



Role of Alternative Assessment Techniques In Improvement EFL learners' Speaking Skill (Iranian EFL Setting)

Abstract

Author

Kaveh Jalilzadeh

Islamic Azad University, Science & Research Branch, Tehran, Iran
Email: Kaveh_j2004@yahoo.com

Adel Dastgoshadeh

English Language Department, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran
Email: Englishadel@yahoo.com

Key words:

Assessment Techniques, Testing, Speaking skill

Research into language assessment is central to any agenda that seeks to address the problems of language teaching and learning. A great deal of research in language testing has developed the technical aspects of measuring language ability for the purposes of informing decisions about individuals, as well as evaluating language programs (Bachman, 2000). The aim of the present study was to investigate the relationship between employing alternative assessment techniques and Iranian EFL learners' speaking ability. To achieve this goal, 60 upper intermediate Iranian EFL learners were selected to participate in this study. Throughout the whole semester, the techniques of self-assessment combined with peer-assessment and teacher-assessment were put into practice in the experimental group. It was in the form of a checklist with five items. All participants received a pre-test containing oral interview and written questions. The written test consisted of 80 items, based on learners' previous book *New Interchange 2*, as well as a post-test, which was in the form of oral interview. An F-test was employed to compare the variances of the two groups in order to check the homogeneity of the two groups, and also a t-test was used to check the homogeneity of the two groups. Inter-rater reliability was calculated in order to check the ratings of the two raters in the oral interview both in pre-test and post-test. Repeated measures of ANOVA were also used to check the differences in the performance of participants within the experimental group on the 5 self-assessments. Oral Interview was used as a pre-test and also post-test for assessing the learners' speaking ability.

Introduction

Language teachers often have the responsibility to select and develop language tests for their classrooms and programs. However, deciding which testing alternatives are most appropriate for a task of particular language education context can be a daunting task, especially given the increasing variety of available instruments, procedures, and practices for language testing. Such alternatives include not only test types with long tradition of use such as multiple choice, matching, true-false, fill in the blank, cloze test and dictation procedures, essay exams and oral interviews, but also innovative assessment rather than procedures including portfolios, self-assessment, peer-assessment, conferencing, dairies and learning logs, and teacher check lists and observations.

In recent years, many areas of instruction and evaluation have undergone reforms in response to changing theories of learning. One such a reform is that the era of testing has changed into an era of assessment (Birenbaum, 1996). Alternative assessment, authentic assessment, and classroom assessment are popular topics which are concerned with employing assessment procedures to raise the standards of assessment and learning. The assumption is then that the objective of teaching spoken language is the development of the ability to interact successfully, and that involves comprehension as well as production.

The field of EFL/ESL has moved from structural teaching approaches to communicative, humanistic and learner centered approaches. These new approaches in

teaching recognize that affective considerations are of vital importance for the acquisition of foreign/second language and suggest methods and techniques that create an anxiety reduced environment for learners (Stevick 1990, Krashen, 1982).

Due to the fact that language teaching and testing are closely related, language testing should enjoy such a shift. According to him, there is a need to shift from strictly summative testing tools and procedures to a more humanistic approach using informal assessment techniques that stress formative evaluation. (Cited in Shaaban, 2001).

Brown et al. (1997) maintain that the students' involvement in the process of assessment has been proved to be pivotal to effective learning. Such informal assessment techniques also involve students in the process of assessment, which consequently improves learning. Self-assessment combined with peer-assessment is characterized by involving students in the process of assessment. And they got three different kinds of feedbacks, such as self, peer, and teacher's feedbacks which were more fruitful for learners. The assumption is made that the aim of teaching spoken language is to develop the ability to interact successfully in that language, and that this involves comprehension as well as production. Speaking also plays a vital role in learning to use language, and also speaking requires students to activate their linguistic knowledge along with their world knowledge in a way that is different from other three skills. Also speaking is one of the most important and fundamental of all four language skills in learning.

Nowadays alternative ways of thinking about learning and assessing learning are needed. A new assessment initiative in foreign and second language study should acknowledge the effect of context on performance and provide the most appropriate contexts in which to assess competence, including ones that involve the individual in making self-assessment. To this end, this study tried to explore whether employment of alternative assessment techniques affect learners' achievement in speaking ability. To be empirically testable, the question was stated into a Null hypothesis form which is read as:

Employing alternative assessment techniques does not improve Iranian EFL learners' speaking ability.

