance. The findings are therefore consistent with the previous findings (Alfallay, 2004; Dlaska & Krekeler, 2008; Hanrahan & Isaacs, 2001; Li, 2001; Orsmond et al., 2000; Patri 2002; Smith et al., 2002; Stefani, 1998; Sung et al., 2005; Taras, 2001, 2002).

The results are also consistent with the findings of Abbasszadeh (2012) who found that self-assessment significantly improve listening and reading performance. However, they believed that self-assessment has the same impact on both intermediate and beginner language learners.

The results also revealed that the Partial Eta Squared of the speaking fluency for intermediate participants was .70; whereas, the Partial Eta squared of the writing test was .63. This value also indicates how much of the variance in the dependent variable is explained by the independent variable. If partial eta squared value is converted into a percentage by multiplying by 100, then the number 70 is obtained. Therefore, we are able 70 percent of the speaking posttest is explained by self assessment, while self assessment can explain only 63percent of the variance of writing posttest. Therefore, it could be strongly argued that the impact of self assessment on speaking skill is higher than its impact on writing skills.

The results also showed that self assessment had significant impact on improving upper-intermediate language learners speaking accuracy and fluency. Partial Eta Squared of speaking and writing tests were .89 and .68, respectively. Therefore, it could also be argued that the impact of self-assessment on speaking fluency of upper-intermediate students is also higher than its effect on speaking accuracy because self-assessment could explain 89 % of the variance of speaking posttest; whereas, it could explain only 68 % of the variance of speaking accuracy.

Such a difference between the impact of self-assessment on speaking fluency and accuracy might be rooted in the difficulties of acquiring speaking accuracy like native speakers in non-English speaking contexts in Iran which needs further investigation.

Another interesting finding of the present study which was not reported by related studies was the difference between Partial Eta Squared of posttests taken by intermediate students and the tests taken by upper-intermediate language learners. Such a controversy is either due to difference between speaking fluency and accuracy (Nunan, 2003; Harmer, 2009) or due to the affective factors and psychological states of beginner and intermediate language learners as well as the rate of their dependence on the teachers. As stated by Richards and Rodgers (2001), beginner and intermediate language learners with low proficiency are more dependent on the teachers than advanced language learners. That is why, less proficient language learners in comparison with or advanced language learners cannot benefit a lot from self-assessment technique.

Therefore, it could be strongly argued that despite the importance of self-assessment on any educational system, teachers must somehow support beginner language learners and direct and monitor the self-assessments done by the language learners. The difference between the Eta of self assessment on speaking is deeply rooted in many other factors. The findings of this study are consistent with McDonald and Boud (2003) who found that when learners assess their own learning, their learning will be promoted to a high extent.

The results of the current study are also supported by Butler and Li (2005) who investigated the effectiveness of self-assessment among young EFL learners and found some positive effects of self-assessment on the students' English performance as well as their confidence in learning English. Moreover, in line with a few recent studies (e.g., Black & William, 1998; Pellegrino & Chudowsky, & Aglaser, 2001) it could be argued that formative self assessment has a significant positive effect on students' learning.

7. Conclusion

Self-assessment impact on EFL learners' fluency and accuracy

In line with the results of the present study, it could be concluded that self-assessment significantly improves language learners' receptive skills. It could be also inferred that self-assessment has a significant effect on intermediate and upper-intermediate language learners' speaking fluency and accuracy. In addition, it could be concluded that intermediate language learners need to receive more corrective feedback from their teachers than upper-intermediate level language learners do. Finally, it is concluded that self-assessment effect on speaking fluency is greater than its effect on speaking accuracy. Therefore, language learners need to have support and corrective feedback from their speaking teachers. Such a difference between self-assessment effect size on speaking accuracy and fluency might be rooted in different variables which need to be explored by the other researchers.

