

The Effect of Testing Frequency on Iranian Pre-intermediate EFL Learners' Language Achievement

Aida Momeni, Shaban Barimani

Science and Research branch, Islamic Azad university of Mazandaran, Iran

Abstract

The present study aimed at investigating whether testing frequency had any effect on language achievement of Iranian EFL learners at pre-intermediate level. Since random assignment was not possible, the nonequivalent group, pretest-posttest design was employed to study three classes of pre-intermediate learners as control and experimental groups. The results of the one-way analysis of variance (ANOVA) from the gain scores of participants indicated a significant difference among the three groups. Additionally, the result of Scheffé's post hoc test showed that those participants who received weekly tests outperformed those who received biweekly tests.

Key words: Testing frequency, Language achievement, Washback, EFL learners

1. Introduction

Testing plays an important and central part in teaching and learning. But does it always have an effect on learning? Is just learning due to nature of the test or can it have another reason? Can frequent testing positively influence teaching and learning?

Testing is not an isolated event. Testing, teaching, and learning are related to each other. The separation of testing from teaching and learning is somewhat impossible. The influence of testing on teaching and learning is called washback effect (Bailey, 1996). According to Alderson and Wall (1993), if teachers use tests to get their students to pay more attention to lessons and to prepare more thoroughly, it is positive washback. If teachers fear poor results and the associated guilt which might lead to the desire for their students to achieve high scores in tests, it might be a reason for teaching to the test. Consequently, teachers narrow the curriculum and produce negative washback. (cited in Djurić, 2008).

Learning and language achievement does not happen suddenly, it occurs continuously and gradually. So evaluation needs the judgment of the process of learning by degrees. Repetitive evaluation of subject matter, called "frequent testing", is also one of the ways to improve student's learning experiences. Frequent testing refers to testing within shorter periods than the commonly used two or three midterms and final exam type evaluations. (Basol, G., & Johanson, G. 2009).

1.1. Purpose of the study

The purpose of the study is to determine what the optimal frequency is. In other words, to what extent it is beneficial and to what extent it is harmful. By knowing the optimal number of tests, teachers can administer tests properly to help learners in higher language achievement.

1.2. Research Questions and Null Hypotheses

In this study four research questions were posed and their corresponding null hypotheses were formulated.

1. Does testing frequency have any effect on language achievement of the Iranian EFL pre-intermediate learners?
2. Is there any difference in language achievement of the Iranian EFL pre-intermediate learners taking weekly quizzes and those taking biweekly ones?
3. Is there any difference in language achievement of the Iranian EFL pre-intermediate learners taking biweekly quizzes and those taking no quizzes?
4. Is there any difference in language achievement of the Iranian EFL pre-Intermediate learners taking weekly quizzes and those taking no quizzes?

Regarding the above mentioned questions, the following null hypotheses are drawn:

1. Testing frequency has no significant effect on language achievement of Iranian EFL pre-intermediate learners.
2. There is no difference in language achievement of the Iranian EFL pre-intermediate learners taking weekly quizzes and those taking biweekly quizzes.
3. There is no difference in language achievement of the Iranian EFL pre-intermediate learners taking biweekly quizzes and those given no quizzes.
4. There is no difference in language achievement of the Iranian EFL pre-intermediate learners taking weekly quizzes and those given no quizzes.

2. Background of the study

A review of the literature revealed, however, that majority of testing frequency research has been undertaken with the fields of mathematics (Zraggen, 2009; Shirvani, 2009; Kika et.al, 1992), psychology (Marcell, 2008; Grover et.al, 1989; Fulkerson & Martin, 1981; Keys, 1934; Turney, 1931), chemistry and other fields. Testing frequency in English context has not received as much attention. Moreover, testing frequency was not examined in Iran, so this research examines the effect of frequency of testing in the field of English in one of the institutes in Gorgan.

The acknowledgment of frequency as a causal variable in language learning, though not prominent in the last 40 years of applied linguistics or second language studies, is certainly not

new. It was a key concept in the theories and practices of American structural linguists such as Harris (1955, 1968). A substantial body of research has been conducted on the effects of frequent testing on students. One of the earliest studies was done by Turney (1931) and is a good example of the typical frequency of testing study.

A description of Turney's work will depict how most such studies have been conducted. Many studies have been conducted since Turney's work. Bangert-Drowns et al. (1986) conducted a thorough meta-analytic review on what they considered to be the most sound research on frequent testing. In conducting this metaanalysis, Bangert-Drowns et al. used 16 variables to identify the most methodologically sound studies including testing procedures, experimental design, classroom setting, and publication histories.

A recent study was conducted by Shafiq et.al (2011) who examined the effect of classroom quizzes on academic achievement of the students. Shafiq et.al (2011) found that students who received regular quizzes outperform those who experienced no quizzes.

