

An Evaluation of Final Conversation Test of BA English Translation Students of Oral/Aural Course in Payam-e-Nour University

Fereidoon Vahdany

Department of Applied Linguistics, Payam-e-Nour University, Tehran, Iran

Fatemeh Amin

Department of Applied Linguistics, Payam-e-Nour University, Tehran, Iran

Somaye Shirani

Department of Applied Linguistics, Payam-e-Nour University, Tehran, Iran

Abstract—The purpose of the present study was to investigate the correlation between teacher-made test and Payam-e-Nour (PNU) final examination for the aural course of the translation students together with a detailed analysis of the PNU final written test items in terms of their item facility, choice distribution, item discrimination, and estimating validity and reliability of the total final written test. Sixty English majors from PNU university of Rasht took part in this study. The result of Pearson product-moment correlation showed that there was a strong positive correlation between oral teacher-made test scores and PNU written test scores ($r = 0.709$). Moreover, Cronbach alpha statistic estimated for the test showed an acceptable reliability index of 0.78. The results of item analysis of the PNU written test items revealed that the written test definitely needed revision. The findings also revealed that PNU written test lacks content validity specially in terms of the number of the items, unequal distribution of the items from two books and throughout the units, the example items, representative tasks, representative sample, the importance or weight given to certain elements, and learning level of the testees. The finding of the study implied close consideration of test content and test specification having the objectives of the course in mind in the process of test making by test developers.

Index Terms—English conversation test, BA English translation students, PNU

I. INTRODUCTION

Evaluation plays a significant role in deciding the learners' learning and the teachers' teaching in schools and universities. It is universally accepted as integral part of teaching and learning. Any changes in assessment policies can be used as a powerful lever for reforming schools and universities (Reardon et. al, 1994, cited in Agrawal, 2004). Evaluation as powerful means of improving the quality of education can use examinations and assessment for curricular reform (Torrance, 1995, cited in Agrawal, 2004).

According to some practitioners, oral-aural skills are ignored in ELT (Podder, 2012). Although communicative English was apparently introduced to the educational curriculum of many schools and universities with the idea of making the students proficient in the four basic areas of language use, especially in oral-aural skills of the school leavers are considered below expectation level. Podder (2012) stated that lack of rigorous aural/ oral assessment system and also lack of resources in evaluating students can be the main reasons of the ignorance because the teachers and also teacher educators are not prepared enough to assess these two skills, i.e. they are not well trained. So, two possible speaking and listening assessment strategies were suggested for English teachers to assess their students' speaking skills: "a) continuous during-teaching and informal assessment, b) self and peer assessment strategy" (p. 1). Therefore, an efficient English language curriculum needs to be revised and improved through inclusion of listening and speaking skills assessment.

Proponents of evaluation reform think that such written tasks as multiple-choice tests have a restricting effect on the curriculum and teaching methods. This kind of evaluation encourages teachers to teach students test-taking skills rather than develop metacognitive skills and understanding (Agrawal, 2004). According to Kaye (2008) phonological features, rules of language, paralinguistic devices, communicative functions, and social meaning can be the elements to be evaluated in speaking test in isolation by analyzing various elements or by using tasks which can measure the speaker's communicative competence in general. Moreover, emotional factors such as stress and nerves, and also available resources such as time, examiners, equipment and space must be taken into consideration.

The ability to speak the target language has become a major focus of language teaching and testing. How to assess college students' speaking ability within EFL teaching programs at universities has become more important during past years. However, it remains a difficult task for many TEFL teachers to conduct a reliable and valid speaking test for

English-majored students. It is newly believed that a task-oriented approach can be a valid assessment of the productive communicative ability of test taker, especially for English-majored students at advanced level (Taeduck & Finch, 1998).

