On the Relationship Between Risk-taking and Self-assessment of Speaking Ability: a Case of Freshman EFL Learners

Mansoor Tavakoli
University of Isfahan, Iran

Behrooz Ghoorchaei
University of Isfahan, Iran

This study investigates the relationship between self-assessment and teacher's rating of speaking ability. Risk-taking (a personality factor) is also singled out to examine its relationship with self-assessment. 79 Iranian EFL Freshmen at the University of Isfahan were given a picture description task to elicit their speech samples. After the speech samples were recorded, they were asked to rewind the tapes and assess their own speaking ability based on a modified version of Farhady et al.'s (1994) oral proficiency scale. Later, the speech samples were rated by their teacher based on the same scale. Moreover, the influence of risk-taking on students' self-assessment was investigated. To analyze the obtained data, Pearson Correlation, Independent samples t-test and the Kruskal-Wallis test were used as statistical procedures. The results revealed that the relationship between self-assessment and teacher's rating of students' speaking ability in the picture description task was not statistically significant. Moreover, there was not any statistically significant relationship between risk-taking and students' self-assessment, however, high risk-takers tended to assess their speaking ability higher than medium and low risk-takers. The findings have some implications for language pedagogy and language testing.

Key words: assessment, risk-taking, self-assessment, teacher's rating, speaking ability

INTRODUCTION

The administration of assessment of oral proficiency is a time-consuming and laborious process which needs trained interviewers and raters. Moreover, the use of conventional tests with marks early on can have a detrimental impact on students (Yorke, 2001). The present study aims at investigating self-assessment as an alternative assessment procedure in determining oral proficiency of EFL learners.

The notion of self-assessment is no longer a new concept in language ability assessment (Brindley, 1989) and its use as a measure of language learning is well researched (e.g., Krausert, 1991; Oscarson, 1978; Wang & Wang, 2007). Nevertheless, there is room for research as to how learners can be involved in oral proficiency assessment because only few studies have attempted to explore this area at the time of writing (Cheng & Warren, 2005). Moreover, in her review of Ekbatani and Pierson's book entitled "Learner-Directed Assessment in ESL", Coombe (2001) notes that "While learner-centered teaching has received a lot of attention in the literature, the importance of involving learners in the assessment process has received scant empirical attention"(p. 1).

Being assessed is an emotional business and being judged about the quality of one's work is potentially a humiliating experience (Clegg & Bryan, 2006). Performance on language tests is affected by attributes of individuals such as age, gender, cultural background, cognitive abilities and affective schemata (Bachman & Palmer, 1996). However, the link between emotional factors such as risk-taking and self-assessment has received scant empirical attention. Thus, the present study aims at investigating on the one hand the role of risk-taking (as a personality factor) in students' self-assessment of speaking ability and on the other hand the relationship between teacher's rating and students' self-assessment of speaking ability.

BACKGROUND

The Assessment of Speaking

In assessment of speaking, the construct refers to the particular kind of speaking that is assessed in the test. In assessing speaking we must analyze the kind of speaking that is needed to assess in a particular context in terms of social and situational needs. Moreover, we should remember in designing rating criteria and procedures that speaking is interactive (Luoma, 2004).

Testing oral proficiency has been one of the most important issues in language testing since the role of speaking ability increased due to the advent of communicative language teaching (Nakamura, 1993). However, testing speaking is difficult and can not be assessed as precisely and easily as other language skills (Harris, 1969). It takes considerable time, effort and training (Hughes, 2003). Farhady, Jafarpour, and Birjandi (1994) however, assert that despite the difficulty of evaluating speaking tests, they should be designed, administered and evaluated regardless of how much cost or effort might be needed. Specifically when a test is high stakes, the allocation of such time and effort might be considered necessary (Hughes, 2003).

There are two types of speaking assessment: Stand-alone and integrated assessment. The first one concentrates solely on spoken interaction or production while in the latter one listening, reading and writing activities are integrated in order to make the language use in the test more authentic (Luoma, 2004).

Self-assessment

The Concept of Self-assessment

Recently there has been movement in language testing away from traditional assessment to alternative assessments that are more authentic in their elicitation of meaningful communication (Brown, 2001, 2004). Self-assessment

as one of the alternative assessment options is defined by McNamara (2000) as "the process by which learners are trained to evaluate their own performances" (p. 136). It helps learners to monitor their own progress and make accurate judgments on the level of their language proficiency (Lim, 2007). It derives its theoretical justification from the principle of autonomy in second language acquisition (Brown, 2004).

According to Brown and Hudson (1998), self-assessment is of three types: performance self-assessment, comprehension self-assessment and observation self-assessment. In performance self-assessment students read a situation and decide how well they would respond in it. In comprehension self-assessment students read a situation and decide how well they would comprehend it. In contrast, in observation self-assessment students listen to audio- or video tape recordings of their own language performance and decide how well they think they performed.

In his review of self-evaluation studies, Heidt (1979; cited in de Bot, 1992) states that self-evaluation is comprised of both cognitive and moral components in that learners on the one hand make a judgment of their own knowledge and skills and on the other hand exercise a degree of willingness to be honest, independent of the outcome.