Review of the Literature

In the literature, a distinction has been made between testing and assessment. "Testing is defined as the use

of tests or the study of the theory and practice of their use, development, evaluation, etc. Testing has traditionally provided a measure of growth or achievement by which the success of students' learning has been evaluated. Additionally, testing provides significant information about student morale and anxiety levels, an opportunity for a special kind of intensive study referred to as reviewing for a test and diagnostic tips that come to the teacher as feedback" (Mousavi, 1999: 404). However, Lambert & Lines (2000) defined language assessment as a fact of life for teachers, part of what teachers do; it is an organic part of teaching and learning; and using assessment evidence is part of the planning process.

Performance assessments require learners to use prior knowledge and recent learning to accomplish tasks that demonstrate what they know and can do. There is a direct link between instruction and assessment. A variety of performance assessments provides a more complete picture of a learner's abilities than can be gathered from performance on a pencil-and-paper standardized test. Language testing is influenced by language teaching, and by the emergence of alternative teaching methodologies, alternative methods of testing also were introduced.

Assessment may be defined as "any method used to better understand the current knowledge that a student possesses." This implies that assessment can be as simple as a teacher's subjective judgment based on a single observation of student performance, or as complex as a five-hour standardized test. The idea of current knowledge implies that what a student knows is always changing and that we can make judgments about student achievement through comparisons over a period of time. Assessment may affect decisions about grades, advancement, placement, instructional needs, and curriculum." (Dietel, Herman, & Knuth, 1991: 1).

Brown & Hudson (1998) described three basic assessment types: (a) selected-response (including true-false, matching, and multiple-choice assessments); (b) constructed-response (including fill-in, short-answer, and performance assessments); and (c) personal-response (including at least conference, portfolio, and self- and peer-assessments).

Rationale for Alternative Assessment

Over the past years, there has been a great interest in using alternative assessment techniques. The interest and support has been justified in the literature by different scholars. Cunnigham (1998) for example,

considers some reasons for the remarkable support for alternative assessment:

1. concern about the negative impact of the use of standardized tests in minimum competency testing;
2. dissatisfaction with existing psychometric models; and
3. The belief that the primary purpose of public schools is the promotion of social justice (p.124).

He, then, asserts that the use of alternative assessment is based on assumptions about how students learn; how best to teach them; and the role of assessment, that are quite different from conventional assumptions in these areas. He continues that alternative assessment can be viewed as rejection of conventional principles of measurement, educational testing, and instruction. Conventional methods are indirect and artificial and teachers face many difficulties trying to prepare students for such tests (p.128).

Moreover it should be mentioned that most alternative techniques emphasize formative assessment and can help decrease the level of anxiety caused by concentration on linguistic accuracy and since they stress communicative fluency, they can increase students' comfort and feeling of success (Shaaban, 2001, p.18).

Aschbacher listed several common characteristics of alternative assessments in Brown & Hudson (1998), stating that they

1. Require problem solving and higher level thinking,
2. Involve tasks that are worthwhile as instructional activities,
3. Use real-world contexts or simulations,
4. Focus on processes as well as products, and
5. Encourage public disclosure of standards and criteria

Alternative Assessment and Language Skills

Alternative assessment includes a variety of measures that are suited for assessing different language skills. However, no single assessment model is suited for every purpose. The issue is not whether one form of assessment is intrinsically better than another, rather, the nature of some techniques are such that they can lend themselves to some skills more than the others. For example, as the literature on the alternative assessment techniques shows, portfolio assessment is used to assess students' reading, writing, and listening skills; journals to assess writing skill; interviews and role-plays to assess speaking skill and written narratives to assess writing skill only.

However, there are some techniques which lend themselves to all language skills. Of these techniques are presentations, conferences, peer-assessments and self-assessments. For the purpose of this study, the three techniques of conferencing, peer-assessment, and self-assessment were selected to see if they have any effect on students' speaking skill. In the next section, first a brief account will be given of the present approaches to speaking skill. Then, the teaching of speaking skill will be dealt with briefly, especially within the perspective of communicative teaching which a prevalent trend in TEFL is.

Method

Participants

100 freshmen majoring in English translation received a pre-test out of them sixty selected based on their scores to participate to act as the main subjects of the study, who were subsequently and equally grouped as experimental and control group.