Beside authenticity and a relaxed atmosphere, there are other important factors in designing an oral test of communicative effectiveness including full local knowledge; a human approach; a suitable balance; and the ability to adapt and improve the test (Underhill, 1987, cited in Taeduck & Finch, 1998). Underhill also mentioned that the test should be designed as a whole, and based on the learners' need. Savignon (1983, cited in Taeduck & Finch, 1998, p. 4) in confirming this 'human approach' stated, "Language is culture in motion. It is people interacting with people...the most effective programs will be those that involve the whole learner in the experience of language as a network of relations between people, things, and events.

Traditionally, the test of spoken English at advanced level is conducted indirectly and invalidly, containing mechanical greeting, dialogues, sentence reading, storytelling, passage recitation, situational conversations, or a prepared speech based on a given topic. However, such an assessment of students' speaking ability can be criticized for lack of authenticity because of the artificiality of much testing material, and also because of the negative wash back effect in producing well-memorized answers in a prepared speech topic (Gong, n.d). Hughes (1989) pointed out, "We want to set tasks that form a representative sample of the population of oral tasks that we expect candidates to be able to perform. The tasks should elicit behavior which truly represents the candidates' ability and which can be scored validity and reliably" (p. 101).

Speaking is "the process of building and sharing meaning through the use of verbal and non-verbal symbols, in a variety of contexts" (Chaney, 1998, p. 13, cited in Kaye, 2006). Speaking is a crucial part of second language learning and teaching. Despite its importance, for many years, teaching speaking has been undervalued and English language teachers have continued to teach speaking just as a repetition of drills or memorization of dialogues. However, today's world requires that the goal of teaching speaking should improve student's communicative skills, because, only in that way, students can express themselves and learn how to follow the social and cultural rules appropriate in each communicative circumstance (Kayi, 2006).

Many universities throughout Asia including Iran have compulsory English conversation courses. The common schedule in these courses usually includes two hours of class per week plus a midterm and final exam. The format of the exam is different in different universities. Sometimes it is a fixed format (i.e. Must be written, must be oral, a one-on-one interview, etc.), but it can also be more flexible depending on the teacher's knowledge and experience. According to Moodie (2008), there are many ways to administer exams in university conversation classes including:

- Written exams (some with listening component)
- Interviews or one-on-one question period with instructor
- Pair work conversations or pair communicative activities

Each of the above formats has some advantages as well as disadvantages. Written exams are preferred especially in large classes in that they can be administered and controlled with few resources, for all students at once. This is a very efficient use of time for the teacher. What is more, an objective written record of answers is at hand to be compared easily. Also, the instructor can test for many language structures since usually up to two hours' time is available in each exam. However, the most important disadvantage of the written exam for conversation courses is lack of construct validity, "written tests ostensibly do not test for speaking ability" (Moodie, 2008, p. 1). This glaring weakness of the written exams can be removed by one-on-one interview examination. But the main problem of this approach is that time constrain prevents the student and instructor have any kind of normal real- world conversation. Lack of enough time to test both speaking and listening comprehension ability is the major obstacle. Moreover, time limitation puts a lot of pressure on the students to act normally. Besides, performing ideally in front of the instructor who is also a tester would make it more stressful for the students. It is also very difficult for the instructor to both have a conversation with a student and evaluate it simultaneously (Moodie, 2008).

According to Moodie (2008) and also some other expertise in the field, task-based pair work exams seemed the most appropriate way of testing in conversation courses. Utilization of pair work activities in conversation course tests can be a way to improve upon one-on-one testing format. This type of activity "frees up the cognitive resource of instructors in order to pay closer attention to the production of each student than if they were participants themselves. Students have longer time to interact, instructors have longer time to evaluate and comment on each student's performance" (Moodie, 2008, p. 2). Moreover, since communicative language teaching methods are the most common methods of teaching nowadays in which pair work is usually the main part of the class, it would be befitting to include similar activities in the exam. Moodie (2008) stated that "for a conversation course, oral pair work exams are much more relevant than written exams or one-on-one interviews", and "there may also be benefits regarding student's motivation" (p. 3).