Self-assessment in ESL/EFL Context

The emphasis on encouraging learners to judge their own ability has led to the interest in the use of self-assessment techniques in ESL classrooms (Dickinson, 1987; Oscarson, 1978; von Elek, 1985). The use of self-assessment in language learning was initiated by Oscarson (1978), an assessment which offers numerical scales and checklists, many with examples connected to each question. As far as the purpose of self-assessment in ESL is concerned, researchers like Bachman (2000) and Oscarson (1989) have identified two types of self-assessment namely performance-oriented and development-oriented self-assessment.

Performance-oriented assessment measures the outcomes related to

selection, certification, placement, achievement, diagnosis, etc. In performance-oriented self-assessment, students are required to evaluate their own language ability. A number of performance-oriented self-assessment studies have shown that learners may have an accurate judgment of their own abilities. Malabonga, Kenyon, and Carpenter (2005) investigated how examinees used self-assessment to choose an appropriate starting level on the COPI (computerized oral proficiency instrument). Results showed that the majority of the examinees were able to use self-assessment to choose the tasks at appropriate levels of difficulty. They reported that only small proportion of the subjects (8%) appeared to choose tasks which were too difficult for them. Also, using scaled descriptions of performance Oscarson (1978) found that adult EFL learners were able to make fairly accurate appraisals of linguistic abilities. Ross (1998) found the largest number of correlations between L2 (i.e., second language) reading criterion variables and self-assessment in reading. Similarly, in Bailey's 1998 study, learners showed moderately high correlations between self-rated oral production ability and scores on the OPI (i.e., Oral Proficiency Inventory). Krausert (1991) made an experiment to see the usefulness of self-assessment for placement decisions with university level ESL students. She found that the subjects do correctly assess their reading, writing and speaking abilities. She recommends that self-assessment be utilized instead of standardized exams for placement in university programs. In contrast, in an attempt to investigate the reliability of self-assessment for reading placement of L2 learners, Brantmeier (2006) found that self-assessment as measured before and after reading via a five-point scale is not an accurate variable predictor for placement or subsequent performance.

Development-oriented self-assessment evaluates the participants for an extended period of time in order to check the patterns of development over time. Studies focus on how the use of self-assessment in ESL classes enhances students' language learning. Using self-report questionnaires and interviews, Wang and Wang (2007) found that self-assessment is welcomed by the majority of their subjects and that it had a positive role in the whole

ESL writing process. They found that self-assessment is more helpful in self-editing and revising than in the composing process of writing. The researchers claim that self-assessment should be introduced into the process of teaching and learning earlier because postgraduate students have already formed a fixed learning habit which is hard to break by just a few questionnaires or interviews. In a longitudinal study done by Ross (2005) it was found that formative assessment practices yield substantive skill-specific effects on language proficiency growth. In a study of self- assessment of oral proficiency by upper intermediate learners of English, Lim (2007) found that the self-assessment activity led learners to focus on specific criteria when learning, which they reported enabled better performances. Nedzinskaite, Svencioniene, and Zavistanaviciene (2006) introduced self-assessment to university students. In this program students reflected and evaluated their own progress during the term. The researchers found that "self-assessment promotes students' autonomy and independent learning skills, makes students more active in judging their own progress and encourages them to see the value of what they have learned" (p. 86). They argued that self-assessment can be integrated in daily classroom activities and that it can give a comprehensive picture of students' abilities and achievements in language learning.

Advantages of Using Self-assessment

According to Brown and Hudson (1998) the advantages of self-assessment are the speed, direct involvement of students, the encouragement of autonomy, and increased motivation because of self-involvement in the process of learning. Porter and Van Dommelon (2005) echo this advantage by stating that "students have a direct hand in determining what they want to get out of the course and thus become more directly involved and responsible for their own learning" (p. 3). Also, self-assessment as a useful skill for students' life long learning is stressed by researchers (e.g., Boud, 2000; Puhl, 1997). Harmer (2001) states that when students are engaged in self-

assessment their understanding of the feed back that their teacher gives them as well as their own awareness of learning process increases. According to LeBlanc and Painchaud (1985), performance-oriented self-assessment eliminates concerns with cheating and security issues. Also, Strong-Clause (2000) mentions that it is time and cost efficient. In a study of self- and peer-assessment of oral proficiency with Brazilian and Korean learners of English, Lim (2007) found that both self- and peer-assessment make learners aware of their own weaknesses and motivate them to confront those weaknesses or achieve high marks according to the criteria given to them. There is empirical evidence which suggests that when learners are experienced in self-assessment and use procedures that focus closely on curriculum content, self-assessment can yield accurate results (Ross, 1998).

Shortcomings of Using Self-assessment

Despite the advantages of the use of self-assessment, it has a number of disadvantages. It is believed that personal feelings are involved in learners' assessments (Dickinson, 1987) and that the criteria used for assessing one's own proficiency in a communicative domain is subjective (Little, 2005). This problem however can be alleviated if the scoring grids the students use to assess themselves describe "clear and concrete linguistic situations in which they are to consider their performance in terms of precisely described behaviors"(Brown & Hudson, 1998, p. 666) and by providing learners with training and guidance in the marking criteria (Freeman, 1995). Some studies show that it is difficult for some learners to cope with self-assessment techniques (Lim, 2007); therefore, learner training should be utilized in order to have an effective use of the instrument (Lim, 2007; Rolfe, 1990). A shortcoming of self-assessment which is difficult to overcome is its subjectivity in situations whereby the consequences of self-assessment become an integral part of the assessment itself. For example, self-assessment might be effective when it is used for research purposes but it may not function well when the aim of self-assessment is placement of the students in

to different study levels (Brown & Hudson, 1998).