Instrumentation

Multiple instruments including a set of achievement tests including grammar, vocabulary and reading sections or subtests in the form of pre and post tests, checklist for feedback purpose and interview were utilized for both treatment and data collection purposes. The rating checklist assessed students' ability in five aspects of their speaking ability, including: preparation, order of ideas, pronunciation, fluency, and accuracy. These aspects were supposed to be assessed by students, peers, and the teacher at the same time on a scale of five. The teacher was supposed to provide feedback at the bottom of the rating checklist. Every three other session, the students presented a two-minute unprepared lecture, rated their own speaking ability, and were also rated by their teacher and peers. In the control group, the students just gave a two-minute lecture every three sessions. They never rated themselves and each other, and there was no rating checklist. In terms of the reliability considerations, necessary steps were taken so that the instruments could enjoy maximum reliability indices reported as follows:

- Pre-test: $r = 0.712$,
- Vocabulary subtest: $r = 0.527$

- Grammar subtest: $r=0.717$
Reading subtest: $r=0.549$

Furthermore, an oral interview test following FSI techniques was used for checking the homogeneity of the participants in terms of their speaking ability which tapped the five following aspects; accent, grammar, vocabulary fluency and comprehension.

At FSI, ratings are assigned in several ways. The testing team consists of a near native speaker and a certified language examiner who may be either an experienced language instructor or a linguist thoroughly familiar with the language (Wild 1975:30)". However, for the pre-test and the post-test of the study, students were rated in the oral interview by two raters. The raters interviewed students one by one. The interview lasted just 8 to 10 minutes. At first, interviewers talked to the interviewees for two minutes. As a warm-up, they asked the students to introduce themselves and talk about general issues. Then the interviewers asked them questions according to the functions and grammatical points that they had studied in their textbook. This stage took 5 to 6 minutes. Two raters (the interviewers) rated the students separately. The inter-rater reliability between the two sets of scores was calculated (See table 1 in appendix).

It should be noted that, the atmosphere of the interview was friendly and unstressful. It was fixed in functions but in different forms. A speaking checklist was used as a treatment and as a means of data collection. The checklist had five items. Every item had five levels. The rating of the checklist was on the scale of 5: 1=Poor, 2=Needs Improvement, 3=OK, 4=Good, 5=Excellent. Based on each item, students assessed their speaking, and they were also assessed by their peers and the teacher.

Procedures

At the outset of the semester, an achievement test consisting of 80 grammar, vocabulary, and reading items was administered as the pre-test in order to check the homogeneity of the two groups with respect to their English language proficiency. From among the 100 students who participated in this study, 60 were chosen as the subjects of the study based on their scores. The cut-point for selection was fifty percent. As mentioned before, through analyzing the results of the pre-test, 20 items were discarded and 60 items were kept as a pre-

test. The papers were rescored after eliminating the poor items and the new scores were used in order to check the homogeneity of the participants. Then they had an oral interview at the beginning of the course. The oral interview was based on the functions and grammatical points that they had studied before.

The subjects were asked to assess themselves by the checklist every session. They were also rated by their peers with regard to their speaking ability. The students in the experimental group were asked to present an unprepared lecture which was about two minutes. The peers and the teacher himself assessed them in terms of their lecture. The teacher tried to assure the learners that the aim was not to criticize each other, rather to learn from each other. The checklist consisted of five items, with five points for each. At the end of each session the teacher collected the checklists and commented on the students' performance based on the learners' weaknesses. (An instance of the checklist and the teacher's comments is available in appendix E). On the whole, the participants had five self-assessments, combined with peer-assessment and teacher-assessment. The justification for implementing five self-assessments through the whole semester was that five speaking activities were performed during the whole semester. In the control group, the learners were supposed to present a two minute unprepared lecture, but there was no assessment. Finally, at the end of course, all participants in the two groups took the oral interview according to FSI. The procedure for rating the oral interview was the same as the oral interview used in the pre-test. The experimental group and the control group were compared on the basis of their mean scores on the oral interview.

Design of the Study

This study involved one independent variable and one dependent variable. The independent variable was alternative assessment techniques, including checklist of self-assessment combined with peer-assessment, and teacher-assessment. The dependent variable was the participants' scores on the written test and the oral interview as post-tests. Regarding the nature of the research question and the hypothesis, the most appropriate design was the quasi-experimental design.

Data Analysis Techniques

In order to test the research hypothesis, the following statistical techniques were utilized.

