ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 2, No. 6, pp. 1289-1297, November 2011
© 2011 ACADEMY PUBLISHER Manufactured in Finland.
doi:10.4304/jltr.2.6.1289-1297

The Effects of Acquaintanceship with the Interviewers and the Interviewees' Sex on Oral Interview as a Test Technique in EFL Context

Mohiadin Amjadian
Kurdistan Medical University, Sanandaj, Iran
Email: mohiadin72@yahoo.com

Saman Ebadi
Allameh Tabatai'e University, Tehran, Iran
Email:Samanebadi@gmail.com

***Abstract*—This study explored the possible relationship between the acquaintanceship of the interviewees with the interviewers, the interviewees' gender and their oral performance in oral language interviews as a test technique in L2 situations. Among 117 participants of the study ,60 university students were chosen based on Michigan test of language proficiency and were randomly assigned into groups (A and B) of familiar and unfamiliar with the interviewer including both males and females. They were interviewed using IELTS interview framework with some modifications, 30 with familiar interviewer and 30 with unfamiliar one. The results of the study indicated that the familiarity and unfamiliarity of the interviewee with the interviewers might not affect the oral performance (scores) of the subjects .The results also evidenced that the effect of the candidates' gender was significant in their scores with the males outperforming the females. This can be due to the sex of the interviewers, all of whom were males for both male and female groups. Although the relationship between sex and oral performance was significant, the interaction effect of the candidates' sex and their acquaintanceship with the test – givers on the oral performance of the subjects was not so significant. The findings of this study have some implications for both speaking instruction and assessment of oral proficiency in EFL context.**

***Index Terms*—oral interview, variation, oral performance, acquaintanceship**

I. Introduction

Testing the language proficiency, especially oral performance, is one of the most complicated issues in the literature of language testing. There are many factors, which might affect different tests of oral proficiency in language testing in EF /SL contexts. Among different factors, which might affect oral performance of the learners, are the effect of the test-givers and test-takers' characteristics on their performance on such tests (Kunnan, 1995; O'Sul1ivan 2002). Some studies have been conducted on the effects of the test givers and test-takers' characteristics such as education level, social status, age, sex and personality on oral performance of different candidates (O' Loughlin 2002; O'Sul1ivan and Porter 1995; Porter and Shen 1991; Farhadi 1982). These studies have examined the effects of multiple background characteristics on language test performance. For example, Farhadi (1982) found significant relationship between sex, university status, academic major and nationality and the subjects' performance on several measures of language ability. Sullivan (2002) explored the possible effect of pair-task performance of test-takers' familiarity with their partner and concludes that learners vary in their oral performance when interacting with familiar or unfamiliar speakers. Gender, social status and interaction style (male / female) were the factors studied in Porter and Shen (1991). Here, fourteen men and fourteen women of various nationalities were interviewed by two interviewers, one male and one female. The results of the research showed no significant difference in the results of the status variable but the gender was significant with achieving higher scores by female interviewers.

What makes the present study different from other studies done in this area is that, unlike other studies (Porter 1991b; O'Loughlin 2002; O'Sullivan 2002) in which the sex of the interviewers and its possible effects on the interviewees' oral performance was studied, here, the sex of the interviewees and also its possible relationship with the acquaintanceship with the interviewer was taken into consideration. Moreover, the number of the subjects (60 university student) is much more than the number of subjects in the other studies (e.g, 32 in O'Sullivan 2002, 16 in O'Loughlin 2002 and 13 in Porter 1991a). This increase in the number of the subjects would have positive effects on the reliability and validity of the outcomes. In this study, the researchers tried to investigate the effect of acquaintanceship of test-takers with test-givers on their oral performance, on one hand, and the interaction effect between this variable and the sex of the test-takers on their performance on the other. O'Sullivan (2000) operationalized familiarity with interviewer as "familiar interviewer is, preferably their own professors or classmates who have been studying in the same class, at least, since the last two years.

The purpose of the study was to clarify the effects of these two variables on the oral performance of learners in oral language tests in L2 contexts. The probable relationships between these variables have been investigated using IELTS interview framework in order to help the L2 language teachers to find the most appropriate ways of testing speaking skill and getting familiar with possible effects of some social variables like sex, familiarity and unfamiliarity which might influence the test takers' performance in speaking tests among Iranian learners. In fact, knowing and predicting the possible effects of such variables in the oral performance of the L2 learners would result in administrating a more reliable and valid test of oral proficiency. The following research questions guided the current study:

1. Does acquaintanceship with the test giver (interviewer) affect the performance of the test- takers (interviewees) in an L2 interview situation?