Contrary to alternative assessment, traditional forms of assessment can not provide teachers with information about "what the students can do in their second language" (Richards & Renandya, 2002, p. 336). Moreover, alternative assessment enjoys valid and reliable procedures that avoid problems of norming, linguistic and cultural biases which are inherent in traditional tests. It also gives us the power to reflect on, discuss and assist a student's journey through the learning process (Huerta-Macias, 2002). Thus, despite all the shortcomings of self-assessment, it can be used not only in self-directed learning but also in the conventional school and university classroom "if implemented systematically and integrated into everyday classroom activities" (Harris, 1997, p. 19).

Risk-taking and Self-assessment

Risk-taking is defined in the dictionary of language teaching and applied linguistics as "a personality factor which concerns the degree to which a person is willing to undertake actions that involve a significant degree of risk" (Richards, Platt & Platt 1992, p. 317). Also, Brown (2001) mentions that self-confidence and language ego lay the ground work for risk-taking. As mentioned in Wen and Clement (2003) good language learners are supposed to be risk-takers. They are willing to guess, willing to appear foolish and willing to try out new structures about which they are unsure. Risk taking is shown to be related to success in language learning.

Ely (1986) for instance, assumed that high risk-taking will yield positive results in second language learning. Also, Evensen and Bednar's study (as cited in Jonassen & Grabowsky, 1993) showed that high risk-takers reported greater perceived depth of communication. Ely (1986) believes that language class risk-taking involves a tolerance of possible incorrectness while using the language. Reiss (1985), after measuring tolerance of ambiguity with three situation-specific items, found a positive relationship between tolerance of ambiguity and university foreign language students' ratings of themselves as

language learners. Using the language class risk-taking scale (a 6-item measure developed by Ely 1986), Liu and Jackson (2008) found that there was a significant and positive relationship between language class risk-taking and students' self-rated proficiency in English reading, writing, listening, speaking and the overall English language proficiency. The more risk-taker the student was the higher the self-reported ratings in listening, reading, writing and speaking, with speaking ability being the most positively correlated to language class risk-taking.

The inclusion of risk-taking as a moderator variable can be due to a number of reasons. First, because it is regarded as a personality factor, we want to ascertain whether it could have any affinity with self-assessment which is itself related to the students' personal decisions. Second, scholars consider it important in boosting students' self-esteem because it has a positive role in successful language learning (cf., Brown, 2000). As cited in Brown (2000, p. 149), the concept of risk-taking is described as related to impulsivity and students' making intelligent guesses. Such factors stress the importance of risk-taking as a personality factor in triumphantly learning a second language. "Learners have to be able to gamble a bit, to be willing to try out hunches about the language and take the risk of being wrong" (Brown, 2000, p. 149). Thus, risk-taking was included in the study to see whether it has any role in self-assessment and to what extent students can make intelligent guesses about their speaking ability.

METHOD

The purpose of the study was to investigate self-assessment as an alternative assessment procedure in determining oral proficiency of EFL learners. Taking into account the importance of assessing speaking ability and that of self-assessment in learner-centered pedagogy, this study investigated self-assessment which is an empirically less-researched issue in language testing. Moreover, risk-taking was chosen as a moderator variable

in order to see whether it has any role in students' self-assessment of speaking ability. As mentioned earlier, risk-taking was included in the present study due to the following considerations. First, we wanted to see whether the personality factor of risk-taking has any affinity with self-assessment which is related to students' personal decisions. Second, as cited in Brown (2000), risk-taking is related to impulsivity and students' making intelligent guesses. Thus, risk-taking is included as a moderator variable to see whether it has any role in self-assessment. In other words, we wanted to see to what extent risk-taking is related to students' making intelligent guesses about their speaking ability. Accordingly, the present study aims at investigating the following research questions:

1. Is there any relationship between students' self-assessment of speaking ability and *teacher's rating* in the picture description task?
2. Is there any relationship between risk-taking and students' self-assessment in the picture description task?

Participants

The participants of the study were 79 first year undergraduate EFL students aged 18-26 at the University of Isfahan. 63 of the participants were females and 16 of them were males. The students were unfamiliar with the concept of self-assessment. So, we introduced for them the concept of self-assessment and its potential usefulness in academic settings.

Instrumentation

In this study, a number of tests and scales were used as the instruments to measure different variables under investigation. Their descriptions will appear in order:

1. A picture description task was administered to elicit the participants'

speech samples.

2. The scale of oral proficiency developed by Farhady et al. (1994) was modified so that it could be useful to assess the speaking ability in the picture description task (see Appendix A)

3. The Venturesomeness subscale of Eysenck's IVE questionnaire validated by Kiany and Pournia (2006) was used in order to determine the subjects' levels of risk-taking (see Appendix B & C).

