

---

# An Investigation into the Effects of Time of Administration and Language of Instruction on Persian Language Learners' Listening Ability

---

**KAMAL HEIDARI SOURESHJANI**

*Islamic Azad University, Shahrekord Branch, Iran*

## **Bio Data:**

Kamal Heidari Soureshjani holds M.A in TEFL from Shiraz University and is now a Young Researchers Club Member. He taught English courses and IELTS at different institutes in Shiraz. He has also published papers in different journals including IJLS & TPLS. He is presently the academic member of Azad University, Shahrekord branch.

## **Abstract**

Among sundry factors influencing an individual score "Environmental factors" and "Test rubric" play significant roles. The present article attempted to investigate the influence of "time of administration" and "language of instruction", each relating to the above-mentioned factors respectively on the Iranian EFL students' listening ability. To do so, 90 male and female language learners were selected and an old version of TOEFL test was given to them in order to ensure their proficiency level. Then, a listening comprehension test, derived from the material they were studying, was distributed among them at different times, each time with specific conditions and purposes. Results of the study revealed that students' performance was better when: The test was given in the morning, and with Persian (the students' native language) instructions. Conducting such studies can add valuable knowledge to the existing body of knowledge regarding the preparation of the best conditions for test takers on the exams.

*Keywords:* time of instruction, language of instruction, listening ability

## **Introduction**

The point that the performance of testees on a test is affected by numerous factors is axiomatic. These assorted factors have been widely taken into consideration and

both theoretical and practical studies have been carried out with regard to these factors (for example, Bachman and Palmer 1981a, 1982a; Brutsch 1979; Clifford 1987, 1981; Shohamy 1983b, 1984a). It is also obvious that there are two major factors interacting with each other while preparing a test: one factor is the "trait" and means the knowledge which is to be measured. And the other is the "method" referring to the procedure by which we assess the trait. In order to assess a given trait, many different methods may be used and as a result, each of them affects the trait in a different way which finally affects the performance of test takers and their scores. In fact, it could be concluded that, as Shohamy (1984) rightly stated, a test is considered as a good one if the method has little effect on the trait. To put another way, if students' performance on a test is the result of the trait being measured rather than the testing method, that test is considered to be a good testing tool.

Besides, traditionally it was claimed that it is possible to develop a valid and reliable language performance test with acceptable scoring and administering conditions just by regarding some predetermined and fixed procedures (Jones, 1979, p. 50). However, nowadays it has been ascertained that the claim is an overly simplified statement for there are multitudes of factors (random and meaningful ones) which may influence the performance on a test.

The other point needs to be asserted here is that teachers, within their teaching process, need to obtain information about the achievement of their students to assess them and improve their teaching by applying the results to their teaching. To use language tests for these purposes and to make decisions, the quality of the information upon which the decisions are based must be reliable and relevant. Thus,

scores from language tests need to be generalizable, and decision makers need to be able to make fair decisions. Therefore, in order to achieve all these purposes is at least partly a function of taking into account all factors which may somehow have an impact on the performance of test takers. One of the best criteria to be considered as a framework for considering these factors can be Bachman's test method framework which is touched upon in the next section of the study.

Further, foreign language listening comprehension, as another aspect of the present study, is a complex process and has a crucial role in developing foreign language competence. In the past, listening comprehension in language learning was overlooked but at present some researchers have devoted some time to listening and they believe that it is an important skill in teaching and learning a second or foreign language. For instance, Nunan (1998) believed that listening is the basic skill in language learning. Without listening skill, learners will never learn to communicate effectively. In fact, over 50% of the time that students spend functioning in a foreign language should be devoted to listening (Nunan, 1998).

### **Background to the Study**

With regard to what was mentioned in the previous section, it becomes clear that test method factors have some sorts of bearings on test taking process and therefore, these factors must be taken into consideration by test constructors while preparing a test. Bachman (1990) stated that test method facets may influence not only the test takers' performance, but their abilities may also be affected by these test method factors. In this section of the paper, some general points about the various factors which may affect the students' performance are first mentioned and then the focus

will be on the variables mentioned in the Bachman's model (1990) which may influence the test-takers' scores.