2. Is there any relationship between the sex of the test-takers (L2 learners) and their oral performance in an interview situation?

3. Is there any interaction effect between the candidates' gender and their acquaintanceship with the test-givers on the oral performance of L2 learners in an interview as a test technique?

In order to come up with scientific results, the researchers has selected the following null hypotheses based on the research questions:

$N0_1$: The Acquaintanceship with the test-giver does not affect the oral performance of the test-takers in an L2 interview situation.

$N0_2$: There is no significant relationship between the sex of the test-takers (L2 learners) and their oral performance in an interview situation.

$N0_3$: there is no significant interaction effect between the candidates' gender and their acquaintanceship with the test givers on the oral performance of the L2 learners in an interview situation.

II. Review of Related Literature

A. *Methods of Testing Speaking (Oral language)*

Oral language in the language classrooms is the most problematic of all the skills to assess. It involves the combination of dimensions that are not necessarily correlated and do not lend themselves well to objective testing, so that a performance may get 7.0 as far as its appropriacy is concerned but get 5.0 in terms of accuracy (Baker ,1989). There seems, not yet, to have clear answers to questions about the criteria of testing speaking and weighting the factors which might be considered to affect this skill. A speaker might produce all the sounds correctly but not make any sense or have great difficulty with phonology and grammar and yet be able to get the message across (Heaton, 1990). Speaking ability can be tested in two ways, directly and indirectly (Farhady, Jafarpoor, Birjandy, 1995). Indirect testing of speaking is carried out through quasi-realistic activities and the direct one is achieved through activities that try to duplicate as closely as possible the setting of a real life situation. Common indirect measures of speaking ability include, picture description, performing commands, retelling stories and role plays. In these tasks, we want the subjects to elicit their language knowledge indirectly and through performing different tasks. The second group of oral performance testing includes interviews and short talks, in which we want the testees to elicit their language knowledge directly and through putting them in real life language settings. Among these groups of speaking techniques oral interviews were used to assess the participants' oral performance in this study.

Oral Interviews as a Test Technique

Interviews are defined as direct conversations between an investigator and an individual or a group of individuals in order to gather information (Longman Dictionary of Applied linguistics 1992). They are situations in which the test-taker and the test-givers carry on a conversation. The advantage of an interview is that, it attempts to approximate a conversation situation. How and in what ways we can prepare tests of interview to assess oral language performance of the learners, is not our concern in the study. As interview seems to be the most valid direct type of speaking tests and appears to offer a realistic means of estimating the overall speaking ability of the learners in a natural speech situation (Farhady, Jafarpoor, BiIjandy; 1995), we were dealing with the effects of external factors which might influence the interview and its results as an oral performance test in this study. Among these factors, is the effect of the interviewers' characteristics like sex, age and acquaintanceship with the candidates on the subjects' oral performance in tests of interview (Weir 1993). Generally speaking, interviews are conducted in different formats among which are IELTS and interviews based on tasks.

The assessment in IELTS speaking takes into account evidence of communicative strategies and appropriate and flexible use of grammar and vocabulary (O'loughlin, 2002). IELTS provides a profile of candidates ability to use English. Candidates receive scores on a band scale from 1 to 9. Another kind of interview which is used for evaluating the oral performance of L2 learners is interview based on tasks or the personal information exchange (PIE) task, which requires the candidates to speak to their partner about a topic relating to their university life. An example of this is: "please tell your partner what were your first thoughts when you started at university" (O'sullivan, 2002).

B. *Holistic vs. Analytical Scoring of Speaking Skill*

As Madsen (1983) states, on a speaking test, getting the students to say something appropriate is only half the job and the scoring procedure is equally challenging. In fact, the complex task of scoring writing composition is the only thing that might match the challenge of scoring a speaking test. The choice of scoring system tends to depend on one of two things: The first one is that, how well trained the rater is to evaluate oral communication and what factors are chosen to be evaluated (Madsen 1983). Generally speaking, the oral language interviews are scored holistic1y or analytically. In both methods, it is better to record the interviews and marking (scoring) should be done by more than one rater in order to check the reliability of the scores given to the interviewees. Oral performance of the candidates can be assessed in two different ways: Holistic and analytic.

A scoring procedure in which an overall impression of the rater, according to which the interviewee receives "excellent" "good" "fail" and "poor" or 'pass / fail" is called a subjective or holistic scoring. While, in an analytical or objective scoring, the rater rates the interviewees" performance separately on scales that pertain to accent, structure, vocabulary, fluency and comprehension. The advantage of holistic grading as Madsen (1983) and Hughes (2003) state is that, it concentrates on communication while not overlooking the components of speech and it is a rapid way of scoring. The limitation is that, a great number of teachers not skillful in analyzing speech, find it confusing to evaluate so many things simultaneously.