Since we were not testing the speaking ability of the subjects in an interview task, we crossed out the comprehension subscale. Moreover, in Farhady et al.'s (1994) scale each subscale has 6 levels with corresponding behavioral statements. The levels were reduced to 4 and instead of the behavioral statements, the words 'weak', 'fair', 'good' and 'excellent' were used to indicate the levels. This was done in order to keep the students away from the complexities inherent in the statements and also to make the process easy for the students to perform. In this study, all the four subscales were equally weighed. The highest score in each subscale was 4 and the lowest score was 1 and a student's overall speaking ability was considered to be the combination of the scores in each scale. Accordingly, the lowest and the highest possible scores on speaking ability were 4 and 16 respectively.

Procedures

First, the students were provided with a picture description task to elicit their speech samples. Then, they were asked to rewind the tapes and assess their own speaking ability based on our four-point Likert scale of oral proficiency (see Appendix A). Moreover, the participants of the study were given the venturesomeness questionnaire to determine their levels of risk-taking. Based on their scores on risk-taking-test, the subjects were divided into three groups of high, medium and low risk-takers if they fell above +0.5 SD, in $\pm$ 0.5 SD or below –0.5 SD respectively. One of the researchers rated the speech samples based on the same scale that the students utilized in self-

assessment. It should be mentioned that the researcher rated the speech samples twice in order to ensure the reliability of his ratings. The intra-rater reliability was shown to be .86.

Speaking ability can be tested via direct and indirect measures. What follows is a brief description of the two types of measures:

Indirect measures: Indirect measures estimate speaking ability by making use of different elicitation techniques. The most common of these measures are using pictures, making transformations, following commands, retelling, explanation, short talks and role-playing (Farhady et al., 1994).

Direct measures: Unlike indirect measures which are carried out through quasi-realistic activities, direct measures of testing speaking attempt to duplicate the real life situations (*ibid*). Oral interview is the most commonly used test of speaking (Chastain, 1988; Harris, 1969).

Speaking performance can be rated in two ways: holistically and analytically. What follows is a brief discussion on the two types of ratings.

Holistic rating: In holistic rating of speaking, a rater exercises his or her subjective judgment. Rapidity of scoring is considered to be the advantage of this type of scoring (Hughes, 2003). Although it has the disadvantage of not being objective, its reliability can be ascertained by having two or more testers mark the testees separately (i.e., inter-rater reliability) (Walker, 1990) or asking the rater to rate two or more times (i.e., intra-rater reliability) and asking the testers "to grade according to very explicit criteria" (Ur, 1996, P. 135).

Analytic rating: In this approach raters assess each of a number of aspects of performance separately using separate rating scales (McNamara, 2000). The ratings given by the scorers are weighed and added up to determine the final score. According to Bachman and Palmer (1996) analytic scales have two advantages. First, "they allow us to provide a profile of the areas of language ability that are rated". Second, they "tend to reflect what raters actually do when rating samples of language use" (p. 211). Despite having these advantages, analytic scoring has the main disadvantage that scoring will

be time-consuming. Moreover, concentration on different aspects may divert our attention from the overall effect of performance (Hughes, 2003).

In the present study, the speech samples of the participants underwent analytic scoring using the modified version of Farhady et al.'s (1994) oral proficiency scale.

DATA ANALYSIS

In this study, various statistical analyses including both descriptive and inferential statistics were used for different purposes. Descriptive statistics such as means and standard deviations were used in order to check the underlying assumptions of the statistical procedures applied in the study.

The speech samples of the participants underwent subjective judgment by both the students themselves and one of the researchers. To enhance the reliability of subjective judgment, the following considerations were made:

- We explained to the students all the terminologies in the oral proficiency scale and guided them in the marking criteria.
- The rater was the instructor and was trained well enough to carry out rating.

Intra-rater reliability was run to check the reliability of the rating of the speech samples. As stated earlier, the intra-rater reliability estimates were found to be .86. For the purpose of testing the hypotheses, inferential statistical procedures were applied. To test the first null hypothesis, Pearson correlation and Paired samples t-test were run using SPSS 11. Far as the second null hypothesis was concerned, the non-parametric test of Kruskal-Wallis was used to find out the nature of the relationship between the variables of the study. Risk-taking with three levels of high, medium and low was the moderator variable and the self-assessment was the dependent variable of the study.

RESULTS AND DISCUSSION

In order to test the hypotheses of the study, several statistical procedures were utilized. The results and discussions are presented in this section.

Testing the First Null Hypothesis

The first null hypothesis of the study posed that there was no relationship between students' self-assessment of speaking ability and teacher's rating in the picture description task. To test this hypothesis, Pearson correlation was used as the statistical procedure. As shown in Table 1, the correlation between students' self-assessment and teacher's rating was .47.

TABLE 1
The Correlation Between Self-assessment and Teacher's Rating

		Teachrat	Selfases
Teachrat	Pearson correlation	1	.476**
	Sig.(2- tailed)	79	.000
	N		79
Selfases	Pearson correlation	.476**	1
	Sig.(2- tailed)	.000	
	N	79	79

Note: **Correlation is significant at the 0.01 level (2-tailed).
Teachrat = Teacher's rating, Selfases = Self-assessment

The above table indicates that there might be much difference between the students' self-assessment and teacher's rating. Paired samples t-test was run to see whether the difference between the means of teacher's rating and self-assessment are statistically significant. Table 2 shows that the means of teacher's rating and self-assessment are 10.20 and 9.07 respectively.