Shirvani (2009) studied testing frequency in high school mathematics classes for six weeks. According to Shirvani (2009), daily quizzing as an assessment strategy would significantly increase student mathematic achievement and he concluded that students' score assignments for the experimental groups significantly outperformed the control group.

Basol et al. (2009) conducted his study through a meta-analytic review on frequent testing. According to the results of Basol et al. (2009), most of the studies on the effect of frequent testing report a positive effect on academic achievement. Also, the effectiveness did not differ according to the frequency level used in high, medium and low frequency group studies (p.119).

In 2009 Zraggen in his research paper "The Effects of Frequent Testing in Mathematics classroom" examined testing frequency to determine whether students retain information better if they are tested on a weekly basis or on a biweekly basis. Zraggen (2009) found that bi-weekly testing is more beneficial for these students than weekly testing.

3.Method

3.1.Participants

The participants of this study were EFL learners at pre-intermediate level in Shokouh institute in Gorgan. All of the learners were female ranging from 16-20 years old. To homogenize the participants and in order to ensure that the members of the three groups belonged to the same population, the test of homogeneity was administered.

Table 3.1. Test of ANOVA for Homogeneity of Variances

	Sum of Squares	Df	Mean Square	F	Sig.
--	----------------	----	-------------	---	------

Between Groups	6.100	2	3.050	.055	.947
Within Groups	4860.800	87	55.871		
Total	4866.900	89			

As we noticed in Table 3.1, one way ANOVA indicates that the sig value ($=.947$) is greater than the p value ($=0.05$). Thus it confirms that the difference between the control and experimental groups was not statistically significant and the three groups belonged to the same level of proficiency.

3.2. Instruments

As it is the case with most studies, tests are among the most commonly used instruments of data collection. Paper and pencil tests were used in this study. Prior to the beginning of the experiment, a 45-item, achievement test was administered to the subjects. The developed test was piloted, with a group of 40 pre-intermediate students studying at the same institute, to determine its reliability. Being aware of the possible underestimation of the NRT reliability measure, for the sake of convenience, the KR-21 formula was applied.

Table 3.2. Descriptive statistics for the piloted test

No. of Participants	No. of items	Mean	SD	KR-21
40	45	36	6.41	0.82

As shown in Table 3.2 above, the test enjoys a reliability estimate of 0.82. The coefficient estimated of the piloted data is at a good level of significance, due to the fact that reliability coefficient ranges from 0.0 (no reliability) to $=1.0$ (perfect reliability).

3.3. Procedure

A quasi-experimental study was conducted for investigating the effect of testing frequency on the learners' language achievement. Then, after having secured the homogeneity of the participants, the researcher started the treatment. Three groups of students participated in the study. Experimental groups were divided into two sections; one received weekly quizzes while the other received biweekly quizzes. Control group received no quizzes except final exam. All groups received pretest and posttest. Experimental groups received classroom quizzes as the treatment. Nine classroom quizzes were conducted for the experimental 1 and four classroom quizzes for the experimental 2. All of these groups received a pretest to determine their entry

behavior and a posttest to measure their terminal behavior. The test used for these purposes has already been piloted to establish its reliability.

4.Results and Data analysis

To test the research hypotheses, the researcher dealt with comparing testing frequency regarding, a parametric technique for analyzing the quantitative data. In this study, independent variable was testing frequency and dependent variable was language achievement. In this way, the study investigated the effect of testing frequency (i.e. independent variable) on language achievement (i.e. dependent variable) of Iranian EFL learners at the pre-intermediate level through one-way ANOVA. Tables 4.1 and Table 4.2 show the statistical analysis of data as follows:

Table 4.1.Descriptive statistics for gain scores among three groups

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Experimental 1	30	17.4000	7.89849	1.44206	14.4507	20.3493
Experimental 2	28	13.0000	5.36104	1.01314	10.9212	15.0788
Control	32	8.9063	4.89476	.86528	7.1415	10.6710
Total	90	13.0111	7.06709	.74494	11.5309	14.4913

Table 4.2.One-way ANOVA across three groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1117.070	2	558.535	14.601	.000
Within Groups	3327.919	87	38.252		
Total	4444.989	89			

The first research hypothesis was “testing frequency hasno significant effects on learners’ language achievement”.As Table 4.1 displays, the mean score and standard deviation for experimental 1 are 17.4 and 7.89respectively,and forexperimental 2 they are 13 and 5.3610. On the other hand, the mean and standard deviation of the control group are 8.9063 and 4.89476. So the difference in the mean and standard deviation of three groups is quite clear. Table 4.2 indicates the results of one-way ANOVA. Before accomplishing one-way ANOVA, the gain

score (the achieved score calculated from subtraction of pretest from posttest) of each individual for each group was calculated. The sig (=0.000) through the one-way ANOVA application which is smaller than 0.05 indicates a difference among three groups. So the first null hypothesis is rejected because testing frequency has significant effects on students' language achievement.