English oral/aural courses at EFL universities are of significant importance because the result of the courses can greatly help students to get along well with their major of study and cope with the responsibilities and objectives of the course of study during the semesters. Among the factors affecting EFL learners' oral/aural abilities, assessment procedure or testing methods, particularly of final tests, are highly important. The result of final conversation test at the end of the terms will inform EFL learners and also teachers' real abilities in speaking and listening and will give them a feeling of satisfaction if the test is really valid and reliable.

The purpose of this study is to evaluate the appropriateness of the final conversation test of a group of BA English translation students in Payam-e-Noor university in which the final oral/aural test consists of two major parts: oral/aural teacher-made test and written test of the course made by the central educational department of Payam-e-Noor University. Moreover, this study is also intended to prioritize the qualitative evaluation of the final conversation test along with quantitative investigation of the test including reliability and correlation. Validity aspects of a national test are superior to its reliability and quantitative information. High correlation between PNU teacher made test and written test, and also high reliability of written test do not guarantee appropriateness of a national test without ensuring its validity. The following research questions were proposed in order to address the purpose of the study:

1. Is there any significant relationship between the two sets of scores (oral teacher-made test scores and PNU written test scores) of the final conversation course?
2. What are the item characteristics of the written test like in terms of item facility, item discrimination, and choice distribution?

II. METHODOLOGY

Participants

This study was conducted based on availability sampling of a group of 60 BA English translation students who took their English conversation course in their first semester of Payam-e-Noor University of Rasht. They were Iranian students whose L1 was Persian, predominantly female, ranging from 19-30 years old. Actually, these students' final written test scores were taken from the question bank of Payam-e-Noor University in Rasht in the second semester of 1390-91. All participants, though heterogeneous in terms of English knowledge and level were the first semester freshmen students majoring English translation course in 1391-92.

Instruments

Teacher-made test and written test scores of PNU students majoring English translation were used. Moreover, for the sake of item characteristics and unavailability of written-test answer sheets of PNU students, written test was taken by a smaller sample group in the form of a paper and pencil test. The written test score was out of 14 and the teacher-made test (practical score) was out of 20. PNU written test consisted of 50 questions including listening comprehension, vocabulary, stress and intonation sections. Surprisingly, one grammar question was also included in the test. Regarding teacher-made test (or practical score) there was no special criterion or a fixed test to evaluate students. That is, the teacher used different ways to improve students' conversation ability based on his own experience and knowledge regarding the material provided, that is, two interchange books by Jack. C. Richards.

Procedures

In the first stage of data collection, the written test scores and the teacher-made test scores of conversation course along with the written test paper and its audio sheet were taken from the question bank of PNU in Rasht. In the next stage, the same written test was taken by a smaller sample students to define the item characteristics in the Fall semester of 2013. Because the researcher was not allowed to have the written test answer sheets of the students by principals of the PNU for the security reasons in order to check each item of the test in terms of item facility, item discrimination, choice distribution, and also reliability index of the test.

The test was administered to 34 PNU students in the lab by the researcher as a proctor. The listening section was played by their own teacher. The objective of the test was explained to the participants first. Then they were given 80 minutes to answer 50 questions plus listening, and reading the audio by the teacher. In order to take the test serious the students were asked by the teacher in advance to be prepared to take it in the last session of their semester.

III. RESULTS AND ANALYSIS

Table 1 presents descriptive statistics for the two primary variables of teacher-made test scores and PNU written test scores. Frequency, mean, and standard deviation of two sets of scores are presented.