References

- Abbasszadeh, S. (2012). *The impact of self- assessment on Iranian EFL learners' writing and speaking*. Unpublished master thesis. Yasouj University, Yasouj, Iran.
- Alderson, C.A. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. New York: NY.
- ALFalla, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32(3), 407-425.
- Alibakhshi, G., & Shahrakipour, H. (2014). The effect of self-assessment on EFL learners' receptive skills. *Jurnal Pendidikan Malaysia* 39(1), 9-17.
- Bachman, L. F. & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

- Bailey, K. (1998). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-79.
- Barbot, M. J. (1991). New approaches to evaluation in self-access learning (trans. form. French). *Études de Linguistique Appliquée*, 79, 77,-94.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Higher Education*, 5 (1), 7-74.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing*, 7(2), 202-29.
- Blatchford, P. (1997). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison. *Educational Psychology*, 17, 354-360.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22 (2), 151-167.
- Brantmeier, C. (2005a). Nonlinguistic variables in advanced L2 reading: Learner's self-assessment and enjoyment. *Foreign Language Annals*, 38 (4), 493-503.
- Brantmeier, C. (2005b). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension. *The Modern Language Journal*, 89 (1), 37-53.
- Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, computer-based testing, and subsequent performance. *System*, 34 (1), 15-35.
- Brown, D. H. (2004). *Language Assessment: Principles and Classroom Practices*. London: Pearson Education, Inc.
- Brown, G., Bull, J. & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Butler. Y. & Li, J. (2003). *The effect of self-assessment among young learners of English*. University of Pennsylvania.
- Calderon, M. (1991). Promoting language proficiency and academic achievement through cooperation. ERIC Document, ERIC #: ED436983.
- Carton, F. (1993). Self-evaluation at the heart of learning. *Le Fran çais dans le Monde* (special number), 28-35.
- Collins, A. (1993). A study of the provision of modern language pupils with special educational needs. Unpublished M.ED. dissertation, The Queen's University of Belfast.
- Davidson, F. & Henning, G. (1985). A self-rating scale of English difficulty. *Language Testing*, 2, 164-169.

Self-assessment impact on EFL learners' fluency and accuracy

- Dearing, R. (1997). *Higher Education in the learning society*. London: HMSO.
- Dickinson, L. (1987). *Self-instruction in language learning*. London: Cambridge University Press.
- Dlaska, A., & Krekeler, C. (2008). *Self-assessment of pronunciation*. Surrey university press.
- Falchikov, N. (1997). Why do lecturers involve students in assessment? In Paper delivered at the 2nd North Umbria Assessment Conference, Encouraging Partnership in Assessing Learning, 3–5 September, University of Northumbria, Newcastle.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59 (4), 395-430.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287-322.
- Fallows, S. & Balasubramanyan, C. (2001). Multiple approaches to assessment: reflections on the use of tutor, peer and self-assessment. *Teaching in Higher Education*, 6, 229-246.
- Freeman, R. & Lewis, R. (1998). *Planning and implementing assessment*. London: Kogan Page.
- Gardner, R. (1985). *Social psychology and second language learning*. London: Arnold.
- Graham, S. (2004). Giving up on modern foreign languages? Students' perceptions of learning French. *The Modern Language Journal*, 88 (2), 171-191.
- Hanrahan, S. & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research and Development*, 20, 53-70.
- Harmer, J. (2009). *The practice of English language teaching*. London: Pearson Longman.
- Heilenman, K. (1991). Self-assessment and placement: a review of the issues. In: R.V. Teschner (Eds.), *Assessing Foreign Language Proficiency of Undergraduates, AAUSC Issues in Language Program Direction* (pp. 93-114). Boston: Heinle & Heinle.
- Higgins, R., Hartley, P. & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education*, 6 (2), 269-274.
- Ivanic, R., Clark, R., & Rimmershaw, R. (2000). What am I supposed to make of this? The message conveyed to students by tutors' written