Based on the framework proposed by Bachman (1990) on different categories of test method facet, there are numerous factors which have an impact on testees' performance on a test. Each of these factors involve some subcategories which must be taken into considerations while test construction and test interpretation. Among the scads of variables which affect language test score and has been mentioned in the Bachman's model is the influence of "test format". Whether test constructors use "multiple-choice", "true-false", "open-ended" or other testing formats in their tests, may influence the test takers' performance (See for example: Alderson, 2000; Bachman & Palmer, 1996; Buck, 2001). As In'nami and Koizumi (2009) argued, since none of test formats is perfect in every context, test constructors must first look into the characteristics of each test format and then make the best selection according to which one serves the most appropriate for the purpose of the specific test to be prepared. With regard to the test format factor, most of the studies focused on the two commonly-used forms: Open-ended and Multiple-choice forms. Shohamy (1984), for example, in one of his studies asserted that in second language reading, multiple-choice formats are easier than open-ended formats. However, Elinor (1997) contradicted the Shohamy's claim and argued that these two formats were considered to be of similar difficulty.

Another factor which may have an impact on the performance of test takers is the "Test topic". Jennings et al. (1999) purported that test constructors must ensure that the test takers are neither advantaged nor disadvantaged in terms of the topic of

the test. There may be some topics which are favored by some specific members over others who are going to take the test (test bias). Take for example a reading test that is about fixing a car. On this test boys may perform better than girls not because they studied more but because the issue of fixing a car is a topic that most boys have more interest and background knowledge in than girls.

Another factor mentioned in this framework is “Environmental factors.” That is, the physical characteristics of the place where the test is administered may influence testees’ performance. Bowen (1987), for example, found that although scores on repeated administrations of an integrative grammar test under different conditions correlate with each other, average levels of performance differed significantly across administrations. This cited remark clearly shows the significance of environmental conditions in test performance. As it was mentioned in the preceding section, each of these factors includes some subcategories. As to this factor, “Time of administration” that is the time when a test is given to testees is one subcategory of environmental factors. It has been claimed that whether we give a test in the morning, at noon, or in the afternoon cause testees to perform differently. Bachman and Palmer (1996) in line with this important factor stated that it is of great significance in assessing the degree to which the time of testing influences testees’ ability to perform at their best (Bachman and Palmer, 1996).

“Test rubric” is also influential in test takers performance. By test rubric it means the principles specifying how testees are supposed to proceed on a given test. Language of questions has been a concern to test constructors, and for some time L2 reading researchers have recommended that the recall be written in the reader’s

native language in order to avoid a test of writing instead of reading (Alderson, 2000; Bernhardt, 1991; Lee & Ballman, 1987; Shohamy, 1984; Wolf, 1993). Like environmental factor, this factor also includes some subcategories one of them is the "Language of instruction." To put it plainly, whether the test instructions are presented in target language (English) or native language (Persian in this paper) may have an impact on how testees perform on a test. Vygotsky (1969) for example, argued that there was a relationship between language of instruction and test-takers' performance. Jekayinfa (1987) also in another study examined the role of competence in the language of instruction as a predictor of performance in secondary school history and concluded that there is positive correlation between competence in the language of instruction and academic achievement in history.

The technology used in the administration of the exam may also affect the performance of test-takers. One of the most commonly-used technologies in test taking is computer. The use of computer-based tests for assessing students has a long, established history. The use of these tests is likely influenced by many reported advantages. Goldberg and Pedulla (2002) supported this idea when they stated that moves toward computerized testing stem from the advantages it offers over the traditional paper-and-pencil format. This kind of test administration has some advantages and disadvantages in comparison of paper and pencil tests. Among its disadvantages is the fact that some of test takers may become for different reasons (like lack of familiarity with computer) demotivated on computerized tests.

The size of the classroom is one of the factors that can affect the performance of test takers. One of the predominant researches into the effect of class sizes on

student attitudes and performances was undertaken in the USA. Feldman's (1984) meta-analysis of this literature found 22 studies showing a negative relationship between class size and student ratings, 11 with curvilinear relationships (better ratings for very large classes than medium sized ones), 2 studies reporting no significant differences, and 2 reporting a positive relationship. Feldman concluded that large class size has a significantly negative influence on student ratings and their performance on different tests.

The item stem can also influence the performance on a test. Buck (1990, 1991) and also Sherman (1997) examined the any effects of item stem preview on test taker performance through making a comparison between the mean scores of groups who previewed item stems and those who did not. Interestingly, neither study found any significant effect for item stem preview on test taker performance or item difficulty. Buck (1991) then asserted that the lack of attention to the questions or to a low level of interest in the content of the text was the cause for this negative result. Buck finally concluded that knowing the question may not motivate listening as much as test developers hope. In another study Freedle and Fellbaum (1987) suggested that the advantages of previewing item stems and answer options may be only restricted to more advanced listeners, putting less able listeners at a comparative disadvantage.