TABLE 1.
STATISTICS FOR THE TEACHER-MADE AND PNU WRITTEN TEST SCORES

	Teacher- made test scores		PNU written test scores	
	Valid	Missing	Valid	Missing
N	60	0	60	0
Mean	14.8250		12.1287	
Median	15.0000		12.6000	
Mode	18.00		12.88	
Std. Deviation	3.65089		1.48922	
Variance	13.329		2.218	
Range	14.00		6.72	
Minimum	6.00		7.28	
Maximum	20.00		14.00	
Sum	889.50		727.72	

Table 2 confirms the normal distribution of both written test and teacher-made test scores. Before running the statistical test namely Pearson correlation, Skewedness analysis was done to check the normality assumption.

TABLE 2.
ONE-SAMPLE KOLMOGOROV-SMIRNOV TEST

		PNU written test	Teacher test
N		60	60
Normal Parameters ^{a, b}	Mean	12.1287	15.1467
	Std. Deviation	1.48922	3.50643
Most Extreme Differences	Absolute	.151	.125
	Positive	.111	.083
	Negative	-.151	-.125
Kolmogorov-Smirnov Z		1.171	.972
Asymp. Sig. (2-tailed)		.129	.302
a. Test distribution is Normal.			
b. Calculated from data.			

The results of the Skewness analysis, as it is signified in table 2 by running Kolmogorov-Smirnov test, confirmed the normality of the distribution for both PNU written test and teacher-made test. The insignificant results indicated that the probability of the Z statistic is higher than 0.05 for both of the tests (.129 for the PNU written test and .302 for the teacher- made test), meaning that the distributions are normal with two parameters of 12.1287 for the PNU written test and 15.1467 for the teacher- made test.

Table 3 presents the statistics related to the correlation between teacher-made test and PNU written test. A correlation analysis was performed between the two sets of scores by using a Pearson product-moment correlation procedure. The researcher used Cohen's (1988) classification of correlation strength. Cohen (1988) suggested that a correlation is considered small when $r = .10$ to $.29$, medium when $r = .30$ to $.49$, and large when $r = .50$ to 1.00 . As it is shown in Table 3, strong positive correlation is reported.

TABLE 3.
CORRELATION BETWEEN TEACHER-MADE TEST AND PNU WRITTEN TEST

		Practical score	Testing score
Teacher- made test score	Pearson Correlation	1	.709**
	Sig. (2-tailed)		.000
	N	60	60
**. Correlation is significant at the 0.01 level (2-tailed).			

At the .01 level, the correlation of .709 is significant between the two sets of scores (oral teacher-made test scores and PNU written test scores) of the test for the conversation course. It is evident that students who scored high on oral teacher-made test scores received very similar scores on PNU written-test scores, and the opposite is also true. The above findings reject the first research null hypothesis that there is no significant relationship between the two sets of scores (oral teacher- made test scores and PNU written -test scores) of the test for the conversation course.

Table 4 summarizes items analyses of the written test in terms of item facility, item_difficulty, and item discrimination. Item analysis was done to examine student responses to individual test items of PNU written test in order to assess the quality of those items and of the test as a whole. After the PNU written tests were administered and scored, a post hoc analysis was performed in order to evaluate the test's effectiveness. This procedure involved an analysis of the individual items on the test and the reliability analysis of the whole test. Item analysis of the PNU written-test scores including item facility, item difficulty, item discrimination and choice distribution were done to examine the efficiency of the items, the results of which are presented in the following sections.

The results of students' performance in PNU written test was then used to determine the item facility index, difficulty index and discrimination index of each item in the respective test. In this study, the item facility index (IF) refers to the percentage of the total number of correct responses to the test item. It is calculated by the formula $IF = C/N$, where C is the number of correct responses and N is the total number of responses (i.e., correct + incorrect + blank responses). To analyze the items, the difficulty level and discrimination index were judged using Farhady et. Al (2006) and the guidelines suggested by Ebel et. Al (1965), who has given the following four guidelines for the interpretations for ID values:

1. If $ID \geq 0.40$: no item revision necessary;
2. If $.30 \leq ID \leq 0.39$: little to no item revision is needed;
3. If $.20 \leq ID \leq 0.29$: item revision is necessary; and
4. If $ID \leq 0.19$: either the item should be completely revised or eliminated (Ebel et.al, 1965, cited in Courville, 2004, p. 40).