In analytical or objective scoring, the speaking items based on the right-wrong arrangement for low level test-takers and partial credit for upper levels, are given marks and then the total sum of the marks will be the speaking test score. In right-wrong system the items are supposed to be right or wrong and get 0 or 1 point. In partial credit system, the test-taker allows 2 points for fully correct answers, 1 for partially correct answers and 0 for unaccepted or unintelligible answers (Underhill 1987). Analytic or objectified scoring is a satisfactory way to evaluate speaking ability, because it can yield more consistent and reliable scores. On the other hand, it might neglect the main goal of speaking, communication, and pays more attention to less important components of speech like grammar, accent etc. In this study, the researchers has tried to develop a local scoring system for L2 learners which have benefited the partial credit system and it uses both holistic and analytic measures to score the candidates' performances.

C. *Familiarity and Gender in Oral Interviews*

A number of researchers distinguish between test features which are irrelevant to the ability which is being measured and those which are relevant to that ability (Porter and Shen Shu Hung, 1991; O'sullivan, 1995; O'sullivan and Porter, 1995; O'Sullivan 2000). O' Sullivan' s (2002) study explored the possible effect of pair-task performance of test-takers' familiarity with their partner and concludes that learners vary in their oral performance when interacting with familiar or unfamiliar speakers. He did the research with 32 Japanese students and maintained that the results might not be generelizable to other contexts. In another study, O' Sullivan (2000) investigated the effect of gender on oral proficiency interview performance. Twelve Japanese learners were interviewed, once by a man and once by a woman. Video tapes of these interactions were scored by trained examiners. Comparisons of the scores indicated that, in all but one case the learners performed better when interviewed by a woman, regardless of the sex of the learners. Analysis of the learners' interviews indicated systematic gender differences, with producing more accurate language by the females. Porter (1991 a) examined the oral performance of thirteen Arab learners by known and unknown interviewers and found no evidence to support his hypothesized interlocutor-acquaintanceship effect. In fact, acquaintanceship with the interviewers seemed not to play a significant role in the candidates' performances. Cholewka (1997) investigated the linguistic behavior of adult second language learners in an oral interview situation. The subjects in this study were six ESL learners (one female and five male) with the same first language. An oral interview conducted twice with a four-day interval, by two different interviewers. The findings of the study revealed that, in unfamiliar setting in an oral interview situation with unknown interlocutors, ESL learners revert to their native language and produce lower level of oral proficiency in the test. This finding suggests that, a task in an unfamiliar real-life situation may elicit significantly higher proportion of language transfer errors than the same task performed in a familiar environment.

III. Methodology

Since, in this study more than one independent variable were involved; acquaintanceship and gender, the hypotheses were investigated in a factorial design and through two-way ANOVA procedures which show not only the main effects of two independent variables on the dependent variable but also their possible interaction effect.

A. *Participants*

The participants of this study included 117 Iranian (B.A) English university students. They were studying at seventh and eighth semesters. A Michigan general proficiency test was used to screen the required number of the candidates who were supposed to take part in the main procedure of the study, oral interviews. Among them, 60 subjects who were supposed to be almost at the same level of general language proficiency were selected; those with the scores between -1.2 SD and + 1.2 SD from the mean score. Then, they were randomly divided into two equal groups of 30, each one consisting of both males and females. In fact, there were four groups of almost 15 in the study: 1. Familiar with the interviewer, female candidates, 2. Familiar with the interviewer, male candidates, 3.Unfamiliar with the interviewer, female candidates and 4. Unfamiliar with the interviewer, male candidates.

B. *Instrumentation*

The following instrumentations were used to tap into the variables at hand in the study:

1. Oral Interview (lELTS)

The kind of oral interview framework used here to measure the candidates' oral production was the oral interview section of the IELTS (the International English language Testing System). It is a four-skill test employed in the selection of prospective non-native speakers of English to universities in such countries as Canada and the UK (O'Loughlin 2002). The version of the speaking sub-test lasts between 10-15 minutes but it took between 15-20 minutes in this study, because in order to get a natural sample of speech, some parts of the official interview were modified; a picture description was used in the third phase of the interview in order to motivate the candidates speak as much as possible. This modification is justified through what Weir (1993) states, "The flexibility of the interview is a major strength; The interview can be modified in terms of the pace, scope and level of the interaction" (P. 56). Choosing IELTS as a test instrument made the researchers able to test their performances through objective scores. Due to not having access to the IELTS speaking scoring scale, a local scoring scale was developed by the researchers to assess the final interviews.