Based on this study it can also be understood another factor which may have an impact on the test-takers' performance is affective factors and especially motivation. Educators and researchers generally agree that motivation is an important factor in school performance and academic achievement (e.g. Atkinson &

Feather, 1966; Bishop, 1989; Brophy, 1983). Atkinson (1966, 1980) argues that motivation influences behavior in two primary ways. First, the strength of motivation controls the amount of time an individual devotes to an activity; second, the strength of motivation determines the individual's efficiency in performing that activity.

The present paper attempts to determine the effects of the factors of: "time of administration" and "language of instruction", on Iranian EFL students' listening ability. To put it plainly, this paper aims at ascribing the best time for administration of a test and the suitable language for instructions of a listening test. Consideration of these issues can help in getting aware of the most appropriate conditions of testing and therefore, cause testees to accomplish the given test at their possible best time.

## **Method**

### **Participants**

Altogether 90 male and female students learning English language as a foreign language in two well-known English language institutes in Shahrekord, Iran took part in the study. They aged between 19 and 29 and were taking Upper-intermediate courses. These language learners, who were selected based on their availability, were then assigned to two equivalent groups based on an Oxford placement test.

### **Instruments**

Two instruments were used in the study. The first one was an old version of TOEFL test which was used with the purpose of forming homogeneous groups of

language learners. The test incorporated three sections: vocabulary, Listening, and reading section.

Then in order to collect the required data on the performance of the participants, a second instrument which was a listening text sample derived, rearranged, and organized by the author of the paper from one of the materials they were studying titled as *Select Reading*, was utilized. The sample was a story text consisting of 6 paragraphs, 20 sentences, and 208 words. The text was appropriate to the level of the students so that; they had no much difficulty in listening and responding the questions during their classes. As to the reliability and validity of the instrument, the estimated reliability index for it turned out to be almost 79 using the Cronbach alpha formulation. By the way, for the validity of the instrument, it was scrutinized by some experienced teachers and experts teaching in the same institutes and universities and was confirmed to be suitable for the present study purpose.

### **Data Collection**

The test was administered two times, each time with specific features and conditions. In the first administration the test was distributed among the two groups in the morning. One group of the learners received the Persian-instruction version of the test and the other group received the English-instruction version. In the second administration which was held with a 14-day time interval, the same test was given at night. In this administration, the same group that took the Persian-instruction version of the test in the first administration received the Persian-instruction version and the group that received the English-instruction version of the test in the morning-administration testing procedure was given the English-instruction

version. The main reason for making a 14-day delay was to ensure that the learners don't make use of their remembrance of the first administration in the second testing administration.

### **Data Analysis**

Having administered the test and gathered the required data, statistical package for social sciences (SPSS) version 17 in general and a two-way ANOVA and two independent sample t-test were run to analyze the data. The purpose of running the two-way ANOVA was to get a general view regarding any probable effect of time of administration, language of instruction on the listening performance of the language learners. Then in the case of the existence of any significant effect, two independent t-tests were run to see exactly which subcategories of the variable(s) cause better listening performance on the part of learners.

### **Results and Discussion**

In this section the results of the study as to the probable effect(s) of time of administration and the language of instruction on the listening performance of language learners are touched upon. Table 1 represents the two variables of the study and also the number of participants for each category along with the Mean and SD of the data. The point worthy of mentioning as to the table is that as it was afore-mentioned, the participants were almost homogeneous.

Table 1

#### *Descriptive Statistics of the Participants of the Study*

time	Lg	Mean	Std. Deviation	N
Night	English	4.63	2.31	90

	Total	4.63	2.31	90
Day	Persian	6.70	2.18	90
	Total	6.70	2.18	90
Total	English	4.63	2.31	90
	Persian	6.70	2.18	90
	Total	5.66	2.46	180

Table 2 that presents the main results of two-way ANOVA is of our main interest. As it is conspicuous from the table, there is a significant difference as to the language of instruction ( $p = .02 < .05$ ). Likewise, a significant difference is observed with regard to the time of administration ( $p = .04 < .05$ ). In contrast, the table also reveals that the interactive effect of the two variables is not statistically significant ( $p = .06 > .05$ ).