TABLE 2
Paired Samples Statistics for Teacher's Rating and Self-assessment

		N	Mean	Std. Deviation	Std. Error of Mean
Pair 1	Teachrat	79	10.203	2.648	.298
	Selfases	79	9.076	1.992	.224

Note: Teachrat= Teacher's rating, Selfases = Self-assessment

As shown in Table 3 below, the p value of .00 is much less than .05. This shows that the difference between the means is statistically significant.

TABLE 3
Paired Samples T-test

	Mean	SD	Std Error of Mean	t	df	Sig. (2-tailed)
Teachrat-Selfases	1.127	2.441	.275	4.102	78	.000

Based on the aforementioned analyses, the first null hypothesis was retained. In other words, there was no relationship between students' self-assessment and teacher's rating. Thus, a tentative conclusion that can be reached is that the participants' self-assessment has not been an accurate measure of testing speaking ability in the picture description task.

Testing the Second Null Hypothesis

The second null hypothesis of the study posed that there was no relationship between risk-taking and students' self-assessment. We had one independent variable with three levels and one dependent variable (i.e., self-assessment). Moreover, the self-assessment scores were not normally distributed. Therefore, Kruskal-Wallis which is the non-parametric equivalent to the one way between groups ANOVA (see Hatch & Lazaraton, 1991) was used to find out the nature of the relationship between risk-taking and students' self-assessment. Table 4 shows that there are differences among the mean ranks of self-assessment scores of the three groups (i.e., high, medium and low risk-takers).

TABLE 4
Mean Ranks of Self-assessment Score on Kruskal-Wallis Test

Risk-level		N	Mean Rank
Selfases	1.00	39	44.50
	2.00	12	41.83
	3.00	28	32.95
Total		79	

Note: Risk-level= Risk-taking level, Selfases= Self-assessment

To see whether the differences among the mean ranks of self-assessment scores are statistically meaningful or not, Table 5 should be examined.

TABLE 5
Results of Kruskal-Wallis Test for Self-assessment Scores

	Selfases
Chi-square	4.344
df	2
Asymp. sig.	.114

Note: a. Kruskal-Wallis Test
     b. grouping Variable: risk-level
     c. selfases= self-assessment

As shown in Table 5, the chi-square value of 4.34 is less than the critical value of 5.99 needed for the df of 2 and the p value of .05 to reject the null hypothesis. In other words, the second null hypothesis of the study is confirmed. The table also depicts this by illustrating that the p value (i.e., .36) is above .05. This means that there is not a statistically significant relationship between risk-taking and students' self-assessment. It should be mentioned however that there is a trend of difference in the sense that high risk-taking subjects performed better than medium risk-taking subjects and medium risk-taking subjects performed better than low risk-taking subjects, though such differences were not statistically significant.

The results showed that the first null hypothesis of the study was retained. There was not a significant relationship between students' self-assessment of speaking ability and teacher's rating in the picture description task. Moreover, risk-taking as a personality factor does not have a significant influence on students' self-assessment.

The results of this study both contradict and echo earlier findings in the literature. It contradicts Krausert's (1991) finding that students correctly assessed their speaking abilities. Also, it contradicts Bailey's (1998) finding which suggests that learners accurately assess their own oral production ability. Results of the present study echo investigations whereby learners do not accurately assess their abilities (Brantmeier, 2006; Hilton, Grandy, Green

Kline, & Liskin-Gasparro, 1985; Wesche, Morrison, Ready, & Pawley, 1990). In line with Little (2005) it could be argued that learners whose experience of formal instruction has been traditional and teacher led can not be expected to assess themselves accurately. Our students were not trained in self-assessment. This may explain lack of significant relationship between self-assessment and teacher's rating. As Harris (1997) interestingly states "While doubts about the reliability of self-assessment have been raised, most of these have been where students have received no training" (p. 18). Also, it should be noted that as shown in Table 2, the difference between the means of teacher's rating and students' self-assessment shows that students' assessment was lower than that of the teacher. This finding is in line with Lim's (2007) finding. These contradictory findings stress the need for more and more researches to explore self-assessment as a tool for assessing language ability.

A closer look at the self-assessment questionnaire reveals that students may not fully know what it means to be "good" or "excellent" fluency or accuracy in speaking ability. Furthermore, since our subjects are all freshman students they may not possess the necessary skills to judge the components of speaking ability (i.e., fluency, accuracy, etc.) which underscores the need for more contextualized, criterion referenced self-assessment instruments for L2 speaking. Freshman students may need training so that they could make reliable judgments on their speaking abilities. This suggestion echoes other studies which call for learner training to help learners have a better judgment of their language abilities (e.g., Cheng & Warren, 2005; Lim, 2007; Rolfe, 1990).

The results also showed that the second null hypothesis of the study was confirmed. In other words, it was shown that there was no relationship between risk-taking and students' self-assessment in the picture description task. However, there was a trend of difference in that high risk-takers tended to perform better than medium and low risk-takers and medium risk-takers tended to perform better than low risk-takers, though these results were not statistically significant. The findings are to some extent consistent with studies like Ely (1986); Evensen and Bednar (as cited in Jonassen, &

Grabowsky, 1993) who found that risk-taking was related to success in L2 learning. Also, the results are to some extent in agreement with Liu and Jackson (2008) who found that there was a significant and positive relationship between language class risk-taking and students' self-rated proficiency in L2 English speaking. As success in language learning is related to students' risk-taking, we assume with Rubin (1994 cited in Brown, 2000) that successful language learners (hence, high risk-takers) make accurate guesses.