The item facility index ranged between 0.35 (items 13 and 23) and 1 (items 16 and 26), with a higher value indicating that a greater proportion of examinees responded to those items correctly, and they were thus easiest items. In this study Farhady et. al (2006) criterion was used to determine the easiness of the items. Items with facility indexes below 0.37

were considered too difficult, and items with facility indexes beyond 0.63 were considered too easy. The ideal item facility index was considered to be 0.50.

Only items 13 and 23 were too difficult. Other items which all had item facility indices higher than (0.63) were found to be too easy. (Items: 1 to 12; 14 to 22; 24 to 32; 34 to 38; 40 to 45; and 48 to 50) .Only four items had normal and acceptable item facility indices. (Items: 33, 39, 46, and 47). The results of Table 4 also show that the higher the facility index value the lower is the difficulty index. The item discrimination procedure suggested by Farhady et. al(2006) was used in this study. Items which show discrimination value beyond 0.40 can be acceptable and the ideal item discrimination index is unity (Farhady et. al, 2006).

The range of the discrimination index found was -0.05 to 0.58 (see Table 4).

TABLE 4.
ITEM ANALYSIS

		Item difficulty	Item discrimination
N	Valid	50	50
	Missing	0	0
Mean		.2126	.2194
Range		.65	.63
Minimum		.00	-.05
Maximum		.65	.58

Items 19 and 20 discriminated negatively, in other words the most knowledgeable examinees got these items wrong and the least knowledgeable examinees got them right. The negative discrimination indexes for items 19 and 20 might indicate that these items measured something other than what the rest of the test was measuring. Table 5 shows the relationship between item difficulty and item discrimination for each PNU written test item administered to 34 participants. Item Difficulty Index (ID) and Discrimination Index (ID) for Each PNU written test items analyzed for 50 test items

It indicates that there is a relationship between item's difficulty index and its discrimination index. In items which had a very high (or very low) facility indexes(lower than 0.37 or beyond 0.63), the value of the discrimination indexes were much less than those items which had a mid-range item facility items (including items 33, 39, 46, and 47 which all had satisfactorily facility indexes). Some items had low discrimination index, but the item's facility index was high (such as items 3, 16, and 26) , these might be interpreted the items as being too easy for almost the entire set of examinees, and probably for that reason not providing much discrimination between high ability and low ability examinees. Only six items (including items 6, 9, 13, 31, 39, and 46) had acceptable item discrimination. In other words the item discrimination indices for 44 items were lower than 0.40. Items 3, 16, and 26 had the worst item discrimination indices (ID = 0).

The reliability of the PNU written test was checked on 34 subjects. Cronbach's alpha statistic was computed for the 50 test items and a reliability of .78 was obtained which is considerably higher than the minimum required value of 0.70.

TABLE 5.
RELIABILITY STATISTICS OF THE PNU WRITTEN TEST

Cronbach's Alpha Based on Standardized Items	N of Items
.783	50

The guideline adopted from Lord 1952 was used to interpret reliability coefficients for the present study. The measure of reliability used was Cronbach's Alpha. This is the general form of the more commonly reported KR-20 and can be applied to tests composed of items with different numbers of points given for different response alternatives. The table 5 presents the distribution of the PNU written test in terms of the skills or sub skills tested.

TABLE 5.
DISTRIBUTION OF THE QUESTIONS OF PNU WRITTEN TEST IN TERMS OF SKILLS OR SUB-SKILL

Skill Sub-skill	Frequency Questions	Percentage
Listening comprehension	30	60%
Stress & Intonation	3	6%
Vocabulary	16	32%
Grammar	1	2%
Total=50		

According to this table, the present PNU written test consists of four parts of listening comprehension, stress and intonation, vocabulary, and grammar. Most of the questions are from listening comprehension part (60%) and grammar part has the least number of the questions, that is, only one question (2%).