1. To check the homogeneity of the two groups based on their language proficiency, an F-test and an independent t-test were performed at the outset of the experiment.
2. To check the homogeneity of the two groups based on their speaking ability, an F-test, and an independent t-test were performed at the outset of the experiment.
3. To check the degree of relationship between alternative assessment techniques and speaking ability, a Pearson correlation was run on students' rating checklists (self-assessment, peer-assessment, and teacher-assessment) and their scores on the oral interview at the post-test.
4. To check the differences in the performance of the participants between two groups on the oral interview as post-tests, an independent t-test was conducted based on the average scores of each learner by two raters.
5. In order to understand whether self-assessment was constructed valid or not a factor analysis was run.

Results

This study was conducted in order to provide an appropriate response to the following question:

“Does employing alternative assessment techniques affect students’ achievement in speaking ability?”

In order to find an answer to this question, several sets of statistical analyses were performed.

To determine whether the samples met the criterion of equality of variances, an F test was used through the following formula:

$$F = \frac{V(\text{larger})}{V(\text{Smaller})} = \frac{4.066}{3.181} = 1.6$$

The F-test value was 1.6. Since the F-value is lower than 3, it can be concluded that the two groups were homogenous in terms of the variances. In other words, the two groups belonged to the same population.

An independent t-test was run to compare the mean scores of the experimental and control groups on the pre-

test. As displayed in Table 1.2, the t-observed value was 1.344, which was lower than the critical value of t, i.e. 2, at 58 degrees of freedom.

Second, to check the homogeneity of the two groups concerning their speaking ability, they were pre-tested by an oral interview. An independent t-test was run to compare the mean scores of the experimental and control groups on the pre-test. As displayed in Table 1.3, the t-observed value was 1.589, which was lower than the critical value of t, i.e. 2, at 58 degrees of freedom.

To check the degree of relationship between alternative assessment techniques and speaking ability, a Pearson correlation was run on students' rating checklists (self-assessment, peer-assessment, and teacher-assessment) and their scores on the oral interview at the post-test. As displayed in table 1.4, there is a high positive correlation between the alternative assessment techniques (Self-assessment, peer-assessment, and teacher-assessment) and speaking ability.

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

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

To check the differences in the performance of the participants in the two groups on the oral interview as the post-test, an independent t-test was conducted based on the average scores of each learner by two raters. The t-value was 5.668, which exceeded the t-critical, which was 2 (table 1.5). So it was concluded that the difference between the experimental and control groups on the oral interview was significant.

In this set of analyses, a factor analysis through the varimax rotation was run to investigate the construct validity of the rating checklist. The SPSS extracted one factor which accounted for 78.828 percent of the variance. Based on the above-mentioned results, it was concluded that all the items in the rating checklist tapped just one factor which was the speaking ability. So the rating checklist had construct validity.

Conclusions and Discussions

After running a correlation coefficient study, it became clear that alternative assessment highly correlated with the speaking ability, because speaking is a process and needs a process-oriented testing tool. Assessment is

process oriented and one of the best tools for testing students' speaking ability. The results of the t-test revealed a significant difference between the performance of the students in the experimental and control groups. Based on the results, it was inferred that the experimental group performed better than the control group due to employing alternative assessment as treatment. Therefore, the null hypothesis of the study was rejected.

“Employing alternative assessment techniques improve Iranian EFL learners’ speaking ability. In other words, employing alternative assessment techniques, namely, self-assessment and peer-assessment leads to a significantly better performance on speaking tests.”

The findings confirm the previous research on the relationship between alternative assessment techniques and students’ achievements on Iranian EFL adult learners’ speaking ability. These findings have some implications and applications.

Utilizing alternative assessment techniques in the experimental group led to a better performance of the students on the students' oral interview in the post-test. So, it can be concluded that alternative assessment can be used for not only testing students but also promoting their learning. Language learning, as any type of learning, can only occur if the experience of the learners is engaged and capitalized on. Every experience is potentially an opportunity for learning, but it does not necessarily lead to learning if there is no active involvement from within the student. Self-assessment plays an important role in helping the student extract meaning from the new experience and also helps him or her to reach an optimal level of performance.

Actively involving students in their own learning and focusing on how to teach students to become more independent learners is a major educational goal in most developed countries. Language assessment is becoming increasingly more authentic and direct as it involves students in tasks that they would normally be involved in.

Alternative assessment as a means of testing students' knowledge and helping them to be aware of their own learning and monitor themselves is gradually coming to the realm of language teaching. As there are some demerits in alternative assessment, the merits overshadow the disadvantages. The most important

demerit is that it cannot be used in high-stake testing situations, but it is more accurate and less threatening.