Table 2

*Test of Between-Subject Effects*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12.0 <sup>a</sup>	1	12.0	2.1	.14
Intercept	4612.8	1	4612.8	818.3	.00
time	.00	0	21.3	21.6	.04
lg	.00	0	35.0	4.1	.02
time * lg	.00	0	18.5	12.1	.06
Error	665.1	118	5.6		

Total	5290.0	240			
Corrected Total	677.2	239			
a. R Squared = .018 (Adjusted R Squared = .009)					

The results of Two-way ANOVA fail to exactly determine whether the performance at night or in the morning at the one hand, and the Persian or English language of interaction cause better performance on the listening test. Therefore, two separate t-tests were run to exactly understand this.

Tables 3 and 4 represent the results of t-test for time of administration. As it is obvious from table 3, the variable of time of administration, as it was proved from two-way ANOVA, causes a significant difference in the performance of the language learners ( $p = .00$ ).

Table 3

*Independent Samples T-Test for Time of Administration*

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	Sig. (2-tailed)		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
				t	df			Lower	Upper
score	Equal variances assumed	.9	.32	6.58	.00	-2.96	.47	-3.90	-2.01
	Equal variances not assumed			6.52	.00	-2.96	.47	-3.91	-2.02

Table 4 shows that the mean of language learners taking the exam in the morning is almost remarkably greater than those taking the same test at night meaning that language learners who took the test in the morning outperformed those who took the same test at night. The point which needs to be mentioned here as to the present finding is that in addition to the language ability of test takers, there are some other factors which may affect the performance of them. Factors like boredom, test wiseness, cognitive features of the test takers, etc. are among these factors. All in all, though this out-of-control factors influence the performance of test takers, their effect is trivial and therefore, it can be claimed that the scores of test takers are, to a great extent, due to their language abilities than other factors.

Table 4

*Group Statistics*

	time	N	Mean	Std. Deviation	Error Mean
score	night	45	5.03	2.10	.38
	day	45	8.00	1.48	.27

As for the second independent variable that is, "language of instruction" as it is clear from table 5, the language of instruction may influence the listening performance of test takers significantly ( $p = .003$ ). This finding is consistent with what was observed from two-way ANOVA.

Table 5

*Independent Samples Test for Language of Instruction*

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower Upper	
score Equal variances assumed	.19	.6	3	58	.00	-1.8	.5	-3.0	-.6
Equal variances not assumed			3	57	.00	-1.8	.5	-3.0	-.6

Now in order to see whether Persian or English instruction can lead to better performance on the listening test, table 6 is a great help. It shows that the mean of those learners who took the test with Persian language was higher than that of those taking the test with English language of instruction meaning that the test takers who took the test with Persian instruction outperformed those taking the same test with English instruction. Like what was mentioned for the previous case, though there may be factors other than the linguistic ability of test takers influencing their performance, the main effect would be the language competence of them. There are some works in the literature that, though not exactly, are consistent with this finding. Vygotsky (1969) for example, argued that there was a relationship between language of instruction and test-takers' performance. Jekayinfa (1987) also in a study examined the role of competence in the language of instruction as a predictor of performance in secondary school history and concluded that there is positive

correlation between competence in the language of instruction and academic achievement in history.

Table 6

*Group Statistics*

lg	N	Mean	Std. Deviation	Error Mean
score English	45	5.0	2.3	.43
Persian	45	6.9	2.1	.39

### Conclusion

The paper was in fact an attempt to investigate the two variables of time of administration and language of instruction on the listening performance of Iranian EFL language learners. The main finding which was gained from the study is that both "time of administration" and "language of instruction" can affect the test takers' performance. The study revealed that when a listening test is given to the test takers in the morning, the scores of students were much better than the time when the same test is given to the same test takers at night. Several reasons may be assigned to this finding. Being in more suitable conditions (mentally and physically) can be a likely reason for better performance in the morning. Psychological studies have repeatedly approved the point that most of people study and perform more adequately in the morning than in the afternoon or at night. The same studies mention that in the morning people's minds are not busy, tired (due to daily activities) and therefore they can function more drastically than any other times. The important point needs to be noted here is that although there may be some who can perform better at night and in the afternoon, it may be as a result of their personal habits, cognitive styles, and many other factors.

Moreover, the study also showed that when a listening test is given to a group of students with the instructions in the test takers' native language (here, Persian language), their scores are by far better than when the same test is given in target language instruction (here, English language). When test instructions are presented in the target language there may be some words and phrases whose meanings are not clear to the test takers, consequently, it may influence their performance. In line with this point, Brindley (1998) rightly assert that the penalty of misunderstanding a question may be just one incorrect item. But the penalty of misunderstanding an instruction can lead to losing many items. Shohamy's study (1984) also argued that depending on the format of the test and language of instruction the students' performance on tests may vary. Bowen (1978) also asserted that the average level of performance of test takers on grammar test in different administrations under different circumstances differ remarkably.