IV. DISCUSSION AND CONCLUSION

The learning situation in the present study (which was distance education) and was different from regular classroom teaching can be considered the next factor. The EFL BA students had to study the whole content of the books

(interchange intro and interchange 1) and participate in both teacher- made and the centralized final exam. The students and the instructor in this study were all aware of the conditions of the present study in distance mode. Orientation programs can also be useful for the staff new to distance education. According to Pritchett (2004), an orientation workshop should include an orientation to the delivery system, analysis of prospective students, instructional design issues, interaction and learning activities, feedback and assessment, hands-on experience, and logistics of the distance learning program at the institution.

Item analysis was used to calculate the discriminatory power, difficulty index, and choice distribution of individual items. The basic purpose of using these methods was to give a numerical value to the relationship between scores for the total test and the score for a single item. This numerical value is the index of the discriminatory effectiveness, easiness, and distribution of the choices for the items. Although there are various similar ways of calculating the discrimination index, the simplified technique of selecting the upper and lower, which has been suggested in Farhady, et al (2006) was used in this study.

The researcher obtained test score distributions and a list of students' scores in percentile rank order. Each item was identified by number and the correct option was indicated. Discrimination indices were computed for the test items which are reported in Table 4 the findings are important in that poor discriminatory items are a valuable sign towards ambiguous wording, grey areas of opinion and perhaps, even wrong keys. Out of 50 items, three items (including items 3, 16, 26) have unacceptable item discrimination indices. This implies that these items are deficient and should be either modified or discarded from the test.

The quality of the PNU written test as a whole was assessed by estimating its internal consistency. Questions in PNU written test were mainly very short and isolated, they were easy to mark and gave reliable results. Thus the given PNU written test produced relatively consistent scores for the same individual on separate items of the test. Reliability is inversely related to the amount of measurement error in test scores. Thus it can be concluded that the test was reliable and had acceptable internal consistency ($r = 0.783$, $N=50$) with a little amount of measurement error. This supports the assumption that the test under analysis is composed of items measuring a single subject area or underlying ability.

Some factors contribute to the positive relationship between the scores obtained in two different administrations in PNU system. One reason for the close relationship between the findings of the two administrations might be due to employing professional PNU instructors who identify the main target of teaching and learning and teach towards them. Here the instructors' expertise which had noticeable positive impact on assessment procedure increased the assessment quality in terms of real reflection of the students' knowledge.

Hence, the second factor can be related to applying suitable textbooks by the instructors for the purpose of teaching and practicing. This factor is justified based on what Motamedi (2009) suggests. He argued that preparing the appropriate content according to the approved topic of every subject by using the advanced scientific findings is one of the main activities of Payam-e-Nour University. These textbooks are regulated by programming instruction to facilitate the learning processes and thus lead to reasonable assessment.

The findings of the study revealed that high reliability and high correlation between PNU teacher – made test and the written test do not guarantee the validity especially content validity of a national test. Training testers and employing well – trained test makers should be taken into consideration by instructional policy makers of PNU system for standardization of the conversation test in line with other national tests, for example, state university and Azad university tests.

The findings have also got some pedagogical implications for test developers to determine the examinees' needs which are going to be analyzed, and clarify the main purpose for testing. This emphasizes the role of using valid and reliable assessment tools. Testers should use assessment tools that are appropriate for the target population. Alternative assessment methods that have less adverse impact can also be considered by test developers.

PNU test designers need to examine different ways to use the test that either reduces or is free of adverse impact. Yet another pedagogical implication of the present study is that when tests are used over a long period of time, it is likely that some items will leak out. The students can easily access the sample final questions on the NET. To help maintain security, test developers can introduce new alternate forms. If alternate forms of the test are available, security can be increased by varying the form used. The result will be higher value for the reliability of the test and preventing measurement error. To improve the reliability, suitable and uniform assessment conditions should be used by test constructors.