Table 4.3. Post Hoc comparison of results among the groups

(I)Testing frequency	(J)Testing frequency	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Experimental 1	Experimental 2	4.40000*	1.62518	.030	.3525	8.4475
	Control	8.49375*	1.57176	.000	4.5793	12.4082
Experimental 2	Experimental 1	-4.40000*	1.62518	.030	-8.4475	-.3525
	Control	4.09375*	1.60047	.043	.1078	8.0797
Control	Experimental 1	-8.49375*	1.57176	.000	-12.4082	-4.5793
	Experimental 2	-4.09375*	1.60047	.043	-8.0797	-.1078

*. The mean difference is significant at the 0.05 level.

As indicated in the above table, a post hoc analysis was done through the application of Scheffé's test to see which pairs of means were different. This table compares experimental 1 and experimental 2 showing the sig is equal to 0.030. Since the obtained sig is less than the ideal value of 0.05, thus the second null hypothesis (there was no significant difference in language achievement of the Iranian EFL pre-intermediate learners taking biweekly quizzes and those taking weekly quizzes) is rejected. In other words, the learners of experimental 1 significantly outperformed learners of experimental 2. It also displays the sig value of experimental 1 and control group as equal to .000. So the third null hypothesis is rejected because learners taking biweekly quizzes significantly outperformed those taking no quizzes. Moreover it presents the sig value of experimental 2 and control group which is .043, so the fourth null hypothesis is rejected because the p value is less than the ideal value of 0.05. Therefore there is a significant difference between students taking biweekly quizzes and those taking no quizzes. It means that students of experimental 2 significantly outperformed control group.

Table 4.4. Tests of Normality

Testing frequency		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Language achievement	Experimental 1	.094	30	.200*	.969	30	.506
	Experimental 2	.154	28	.085	.937	28	.094
	Control	.107	32	.200*	.973	32	.583

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Before accomplishing one-way ANOVA, as shown in Table 4.4 test of normality is performed to determine whether the data are normally distributed or not. The above table presents the results from two well-known methods of normality, namely the Kolmogorov-Smirnov method and the Shapiro-Wilk method. As we see in the Shapiro-Wilk method, the sig value for experimental 1 is .506 and for experimental 2 is .094 and for control group is equal to .583. As it is clear, all of the sig values are greater than ideal value of 0.05. Thus, it is concluded that the data are normally distributed.

5. Discussion and Conclusion

5.1. Discussion related to the first null hypothesis

In order to investigate the first null hypothesis, stating that, testing frequency has no significant effect on language achievement of the Iranian EFL pre-intermediate learners, a one-way ANOVA among three groups was performed. As it was evident from results in Tables 4.1 and 4.2, there was significant difference among the means of the three groups. Thus, the first null hypothesis is rejected ($\text{sig} = .000 < 0.05$) and the researcher concluded that testing frequency has effect on language achievement of pre-intermediate students.

Like previous findings of several other researchers (e.g. Basol, et.al., 2009; Dineen, Taylor, & Stephens, 1989; Dustin, 1971; Key, 1934; Kika, et al., 1992; Shafiq, et.al., 2011; Shirvani, 2009; Turney, 1931), which indicates that learners given frequent testing outperformed learners not given infrequent testing, our results also support such advantage.

5.2. Discussion related to the second null hypothesis

To answer the second research question, it was hypothesized that there was no significant difference in language achievement of Iranian EFL pre-intermediate learners taking biweekly quizzes and those taking weekly quizzes.

As Tables 4.1 and 4.2 revealed, weekly given short quizzes learners had higher language achievement than those given no short quizzes. As assessed by Post hoc comparisons in Table 4.3, the probability between experimental 1 and experimental 2 was equal to 0.030, so the second

null hypothesis is rejected because the p value is less than the ideal value of 0.05. Therefore, there is a significant difference between learners taking biweekly quizzes and those taking weekly quizzes.

This finding is similar to that of Dinean et al. (1989) who reported learners taking daily quizzes outscored the learners taking weekly quizzes. The results of the findings of the Kikaet. al (1992) study, showed that learners who were tested weekly outscored their bi-weekly tested counterparts support the results of this study.

The results of this study are also consistent with the findings of Dinean et al. (1989) and Shirvani (2009), in which more frequent testing, is more effective than less frequent testing. Both Dinean et al. (1989) and Shirvani (2009) compared daily quizzes with weekly quizzes while the present study examined weekly quizzes with biweekly quizzes.

Unlike the present study, Zraggen (2009) focused on retention of material and found that biweekly testing is more beneficial for learners than weekly testing.