- comments. In: Lea, M.R., Stierer, B. (Eds.), *Student Writing in Higher Education: New Contexts*. Buckingham: Open University Press.
- Janssen-van Dieten, A. (1989). The development of a test of Dutch as a foreign language: the validity of self-assessment by inexperienced subjects. *Language Testing*, 6 (1), 30-46.
- Jawah, C., Macfarlane-Dick, D., Matthew, R., Nicol, D., Ross, D. & Smith, B. (2004). Enhancing student learning through effective formative feedback. In: The Higher Education Academy. LTSN, London.
- Jones, L. & Fletcher, C. 2002. Self-assessment in a selective situation: An evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology*, 75, 145-161.
- Lejk, M. & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education*, 26, 62-72.
- Lewkowicz, J. A. & Moon, J. (1985). Evaluation: A way of involving the learner. In J.C. Alderson (Eds.), *Lancaster Practical Paper in English Language Education* (Vol. 6: Evaluation), (pp. 45-80). Oxford: Pergamon Press.
- Li, L. (2005). Some refinement on peer assessment of group projects. *Assessment and Evaluation in Higher Education*, 26, 5-18.
- Lindholm-Leary, K. & Borsato, G. (2002). Impact of two-way immersion on students' attitudes toward school and college. Eric Digest. ERIC Document # ED464541.
- Livesey, D. (1992). An application of the theory of reasoned action for relating attitude, social support and behavioral intention in an EFL setting. Paper presented at the Annual meeting of the Teachers of English to Speakers of Other Languages 26th, Vancouver, BC, Canada, 3-7 March, 1992.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26 (1), 75-100.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10, 209-220.
- Morton, L., Lemieux, C., Diffey, N. & Awender, M. (1999). Determinants of withdrawal from the bilingual career track when entering high school. *Guidance and Counseling* 14, 1-14.

Self-assessment impact on EFL learners' fluency and accuracy

- Mousavi, S. A. (2012). *An encyclopedic dictionary of language testing*. Tehran: Rahnama Press.
- Nunan, D. (2003). *Practical English language teaching*. Singapore: McGraw Hill.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 25, 23-38.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6 (1), 1-13.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *Language testing and assessment*, 7, (pp.175-87). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Oscarson, S. M. (1989). Self-assessment of language proficiency: Rationale 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, 109-131.
- Peirce, B. M., Swain, M., & Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics*, 14, 25-42.
- Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academic Press.
- Richards, J.C., & Rodgers, T.S. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78, 450-462.
- Shrauger, J.S. & Osberg, T.M. (1981). The relative accuracy of self-predictions and judgments by others of psychological assessment. *Psychological Bulletin*, 90, 322-351.
- Stankowv, L. (1998). Calibration curves, scatter plots, and the distinction between general knowledge and perceptual tests. *Learning and Individual Differences*, 8, 28-51.
- Stefani, L. A. J. (1994). Peer, self, and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19, 69-75.
- Stefani, L.J. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23 (4), 339-350.
- 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 and Education: An International Journal*, 45, 187-202.

- Sung, Y.T., Chang, K.E., Yu, W.C., & Chang, T.H. (2009). Enhancing teachers' learning and reflection through structured digital portfolios. *Journal of Computer Assisted Learning*, 2(1), 21-35.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Toward transparency for students and for tutors. *Assessment and Evaluation in Higher Education*, 26, 605-614.
- Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment and Evaluation in Higher Education*, 27, 501-510.
- Teweles, B. (1995). Motivation as a two-sided coin: motivational differences between college-level Chinese and Japanese learners of EFL. *Texas Papers in Foreign Language Education*, 2, 1-22.
- Thomson, C.K. (1996). Self-assessment in self-directed learning: issues of learner diversity. In Pemberton, R.; Li, E.; Or, W. Pierson, H. (Eds.). *Taking control: Autonomy in language learning*. Hong Kong: Hong Kong University Press.
- Todd, R. W. (2002). Using Self-Assessment for Evaluation. *English Teaching Forum*, 40(1), 16-19.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.
- Upshur, J. (1975). Objective evaluation of oral proficiency in the ESOL classroom. In: L. Palmer, & B. K. Spolsky (Eds.), *Paper on Language Testing* (pp.1967-1974). TESOL, Washington, DC.
- Von Elek, T. (1987). A test of Swedish as a second language: An experiment in self-assessment. In: Li, Y., Fok, A., Lord, R., Low, G. (Eds.), *New Directions in Language Testing* (pp. 47-57). Oxford: Oxford University Press.
- Wesche, M., Morrison, F., Ready, D., & Pawley, C. (1990). French immersion: Postsecondary consequence for individuals and universities. *Modern Canadian Language Review*, 46, 430-451.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(4), 186-193.