2. Local Scoring Scale

To assess the oral interviews, some reliable scoring scales were needed. Since the IEL TS oral interview scoring system is not public, the researchers developed a scoring system using some international and local scoring systems like FCE and ALIGU tests. The rubric, the researchers developed in this study, is not a task -specific one and it can be used to assess different oral tasks given to the subjects. On the other hand, the researchers tried to develop a speaking scale which is suitable for S/FL testing situations and can be used by the ordinary teachers and practitioners. It does not require trained raters and it is supposed, based on the pilot study results, to be valid and reliable. To test its reliability and validity, one hundred candidates' oral performances were, firstly, rated by two qualified teachers holistically and then for the second time their performances were rated using the developed local scoring scale. Then, the correlation of the two sets of scores were calculated and the researchers found an acceptable correlation (0.79) between the two measures(scores), therefore, the researchers were convinced that the scoring scale benefits from an acceptable degree of inter-rater reliability. On the other hand, it is a combination of holistic and analytic scales together. In fact, in this rubric, conducting holistic and analytic scoring of the oral performance separately at two different times was not required. It is holistic, because it includes well defined levels for each component of the speaking skill, fluency, accuracy, vocabulary, pronunciation and gives a general impression of the oral performance targets to the rater or teacher. However, it is supposed to be analytic and objective, because in addition to having well and clearly defined levels, each level is given a numerical value instead of giving them nominal values like fair, good, adequate and sum of these numerical values for various components is the subjects' total score based on which he/she is ranked in a group of subjects taking oral language interview as a test. Each component is assessed from 1 to 5 and the whole speaking skill, including five components, is assessed from 1 to 25. (see Appendix A for more details).

3. Michigan Test

A general language proficiency test was required to screen the qualified subjects to take part in the oral interviews. A Michigan Test (1990) including 90 multiple-choice items of vocabulary, grammar and reading comprehension, was given to 117 English university students for 70 minutes to accomplish in order to come up with a homogeneous number of subjects. Then, after scoring the test, 60 subjects were selected to take part in the interviews, those between 1.2 SD above and below the mean score.

IV. PROCEDURES

At the outset of the study, a Michigan Test of language proficiency was given to 117 English university students to screen a qualified number of the subjects to take part in the oral interviews. Sixty subjects, males and females, were selected among those whose scores were between 1.2 SD above or below the mean score. They were ranked from the highest score to the lowest and they were divided into two equal groups of thirty, odds for group A and evens for group B. The number of males and females were almost equal for both groups. Group A, by chance, was chosen to be interviewed by an unfamiliar interviewer, who was the researchers himself, and group B was chosen to be interviewed by familiar interviewers, who were the subjects' classmates or professors chosen by the candidates themselves and they were with each other, at least, for the last two years. Of course, the familiar interviewers took part in a four-hour session of TTC in order to be made familiar with the IELTS oral interview framework and to conduct the interviews consistently and skill fully over various subjects. After conducting the interviews**,** they were tape recorded and finally they were rated by a trained rater twice with an interval time (one week) using a local rating scale (see Appendix A for the scale). Intra-rater reliability of the two sets of scores was calculated through correlational procedures and a correlation degree of 0.83 convinced the researchers that the rating system was reliable. The mean of the two scores as their final score for both groups were investigated, using two-way ANOV A procedures, in order to find possible relationships between the independent variables and the scores. Moreover, the scores for the different parts of the spoken data based on the scale including pronunciation, fluency, accuracy, comprehensibility and vocabulary, for each candidate's performance were taken out. Then, using factorial ANOVA, the performances of the candidates in each group, male and female familiar with the interviewer and male and female unfamiliar with the interviewer, were compared and contrasted.

V. DATA ANALYSIS

To begin to investigate the link among the variables, the tape-recorded interviews were scored to study the candidates' performances through their raw scores. The data gathered through the scores given to the candidates were analyzed using two-way ANOVA procedure with follow up mean score calculations to indicate any possible relationships between the variables. In order to investigate the possible relationship between the acquaintanceship of the interviewees with the interviewer and the candidates' oral performance in research question 1, a two-way ANOVA procedure was run at significance level of et. = 0.05. The results (table 1) shows that, although there is a trend that the familiar groups' mean score is higher (Table 2), the relationship is not so significant at this level (The amount of observed F does not exceed the critical F 2.144 < 4.02), to reject this null hypothesis. Put another way, we can say that there is not a significant relationship between these two variables and familiarity or unfamiliarity with the interviewer has not influenced the oral performance of the subjects in both groups**.** This result, unlike that of O'Sullivans' (2002) which claims that there is a significant relationship between the acquaintanceship with the interviewer and the subjects' oral performance in oral language interviews, is in line with Porter's (1991 a) study in which the effects of familiarity and unfamiliarity with the interviewer was investigated and no evidence was found to show a significant effect.