The present study suffers from a set of limitations. The first and foremost is the small number of participants. In this study just 90 language learners taking the upper-intermediate level courses were utilized. However, in order to be able to take the current study findings with a more confidence there is a need to achieve some other studies with more number of participants with various levels of proficiency. The second limitation which can be assigned to the paper is about the way data were gathered. In order to collect the data, the author distributed the instruments in two times: once in the morning with Farsi and English instructions. And the second time at night again with both Farsi and English instructions. In each of the administrations specific learners took the test. That is, no random-selection

procedure was employed to choose the students. Therefore, there was no language learner who may take the same test (for example, one with English instruction) twice. Some may think that if random procedures were utilized the results may be more reliable. However, in order to prove this claim there should be carried out studies taking these procedures and then by comparing their results with the present study results it would be possible to judge as to the effectiveness of the data-gathering methods.

### References

- Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Atkinson, J. W. (1966). Motivational determinants of risk-taking behavior. In J. W. Atkinson & N. T. Feather (Eds.), *A theory of achievement motivation*. New York: John Wiley & Sons, Inc.
- Atkinson, J. W. (1980). Motivational effects in so-called tests of ability and educational achievement. In L. J. Fyans, Jr. (Ed.), *Achievement motivation: Recent trends in theory and research*. New York: Plenum Press.
- Atkinson, J. W., & Feather, N. T. (Eds.). (1966). *A theory of achievement motivation*. New York: John Wiley & Sons, Inc.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, research and classroom perspectives*. Norwood, N.J.: Albex.

- Bishop, J. (1989). *Incentives for learning: Why American high school students compare so poorly to their counterparts overseas*. NY: Cornell University, Center for Advanced Human Resource Studies and New York State School of Industrial and Labor Relations.
- Bowen, J. D. (1978). The effect of environment on proficiency testing. *Work paper in Teaching English as a Second Language XII*. Los Angeles : University of California.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics, 18*, 171-191.
- Brophy, J. (1983). Conceptualizing student motivation. *Educational Psychologist, 18*, 200-215.
- Buck, G. (1990). Testing second language listening comprehension. Unpublished Ph.D dissertation, University of Lancaster.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing, 8*, 67-91.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Elinor, S. H. (1997). *Reading native and foreign language texts and tests: The case of Arabic and Hebrew native speakers reading L1 and English FL texts and tests*. Retrieved from [www.eric.com](http://www.eric.com)
- Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education, 21*, 45-116.
- Freedle, R., & Fellbaum, C. (1987). An exploratory study of the relative difficulty of TOEFL's listening comprehension items. In Freedle, R. & Duran, R. (Eds.),

*Cognitive and linguistic analyses of test performance*. New Jersey: Ablex Publishing Corporation.

Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to the test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, 62(6), 1053-1067.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26, 219-244.

Jekayinfa, A. A. (1987). Competence in the language of instruction as a predictor performance in secondary school history. *Language Testing*, 46, 43-65.

Jennings, M., Fox, J., Barbara, G., & Shohamy, E. (1999). The test takers' choice: Investigation of the effect of topic on language-test performance. *Language Testing*, 16, 426-456.

Jones, R. (1979). Performance testing of second language proficiency. In E. Briere & F. Hinofotis (Eds.), *Concepts in language testing* (pp. 50-57). Washington, DC: TESOL.

Lee, J. F., & Ballman, T. L. (1987). Learners' ability to recall and rate important ideas of an expository text. In Van Patten, B., Dvorck, T. R., & Lee, J. F. (Eds.), *Foreign language learning: A research perspective* (pp. 108-117). Rowley, MA: Newbury House.

Nunan, D. (1998). *Approaches to teaching listening in language classroom*. Proceedings of the 1997 Korea TESOL Conference. Tie-on, Kotesol.

Sherman, J. (1997). The effect of question preview in listening comprehension tests.

*Language Testing*, 14, 185-213.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading

comprehension. *Language Testing*, 1, 147-170.

Vygotsky, L. S. (1969). *Thought and Language*. Leeds E. J. Arnold and Sons Limited.

Wolf, D. (1993). A comparison of assessment tasks used to measure FL reading

comprehension. *Modern Language Journal*, 77, 473-89.