It might be suggested that self-assessment as an alternative assessment procedure can be utilized if teachers are concerned with long term educational benefits rather than students' immediate success or failure to imitate or supplement teacher assessment. Also, it could be suggested that self-assessment be used in low stakes decisions such as placement tests whereby as Bachman (2004) asserts the "decision errors are easy to correct" (p. 12).

Taking into account the need to "offer assessment models that are more educational, democratic, ethical and at the same time valid" (Shohamy, 2001, pp. 389-90) we suggest that self-assessment and teacher assessment work together instead of being mutually exclusive and the combination of the two should be preferred to the use of either one alone. It might be suggested that teachers use multilevel model of second language assessment which attempts to integrate quantitative data with qualitative descriptions. Using multiple sources of information might show us a better picture of students' achievements. As Brown and Hudson (1998) put it "using multiple sources of information of varying types increases the collective reliability of that information and of any decisions that may result from interpreting it" (p. 671). Moreover, to be effective, self-assessment should be integrated with every day classroom activities so that self-assessment becomes part of regular procedures for dealing with speaking (Harris, 1997).

It was shown that low risk-takers have a low perceived self-competence in English speaking. To help these students increase their self-perceived competence in speaking ability, teachers may be recommended to enhance

the students' self-confidence by focusing on their achievements (Liu & Jackson, 2008). Also, during the course teachers can try different means of improving students' interest and motivation to use the language which may lead to more use of the language which in turn may help students increase self-ratings of their speaking ability. However, the point to be noted here is that if teachers want to improve their students' self-assessment of speaking ability, they should encourage students to make guesses willingly and "value them as persons for those risks that they take" (Brown, 2000, p. 150).

CONCLUSION

The present study examined the relationship between self-assessment and teacher's rating of freshman EFL students' speaking ability. Moreover, the relationship between the personality factor of risk-taking and self-assessment was investigated. The results revealed that self-assessment as measured by non-descriptive self-assessment items with freshman students can not be an accurate and dependable measure of speaking ability. However, as suggested earlier self-assessment can be used along with other alternative measures in language assessment to have a better picture of our students' achievements. As Harris (1997) states, this is particularly important in large classes in which it is impossible to continually monitor individual performance.

There was not a significant relationship between risk-taking and self-assessment. However, high risk-takers tended to assess themselves better than medium or low risk-takers and medium risk-takers tended to assess themselves better than low risk-takers. Thus, it might be recommended that teachers help their students have an optimal level of risk-taking so that they can make better assessment of their speaking ability. Overall, the findings stress the need for more investigations concerning students' self-assessment of speaking ability and the factors-specifically personality factors-which might influence self-assessment.

Few studies have been done on the notion of self-assessment and speaking

ability and there has been scant research to date which seeks the role of risk-taking in students' self-assessment of speaking ability; therefore, there are several areas of potential research. In this study, the modified version of Farhady et al.'s (1994) scale for oral proficiency was used to measure the speaking ability. Other kinds of tests can be used as measurement tools for speaking ability. The interaction of age, gender, language proficiency and self-assessment that was left untouched in the present study is another potential area that can be handled in further research.

The present research aimed at investigating on the one hand the relationship between teacher's rating and self-assessment of speaking ability and on the other hand the relationship between risk-taking and self-assessment. However, owing to some constraints there were some limitations for this study. First, speaking ability was measured based on students' performance on a picture description task. This is a limitation of the study as Farhady et al. (1994) mentioned oral interview is the most face valid test of speaking. Also, oral interview is the most commonly used test of speaking (Chastain, 1988; Harris, 1969). Second, the inferences drawn from this study is limited due to the nature of the participants. The participants of the study were 79 freshman EFL students at the University of Isfahan. Other researches with larger sample size and with students of various proficiency levels at different universities can be done to enhance the generalizability of the findings. Third, the self-assessment rating scale utilized in the present study has 4 subscales, each with 4 levels of "Weak", "Fair", "Good" and "Excellent". As stated earlier, there is a need for more contextualized, criterion-referenced self-assessment instruments for L2 speaking.

THE AUTHORS

Mansoor Tavakoli is an assistant professor of English and has taught TEFL courses at three levels of B.A, M.A, and Ph.D. in the Department of English Language at the University of Isfahan for 12 years. His research interests are

language teaching and assessment. His recent publications include *Genre Analysis: An Investigation of the Discussion Sections of Applied Linguistics Research Articles* with Z. Amirian and Z. Kassaian (2008) and *Assessment in a multilingual world: English as an International Language (EIL)* with N. Shomoossi (2008).
Email: mr.tavakoli14@gmail.com

Behrooz Ghoorchaei is a PhD candidate in TEFL at the University of Isfahan. He has taught English courses in different institutes and universities in Iran. His research interests are language testing and assessment, language teaching and Sociolinguistics. His recently published article is *Analysis of Borrowed Words in the Spoken Discourse of Turkmen Native Speakers in Iran: A Sociolinguistic Perspective* with A. Eslami Rasekh and N. Shomoossi (2008).
Email: behroozghoorchaei@gmail.com

ACKNOWLEDGMENT

REFERENCES

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. New York: Heinle & Heinle.
Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151-167.

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System, 34*(1), 15-35.

Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum.* Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

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

Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). New York: Longman.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. New York: Longman.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, New York: Longman.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(3), 93-121.

Chastain, K. (1988). *Developing second language skills: Theory and practice* (3rd ed.). San Diego: Harcourt Brace Jovanovich.

Clegg, K., & Bryan, C. (2006). Reflections, rationales and realities. In C. Bryan, & K. Clegg, (Eds.), *Innovative assessment in higher education* (pp. 216-227). London: Routledge.

Coombe, C. (2001). Learner-directed assessment in ESL. *TESL-EJ, 5*(1), 1-3.

de Bot, K. (1992). Self-assessment of minority language proficiency. In L Verhoeven, and De Jong, John. H. A. L. (Eds.), *The construct of language proficiency: Applications of psychological models to language assessment.* (pp. 137-146). Amsterdam, Netherlands: John Benjamins Publishing Company.

Dickinson, L. (1987). *Self-instruction in language learning.* Cambridge: Cambridge University Press.

Ely, C. M. (1986). An analysis of discomfort, risk-taking, sociability, and motivation in the L2 classroom. *Language Learning, 36*(1), 1-25.

Farhady, H., Jafarpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice.* Tehran: SAMT Publications.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education, 20*(3), 289-300.

Harmer, J. (2001). *The practice of English language teaching* (3rd ed.). Harlow: Longman.

Harris, D. P. (1969). *Testing English as a second language.* New York: McGraw-Hill.

Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal, 51*(1), 12-20.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle & Heinle.

Hilton, T., Grandy, J., Green Kline, R., & Liskin-Gasparro, J. E. (1985). *Final report: The oral language proficiency of teachers in the United States in the 1980s- an empirical study*. Princeton, NJ: Educational Testing Service.

Huerta-Macías, A. (2002). Alternative assessment: Responses to commonly asked questions. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 338-343). Cambridge: Cambridge University Press.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Jonassen, D. H., & Grabowsky, B. L. (1993). *Handbook of individual differences, learning and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kiany. G. R., & Pournia, Y. A. (2006). The relationship between risk-taking and the syntactic complexity and grammatical accuracy of the Iranian intermediate EFL learners' descriptive and expository writing. *Pazhuhesh-e Zabanha-ye Khareji, 27(Special issue, English),* 143-64.

Krausert, S. R. (1991). *Determining the usefulness of self-assessment of foreign language skills: Post-secondary ESL students' placement contribution.* Ph. D Dissertation, University of Southern California.

Le Blanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*(4), 673-687.

Lim, H. (2007). A study of self- and peer-assessment of learners' oral proficiency. *CamLing Proceedings,* 169-176. Retrieved from the Web April 20, 2008. http://www.ling.cam.ac.uk/camling/Manuscripts/CamLing2007_Lim.pdf

Little, D. (2005). The common European framework and the European language portfolio: Involving learners and their judgments in the assessment process. *Language Testing, 22*(3), 321-336.

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, *92*(1), 71-86.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Malabonga, V., Kenyon, D. M., & Carpenter, H. (2005). Self-assessment, preparation and response time on a computerized oral proficiency test. *Language Testing*, *22*(1), 59-92.

McNamara, T. (2000). *Language testing*, Oxford: Oxford University Press.

Nakamura, Y. (1993). *Measurement of Japanese college students' English speaking ability in a classroom setting*. Unpublished Ph.D. Dissertation, International

Christian University, Tokyo.

Nedzinskaite, I., Svencioniene, D., & Zavistanaviciene, D. (2006). Achievements in language learning through students' self-assessment. *Studies about language,* 8, 84-7. Retrieved from the Web April 16, 2008. http://www.ceeol.com/aspx/ getdocument.aspx?logid=5&id=337ad699-f307-4044-8f10-3bc8f9447378

Oscarson, M. (1978). *Approaches to self-assessment in foreign language learning.* Oxford: Pergamon Press.

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

Porter, P., & Van Dommelon, D. (2005). Integrating assessment with grammar-for-writing instruction. Proceedings of the CATESOL state Conference, Retrieved from the Web May 3, 2008. http://www.catesol.org/Porter_vanDommelen.pdf

Puhl, C. A. (1997). Develop, not judge: Continuous assessment in the ESL classroom. *English Teaching Forum*, *35*(2), 2-9.

Reiss, M. A. (1985). The good language learner: Another look. *The Canadian Modern Language Review, 41*, 511-523.

Richards, J. C., Platt, J., & Platt, H. (1992). *Dictionary of language teaching and applied linguistics* (2nd ed.). Harlow, Essex: Longman.

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice.* Cambridge: Cambridge University Press.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, *15*(1), 1-20.

Ross, S. (2005). The impact of assessment method on foreign language proficiency growth. *Applied Linguistics*, *26*(3), 317-42.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing, 18* (4), 373-91.

Strong-Klause, D. (2000). Exploring the effectiveness of self-assessment strategies in ESL placement. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 49-73). New Jersey: Lawrence Erlbaum Associates, Inc.

Ur, P. (1996). A *course in language teaching practice and theory*. Cambridge: Cambridge University Press.

von Elek, T. (1985). A test of Swedish as a second language: An experiment in self-assessment. In Y. Lee, A. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 47-57). Oxford: Oxford University Press.

Walker, C. (1990). Large scale oral testing. *Applied Linguistics, 11*(2), 200-219.