The Results are briefed as follows:

TABLE (1)
TESTS OF BETWEEN-SUBJECTS EFFECTS
Dependent Variable: Test Scores (Oral Interview Scores)

Source	Type III sum of squares	df	Mean square	F	Sig.	Eta Squared
Corrected model	105.906	3	35.302	3.696	.017	.165
Intercept	14877.214	1	14877.214	1557.743	.000	.965
Acquaintanceship	20.479	1	20.479	2.144	.149	.37
Sex of candidate s	66.699	1	66.699	6.984	.11	.111
Acquaintanceship*sex	7.842	1	7.842	.821	.369	.014
Error	534.828	56	9.550			
Total	15556.000	60				
Corrected total	640.733	59				

a R Squared= 165 (Adjusted R Squared=121)

LEVENE'S TEST OF EQUALITY OF ERROR VARIANCES
Dependent Variable: Scores

F	df1	df2	Sig.
1.737	3	56	.170

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: intercept +ACQAUAIN+SEX+ACQAUAIN*SEX

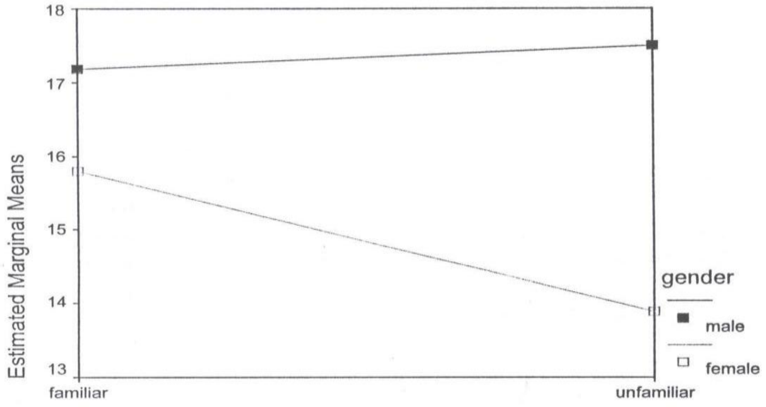
To investigate the second hypothesis, the possible relationship between the sex of the subjects and their oral performance in oral language interviews, a two-way ANOV A was run at the significance level of $a = 0.05$ in SPSS. Based on the results in table (1), the researchers could claim that the relationship between these two variables was significant (Observed F exceeds Critical F 6.94 > 4.02), therefore, null hypothesis two was rejected. In fact, we can claim that there is a significant relationship between the sex of the subjects and their oral performance. To investigate whether the male or female group outperformed in this respect, the groups' mean scores were calculated and it showed that the males with a mean score of 17.33 outperformed the females with a mean score of 14.75. The results were briefed as follows:

TABLE 2 (MEAN SCORES)
DESCRIPTIVE STATISTICS
Dependent Variable: Scores

Acquaintanceship	gender	mean	Std. Deviation	N
Familiar	Male	17.2000	3.5295	15
	Female	15.8000	3.1214	15
	Total	16.5000	3.3502	30
Unfamiliar	Male	17.5000	3.5032	14
	Female	13.8889	2.3736	16
	Total	15.3333	3.3460	30
Total	Male	17.3333	3.4530	29
	Female	14.7576	2.8617	31
	Total	15.9167	3.3713	60

To delve into the investigation of the third hypothesis, the possible relationship between any probable interaction effect of the sex of the subjects and their acquaintanceship with the interviewers on their oral performance, a two-way ANOVA was run at the significance level of $a = 0.05$. The results in table (1) indicated that there was not a significant relationship between them (Observed F is smaller than Critical F, 621 < 3.17). In other words, this null hypothesis is not rejected at this significance level. It can be claimed that the effect of sex, which was significant alone, has been moderated by effect of acquaintanceship with the test-givers. However, females' performance in unfamiliar group was lower than that of familiar group, but the males' performance in both groups, familiar and unfamiliar, was, to some extent, similar. It can be explained by the fact that, the females were under the influence of two different factors at the same time, the sex of the interviewer and unfamiliarity with the interviewer in the unfamiliar group, while the sex of the interviewers (all males), might have been a help to the males' performance. The mean differences in the performances of the groups were briefed in the following graph:

Mean Graph for (Un/familiar & Fe/male groups)
Estimated Marginal Means of Scores



Acquaintanceship

VI. Discussion and Pedagogical Implications

The results for the first hypothesis, which stated that there was not a significant difference in the performance of the familiar group and unfamiliar one, is in line with the findings of Porter (1991a) but it rejects the findings of O'Sul1ivan (2002) which stated that " learners vary In their oral performance when interacting with familiar or unfamiliar speaker". The reason might be the cultural burden of the term familiarity and unfamiliarity, which is different among various L2 learners, so that L2 learners from different cultural backgrounds perform differently in unfamiliar and familiar settings. Moreover, what seems to differentiate L2 learners' performances seriously, is their language proficiency knowledge which has been controlled through homogenizing the candidates' proficiency level in this study.

The findings of the second hypothesis are in line with the findings of most of the other studies (Buckingham, 1997, O'Sul1ivan, 2000 and Porter and Shen 1991) in which the effect of both interviewers and interviewees' gender on the candidates' performances have been emphasized. The males' outperformance in the interviews is justified through the interviewers' gender. In fact, all of the interviewers were males and as Buckingham (1997) states "female interviewees achieve higher scores when interviewed by a woman and the male interviewees perform better when interviewed by a man". Moreover, the cultural background of the interviewees in the research area might be another reason why the females did worse than the males in the unfamiliar setting.

The investigation of the third hypothesis showed no evidence of interaction effect of gender and acquaintanceship on the candidates' performance. Although, the gender was found to be effective on the candidates' performances, it seems that acquaintanceship with the test-givers has moderated its effect, so that their interaction effect was not significant.

The results of the study are believed to have implications for second language pedagogy in two different ways: second language teaching and second language testing.

In the area of language teaching, being familiar with the learners without knowing about their needs might not help the teachers to be successful in teaching oral skills. In fact, teachers’ familiarity with his/her learners has a two-fold role in teaching oral skills. It may facilitate the fluency of their speech, due to decreasing psychological barriers on one hand, and increase inaccuracies in their speech on the other. The more the learners are free to take risks, the more they might make grammatical mistakes. Moreover, it might develop a kind of informal spoken language rather than standard dialect of the language as shown in the linguistic variant analysis.

This study is also relevant to language teaching in other ways. Class group discussions might be more useful, if the teacher tries to use implications of such researches in choosing pair-work oral tasks and see how the learners might act in oral discussion groups when they are familiar (friend) with their partners and when they are unfamiliar with each other. Outperforming the males from the females in their oral performance, which might be due to using male interviewers in both groups, shown that the sex of the teachers might be a factor to oral language teaching in which the learners achieve high scores when paired with the interviewer of the same gender (Buckingham, 1997).

As this study has been done in the area of oral language assessment and the possible effect of both test-givers and candidates’ characteristics on the candidates’ performance, it is believed that it would have more implications in the area of L2 language testing. Considering the difficulty of taking oral proficiency tests, investigating the effects of acquaintanceship of the interviewees with the interviewers on an oral interview test helps both educators and teachers to be more aware of the test-takers' personality effects on the results of an oral interview test and try to diminish the bias effects of such variables in such tests. Moreover, the effect of gender of interviewers and interviewees, although not a clear-cut one, on the oral performance of the candidates should be taken into consideration, especially in our settings in which gender plays a great role in the social behavior of the learners. One way of decreasing the effect of sex on the oral performance of the test-takers in L2 language testing situations is to pair the same gender together in oral interviews as tests of language. This is consistent with recent thinking in the fields of both gender studies and applied linguistics suggesting that gender competes with other aspects of an individuals' social identity in a fluid and dynamic fashion (O'Loughlin , 2002).

Which the results of this study may not be readily generalizable to other contexts, it suggests that a test-candidate's degree of acquaintanceship with his or her interlocutor as well as the sameness of the sex of the interviewer with the interviewees might have, to some extent, predictable effects on the candidates' performance in L2 settings.

Moreover, in important, large-scale oral interviews like oral interviews of entrance examinations of universities in postgraduate levels (M.A & Ph.D), the results of such studies, might help the administrators in how to choose interviewers for such interviews and what to do in order to diminish the effects of personality factors of the interviewers on the results of these oral tests. The last but not the least implication of the study in the area of language testing is the use of the local scale (see appendix A) developed by the researchers for scoring oral proficiency test in the universities, language institutes and schools. It has been validated and it does not need skillful and trained raters. In fact, it has been localized for Iranian teachers and students.