REFERENCES

- [1] Agrawal, M. (2004). Curricular reform in schools: The importance of evaluation. *Curriculum Studies*, 36(3), 361-379. Retrieved April 20, 2012 from: <http://www.tandf.co.uk/journals.pdf>.
- [2] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- [3] Courville, T.G. (2004). An empirical comparison of item response theory and classical test theory item/person statistics, Unpublished PhD thesis, Texas A & M University, Texas. Retrieved April 20, 2012 from: <http://repositories.tdl.org/tld/handle/1969.1/1064>.
- [4] Farhady, H., Ja'farpour, A. & Birjandi, P. (2006). *Testing language skills: from theory to practice*. Tehran: SAMT.
- [5] Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.

- [6] Gong, B. (n.d). Considerations of spoken English tests for advanced students. Retrieved April 20, 2012 from www.iaea.info/documents/paper-4pd34afa.pdf.
- [7] Hughes, A. (1989). Testing for language teachers (2nd ed). Cambridge: Cambridge University Press.
- [8] Kaye, P. (2008). Evaluating speaking-part 2. Retrieved April 20, 2012 from: mhmtl:file//www.teachingenglish.org.uk/articles/evaluating-speaking-part2.
- [9] Kayi, H. (2006). Teaching speaking: Activities to promote speaking in a second language. *The Internet TESL Journal*, 13(11). Retrieved April 20, 2012 from: <http://iteslj.org/articles/kaye-teachingspeaking.html>.
- [10] Kent, D. B. (2001). A method for oral testing in university English programs at Korean universities. *The internet TESL Journal*. Retrieved April 20, 2012 from: iteslj.org/techniques/kent-teaching.html.
- [11] Moodie, I. (2008). Using pair work exams for testing in ESL/EFL conversation classes. *The Internet TESL Journal*, 14(8). Retrieved April 20, 2012 from: <http://iteslj.org/techniques/moodie-paiworktesting.html>.
- [12] Podder, R. (2012). Oral-aural skills ignored in ELT. Retrieved April 20, 2012 from: www.thenewantion.bd.com/newsdetails.aspx?newsid=32870.
- [13] Ross, J. (2006). ESL listening comprehension: Practical guidelines for teachers. *The Internet TESL Journal*, 12(2). Retrieved April 20, 2012 from: <http://iteslj.org/techniques/rosslisteningcomprehension.html>.
- [14] Taeduck, H. & Finch, A. E. (1998). Oral testing and self-assessment-the way forward? Retrieved April 20, 2012 from www.3telus.net/linguisticsissues/oraltesting.html.
- [15] Tsang, W. K. & Wong, M. (2002). Conversational English: An interactive, collaborative, and reflective approach. In J. C. Richards & W. A. Renandya (eds.), *Methodology in language teaching*. Cambridge: Cambridge University Press.

Fereidoon Vahdany was born in Rasht, Iran in 1965. He received his PhD. in TEFL from Tehran University. Currently, he is assistant professor at Payame Noor University in Iran. His research interests include testing, teacher education, CALL and MALL.

Fatemeh Amin was born in Tonekabon, Iran in 1981. She received her M.A degree in Teaching English Language from Payam-e- Nour University of Rasht, Iran in 2013. She has been teaching English in English Language Institutes for more than ten years in Tonekabon and Rasht, Iran.



Somaye Shirani was born in Esfahan/Iran in 1989. She did an MA in teaching English from Gheshm International PNU University (Payam-e-Nour) in 2015.

She is currently an English teacher in English academy. She has taught English for about 7 years and she passed so many different TTC courses.

She has lived in Karaj for 8 month. She did PHD pre university exam and she is waiting for the interview.