Wang, H., & Wang, W. (2007). The addition of an affect test and self-assessment into ESL writing assessment: Process and effect. A case study in a non-English

major postgraduate ESL writing class. *Asian EFL Journal, 20.* Retrieved from the Web April 20, 2008. *http://www.asian-efl-journal.com/pta_May_07_hwyw.php*

Wen, W. P., & Clement, R. (2003). A Chinese conceptualization of willingness to communicate in ESL. *Language, Culture and Curriculum, 16*(1), 18-38.

Wesche, M. B., Morrison, F., Ready, D., & Pawley, C. (1990). French immersion: Post secondary consequences for individuals and universities. *The Canadian Modern Language Review, 46*(3), 430-451.

Yorke, M. (2001). Formative assessment and its relevance to retention. *Higher Education Research and Development, 20*(2), 115-126.

APPENDIX A

Self-assessment rating scale

Name: _____

Native language: _____

Years of formal study: _____

Age: _____

Instruction

Please think about your skill in speaking English and select the level that most closely matches your skill in spoken production.

1 = Weak   2 = Fair   3 = Good   4 = Excellent

Now mark your answers here.

A:	Pronunciation	1	2	3	4
B:	Structure	1	2	3	4
C:	Vocabulary	1	2	3	4
D:	Fluency	1	2	3	4

APPENDIX B

Persian version of venturesomeness subscale of Eysenck's IVE questionnaire (Kiany & Pournia 2006)

پاسخگوی محترم: لطفاً گزینه ای را که در مورد شما صدق میکند، علامت بزنید.

1= همیشه	2= بیشتر اوقات	3=گاهی	4= به ندرت	5= هیچ وقت

5	4	3	2	1		
5	4	3	2	1	از اسکی روی آب لذت می برم.	1
5	4	3	2	1	در موقع خرید به جای اینکه مارکهای جدید را به امید یافتن جسی بهتر امتحان کنم ترجیح می دهم مارکهایی را انتخاب کنم که میدانم قابل اعتمادند.	2
5	4	3	2	1	از خطرپذیری لذت می برم.	3
5	4	3	2	1	از پرش با چتر نجات(چتربازی) لذت می برم.	4
5	4	3	2	1	فکر می کنم پیاده طی کردن مسافتهای طولانی و جلوی اتومبیلهای در حال عبور را گرفتن برای مجانی سوار شدن راه خیلی خطرناکی برای مسافرت است.	5
5	4	3	2	1	شیرجه زدن در استخر از روی تخته بلند شیرجه را دوست دارم.	6
5	4	3	2	1	از تجربیات و احساسات جدید و هیجان انگیز استقبال می کنم حتی اگر تا حدودی ترسناک و غیر متعارف باشند.	7
5	4	3	2	1	دوست دارم چگونگی پرواز با هواپیما را یاد بگیرم.	8
5	4	3	2	1	درک افرادی که جان خود را به خاطر کوهنوردی به خطر میاندازند برایم مشکل است.	9
5	4	3	2	1	انجام کارهایی را که تا حدودی ترسناک هستند دوست دارم.	10
5	4	3	2	1	ترجیح میدهم تدریجاً وارد آب سرد دریا شوم تا اینکه یکباره درون ان بپرم و یا درون آن شیرجه بزنم.	11
5	4	3	2	1	از هیجان اسکی بازی با سرعت خیلی زیاد در شیب یک کوه بلند لذت می برم.	12
5	4	3	2	1	غواصی کردن را دوست دارم .	13
5	4	3	2	1	از سریع رانندگی کردن لذت می برم.	14

5	4	3	2	1	غارنوردی را دوست دارم.	15
5	4	3	2	1	از شغلی که تا حدودی با خطر همراه است دوری می جویم.	16

APPENDIX C

English translation of Persian version of Venturesomeness subscale of Eysenck's IVE questionnaire (Kiany & Pournia, 2006)

Please mark the choice which best indicates the extent to which you agree or disagree with the statement.

1 = Always   2 = Usually   3 = Sometimes   4 = Rarely   5 = Never

1	I enjoy water skiing.	1 2 3 4 5
2	I usually prefer to stick to brands I know are reliable, rather than trying new ones on the chance of finding something better.	1 2 3 4 5
3	I enjoy taking risks.	1 2 3 4 5
4	I enjoy parachute jumping.	1 2 3 4 5
5	I think hitch-hiking is too dangerous a way to travel.	1 2 3 4 5
6	I like diving off the high-board.	1 2 3 4 5
7	I welcome new and exciting experiences and sensations, even if they are a little frightening and unconventional.	1 2 3 4 5
8	I like to learn to fly an airplane.	1 2 3 4 5
9	I find it hard to understand people who risk their necks climbing mountains.	1 2 3 4 5
10	I like doing things that are a little bit frightening?	1 2 3 4 5
11	I prefer to enter cold sea water gradually rather than diving or jumping straight in?	1 2 3 4 5
12	I enjoy the sensation of skiing very fast down a high mountain slope.	1 2 3 4 5
13	I like to go scuba diving.	1 2 3 4 5
14	I enjoy fast driving.	1 2 3 4 5
15	I like to go exploring in caves.	1 2 3 4 5
16	I am put off by a job involving quite a bit of danger.	1 2 3 4 5