APPENDIX A

How to assess speaking skill in S/F L testing contexts

Oral assessment is one of the controversial issues in S/F L testing contexts. Different ways have been suggested to assess oral language among which are, role-plays, retelling stories, picture description and oral interviews. These are the ways, we trigger the subjects to speak out, but how to score or rank them based on their performance is another problem.

To come to an academic solution, some rubrics (scales) have been presented by various researchers. Rubrics, scoring guidelines, are terms that refer to the guides used to score subjects' performance in a reliable, fair and valid manner. The researchers can use different types of scales to document oral performance of the subjects for example, numerical scales such as those varying from 1 to 10, qualitative one, which assigns words to the various levels of performance like inadequate, adequate and excellent to clarify the subjects' oral performance. Sometimes simple checklists are used to document the presence or absence of a variety of behaviors and then the general performance is scored.

Technically, the rubrics should have the following characteristics, if they are to yield consistent scores:

1. Continuous: the degree of difference between a 5 and 4 levels should be the same as between 1 and 2 levels.
2. Parallel: Similar language should be used to describe each level of the performance.
3. Coherent: The rubric must focus on the same achievement target throughout the whole assessment process. For example if the purpose of the performance assessment is to measure organization in writing or speaking, then each point on the rubric must be related to different digress of organization not factual accuracy or creativity.
4. Highly descriptive: the rubric should describe and clarify each level of performance in order to help teachers and raters recognize the salient features of each level.
5. Validity: what is scored should be what is central to performance not what is easy to see or score. They should reduce the likelihood of biased judgments of subjects' work.
6. Reliability: A reliable performance assessment rubric should enables: a. Several judges to rate a subject's performance on a specific task to assign the same score or, b. Each judge should be able to rate or score the same performance in different times, giving consistent scores.

The rubric, the researcher has developed in this study, is not a task-specific one and it can be used within different oral tasks, given to the subjects. Due to some limitations the rubrics for native speakers of language have like high standards and the need for native language raters, the researcher has tried to develop a scale which is more suitable for S/F L situations and can be used by English language teachers and non-native raters and practitioners. It does not need trained raters and it has been validated through administrating by two skillful raters who scored the oral performance of 100 Iranian English university student using the rubric. In fact an inter-rater reliability was accounted using spearman correlation and the results showed a high positive correlation (+.79) between them. Moreover it is at hand and not as expensive as the original ones.

It is a combination of holistic and analytic scales together. In fact it does not need to conduct holistic and analytic scorings separately and in two different times. It is holistic because it has levels for each component of the skill like accuracy, fluency… which have been described and clarified to the raters. It is supposed to be analytic, because each level, in addition to its description, is given a numerical value (1-5) and the sum of these values (scores) is the score for that component. Finally sum of the numerical values given to the different components of the spoken data, accuracy pronunciation, fluency, comprehensibility and vocabulary, is the subjects' total score based on which H/She is ranked (between 1 to 25).

The scale components and numerical values are given in the following tables (Table 1 to 5):

TABLE ONE (FLUENCY)

Very poor	**poor**	**Adequate**	**Good**	**Very good**
Full of hesitation, stops and a very slow rate of delivering of language **Total Score=1**	Slow rate of delivery of language with errors **Total Score=2**	Acceptable rate of delivery but with frequent errors **Total Score=3**	Reasonable rate of language delivery with relatively infrequent hesitation **Total Score=4**	Acceptable, native-like rate of delivery whit no hesitation **Total Score=5**

TABLE TWO (PRONUNCIATION)

Very poor	**poor**	**Adequate**	**Good**	**Very good**
Barely comprehensible, Extremely Anglicised pronunciation **Total Score=1**	Comprehensi ble but with some difficulty and frequent serious mispronunciation **Total Score=2**	Inconsistent with some notable errors in pronunciation and intonation **Total Score=3**	Mostly correct intonation and pronunciation using different sound patterns Easily **Total Score=4**	Only occasional errors of Pronunciation and intonation **Total Score=5**

TABLE THREE (ACCURACY)

Very poor	**poor**	**Adequate**	**Good**	**Very good**
Very little evidence of grammatical awareness. Very short utterances whit frequent inaccuracies and interfere with communication **Total Score=1**	Frequent serious errors or few serious errors but only uses simple language and rely on short utterances **Total Score=2**	Weak in more difficult areas, some ability to self correct with inaccuracies but do not impede with communication **Total Score=3**	Mostly accurate in more complex language with some mainly minor errors. Syntax nearly always correct with increased length of utterances **Total Score=4**	Highly accurate in more complex language with only a few errors. Use lengthy utterances with complex structure **Total Score=5**