Pedagogical Implications

Since students are nowadays test-oriented and just study for the test, alternative assessment could be useful in two ways. The first merit is that it is continuous, so students have to study during the whole term for the course. They learn without trying to learn and learn everything gradually. The second merit is that, as they are doing assessment tasks, they will learn, so they understand that test taking is not as scary as they think; they find out that everything is for learning; thus, they should not be worried about the test. Furthermore, assessment plays a key role as a consciousness-raising task. When we perform an assessment task e.g. self-assessment, students become aware of their learning. They understand what they are doing, and can monitor themselves. Therefore, the findings of the current study can be of use in all educational centers. They have direct or indirect implications for and applications to teaching, learning, test development, syllabus design and material development. Considering the important decisions which are made based upon the tests and their possible influence on students’ fate, there is no doubt that alternative procedures as formative, motivating, and anxiety-reducing methods, should be incorporated into the syllabi and lesson plans of the educational centers. In recent years the focus has shifted from the product of instruction to the process; therefore, alternative assessment can contribute to the development of process-oriented curricula. Hence, course designers, considering the instructional as well as evaluation value of alternative assessments and their positive effects, should plan for their use in instruction.

References

- Birenbaum, M. (1996)** Assessment 2000: Towards a Plurastic Approach to Assessment, in: M.Birenbaum & F. Dochy (Eds). *Alternatives in Assessment of Achievement, Learning, Processes and Prior Knowledge*, pp.3-31 (Boston, Kluwer Academic).
- Brown, D.J., & Hudson, T. (1998)** *The Alternatives in Language Assessment*. TESOL Quarterly, 32, 653-675
- Brown et al. (1997)**. *Assessing Student Learning in Higher Education*. London & New York: Routledge.
- Dietel, R.J., Herman, J.L., & Knuth, R.A. (1991)** *What Does Research Say about Assessment?*

Retrieved February 10, 2005, from
http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Cunningham, G. K. (1998) Assessment in the Classroom. London: Flamer Press

Lambert, L., & Lines, D. (2000) Understanding Assessment: Purposes, Perceptions, Practice. Routledge, Falmer

Mousavi, A. (1999) A Dictionary of Language Testing (2nd Ed). Tehran: Rahnama publication.

Shaaban, K. (2001) Assessment of Young Learners, English Teaching Forum, 39, 4:16-22.

Appendix

Table 1.1. Inter-rater Reliability of the Interview

precb	preca		
.761(**)	1.000	Correlation Coefficient	preca Spearman's rho
.000	.	Sig. (2-tailed)	
120	120	N	
1.000	.761(**)	Correlation Coefficient	Precb
.	.000	Sig. (2-tailed)	
120	120	N	

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

Table 1.2. Independent t-test of the Pre-test

t-test for Equality of Means					Levene's Test for Equality of Variances (F-test)		
Std. Error Difference	Mean Difference	Sig. (2-tailed)	df	t	Sig.	F	
.943	1.267	.184	58	1.344	.131	2.346	Equal variances assumed Control
.943	1.267	.185	54.829	1.344			Equal variances not assumed

Table 1.3 Independent t-test of the Pre-test (oral interview)

t-test for Equality of Means					Levene's Test for Equality of Variances		
Std. Error Difference	Mean Difference	Sig. (2-tailed)	df	t	Sig.	F	
.682	1.083	.118	58	1.589	.420	.660	Equal variances assumed C1
.682	1.083	.118	57.143	1.589			Equal variances not

assumed

Table 1.5. Independent t-test for the post-test oral interview

t-test for Equality of Means										Levene's Test for Equality of Variances	
95% Confidence Interval of the Difference		Std. Error Difference	Mean Difference	Sig. (2-tailed)	df	t	Sig.	F			
Upper	Lower										
4.375	2.092	.570	3.233	.000	58	5.668	.020	5.713	Equal variances assumed	Ex2	
4.378	2.089	.570	3.233	.000	51.725	5.668			Equal variances not assumed		

Table 1.6. Total Variance Explained

Extraction Sums of Squared Loadings			Initial Eigenvalues			Component
Cumulative %	% of Variance	Total	Cumulative %	% of Variance	Total	
78.828	78.828	3.941	78.828	78.828	3.941	1
			86.008	7.179	.359	2
			92.357	6.349	.317	3
			96.988	4.631	.232	4
			100.000	3.012	.151	5

Table 1.7.6 Component Matrix(a)

Component	
1	
.912	Q 1
.904	Q 2
.875	Q 3
.874	Q 4
.874	Q 5