TABLE FOUR (VOCABULARY)

Very poor	**poor**	**Adequate**	**Good**	**Very good**
Very poor and limited. Vocabulary. Frequent incorrect words and borrowing from other languages. Unable to use alternate phrasing. **Total Score=1**	Poor, lexical limitation, frequent repetition of common words and phrases and occasional borrowing of vocabulary. **Total Score=2**	Generally sufficient lexical knowledge and ability to produce synonyms. Repetition of less common vocabulary. **Total Score=3**	Adequate lexis to handle familiar tasks. Varied and interesting vocabulary. Little or no repetition and a good range of specific and general vocabularies. **Total Score=4**	In addition to the features of good performances, demonstrates idiomatic usage of vocabulary with ability to select unexpected words. **Total Score=5**

TABLE FIVE (COMPREHENSIBILITY, ORGANIZATION OF IDEAS)

Very poor	**poor**	**Adequate**	**Good**	**Very good**
Very poor information, inadequate and not organized. Ideas introduced into no apparent order. **Total Score=1**	Information is basic and inconsistently organized. Some progress of organized ideas but with a simplistic approach. **Total Score=2**	Ideas mostly well-worn. Pedestrian and not always linked and some attempt to give opinion or develop an argument. **Total Score=3**	Information is effectively chosen and clearly organized. Clear structure with logical presentation of ideas and development of arguments. **Total Score=4**	Information is thorough and exceptionally organized. Ideas clearly linked and well developed. Interesting insights into the subjects are given. **Total Score=5**

REFERENCES

[1] Backman, L. F. (1990). funder mental Considerations in Language Testing. Oxford University Press, Oxford.

[2] Baker, D. (1989). Language Testing; A Critical Survey and Practical Guide. WClE 7DP, London.

[3] Buckingham, A. (1997). Oral language testing: Do age, status and gender of the interlocutor make a difference? MA dissertation abstract, University of Reading.

[4] Bygatc, M. (1983). Speaking. Oxford University Press, Oxford.

[5] Cholewka, Z. (1997). The influence of the setting and interlocutor familiarity on the professional performance of foreign engineers trained in English as a second language. Spring Vale, Australia.

[6] Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly*, 16 (1), 43-50.

[7] Farhady, H. & Jafarpoor, P. & Birjandy, P. (1995). Testing language skills from theory to practice. SAMT Publications, Tehran.

[8] Hughes, A. (2003). Testing for Language Teachers. Cambridge University Press.UK.

[9] Kunnan, A.J. (1995). Test—taker characteristics and test performance. Cambridge university press, UK.

[10] Madsen, HS. (l983).Techniques in Testing. Oxford University Press, Oxford.

[11] O' Sullivan, B. (2000). exploring gender and oral proficiency interview performance. *System* 28, 373-88

[12] O' Sullivan, B. Porter, D. (1995). Speech style, gender and oral proficiency interview performance. Paper presented at the RELC Conference, Singapore (Net).

[13] O'loughlin, K. (2002). The impact of gender in oral proficiency testing. MA Unpublished thesis, The university of Melbourne

[14] O'SuIlivan, B. (2002). The effect of candidates acquaintanceship on OPI performance. *Language Testing* 19, 277-295.

[15] Porter, D. & Shen Shu Hang. (1991). Sex, status and style in the Interview. *Ahrhus*, pp. 111-128.

[16] Porter, D. (1991b). Affective factors in the assessment of oral interaction: Gender and Status. *Anthology Series 25*, pp. 92-102.

[17] Porter, D. (1991a). Affective factors in language testing. *Language Testing in the 1990s,* London, pp. 32-40.

[18] Richards, J.C. , Platt, J. , Platt H. (1992). Dictionary of Language Teaching and Applied linguistics. Longman Publications, London.

[19] Underhill, N. (1987). Testing Spoken Language. Cambridge University Press, London.

[20] Weir, C. (1993). Understanding and Developing Language Tests. Prentice Hall International LTD Press, New York.

Mohiadin Amjadian is a faculty member at Kurdistan medical university. He got his MA in TEFL at Tarbiat Modares University. He is interested in ESP, language testing and material development.

Saman Ebadi is a PhD candidate of TEFL at Allamaeh Tabatabei University, Tehran, Iran. His areas of interest are sociocultural theory, dynamic assessment, CALL, CMC, language acquisition, and syllabus design. He has presented in different national and international conferences on ELT and published articles in scholarly journals.