5.3. Discussion related to the third null hypothesis

The third null hypothesis was as follow:

“There is no difference in language achievement of Iranian EFL pre-intermediate students taking biweekly quizzes and students taking no quizzes.”

The findings related to the third research question were statistically examined. Simply put, according to the results of the study (as shown in Table 4.3) the sig value was $.000 < 0.05$. Thus the third null hypothesis is rejected because learners of the experimental 1 significantly outperformed learners of the control group.

5.4. Discussion related to the fourth null hypothesis

As the forth hypothesis the researcher was interested to know if students taking weekly quizzes and students taking no quizzes were different with regard to language achievement. According to the findings, the data obtained from Post hoc comparisons (Table 4.3) clearly indicate that the sig value of experimental 1 and control group as equal to 0.43. So the fourth null hypothesis is rejected because the p value is less than the ideal value of 0.05. Thus learners taking weekly quizzes (experimental 2) outperformed those taking no quizzes (control group).

5.5. Conclusion

As it has been indicated several times throughout this study, the main goal of the research has been to find out the effect of testing frequency on language achievement of pre-intermediate learners.

The findings of this study lead to two major conclusions. First, overall testing frequency has a positive washback effect on language achievement. There is enough evidence that shows that the use of frequent testing would significantly increase learners' language achievement. In other words, testing frequency has washback effect on learners' language achievement because

frequent and short quizzes help students to pay more attention to lessons, to study more, and to prepare more thoroughly.

A second conclusion is that, administering more short quizzes in comparison to less quizzes leads to higher language achievement. This conclusion stems from our findings revealing weekly testing is more beneficial for pre-intermediate learners than those taking bi-weekly testing. Findings in this research had also shown that administering short and frequent quizzes promote students' achievement and learners tested more frequently had higher achievement than those tested less frequently. Thus, when the number of tests increased learners encouraged to study more and to progress better.

6. Implications

1. The results from this study have several important implications for English language institutes. Since testing frequency has significant effect on language achievement of Iranian EFL learners, syllabus designers and supervisors in institutes ought to take testing frequency into careful considerations in lesson plan in order to promote learners' language achievement. They should inform teachers in workshops and training seminars being held in institutes to consider short and frequent quizzes after each period they taught the lesson. Teachers in institutes could benefit from the results of this research. Most of the teachers devote the class time to teaching and neglect testing and administer just one or two quizzes during term. By knowing the optimal number of testing, teachers could administer short quizzes properly.
2. Doing this research has also benefit for universities. Because learners in universities don't study their courses until the last night before the exam, instructors should divide the course into manageable segments and administer more short quizzes during term because it can help them to get better prepared for the final exam.
3. The results of the study may also have benefit for schools. In school, teachers usually administer one or two exams and students don't study their lesson during term. Therefore, by administering more short quizzes (e.g. weekly quizzes), teachers can motivate their students to study more rather than cramming at the end of the term.

References

- Alderson, J.C., Wall, D. (1993). Does washback exist? *Applied Linguistics*, 4, 115-129.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.

- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1986). Effects of frequent classroom testing (Report No. TM-860-524). San Francisco, CA: Paper presented at the Annual Meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 274 672).
- Başol, G., Johanson, G. (2009). Effectiveness of frequent testing over achievement: a meta analysis study *International Journal of Human Sciences* [Online]. 6:2. Available: <http://www.insanbilimleri.com/en>
- Dineen, P., Taylor, J., & Stephens, L. (1989). The effect of testing frequency upon the achievement of students in high school mathematics courses. *School Science Mathematics*, 89(3), 197-200. Retrieved September 20, 2008, from the ERIC database, No. EJ391203. dissertation, Kansas State University, Manhattan.
- Djuric, M. (2008). Dealing with Situations of Positive and Negative Washback. *Scripta Manent* 4(1), 14-27
- Dustin, D. S. (1971). Some effects of exam frequency. *The Psychological Record*, 21, 409-414. Retrieved September 20, 2008, from the PsycINFO database, No. 1972-09643-001.
- Fulkerson, F. F., & Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructor, and test anxiety. *Teaching of Psychology*, 8(2), 90-93. Retrieved September 20, 2008, from the Professional Development Collection database, No. 6803000.
- Grover, C. A., Becker, A. H., & Davis, S. F. (1989). Chapters and units: Frequent versus infrequent testing revisited. *Teaching of Psychology*, 16(4), 192-194.

Harris, Z. (1955). From phoneme to morpheme. *Language*, 31, 190–222.

Keys, N. (1934). The influence on learning and retention of weekly as opposed to monthly tests. *Journal of Educational Psychology*. 25(6). 427-436. Retrieved September 24, 2008, from the PsycINFO database, No. edu-25-6-427.

Kika, F. M., McLaughlin, T. F., & Dixon, J. (1992). Effects of frequent testing of secondary algebra students. *Journal of Educational Research*, 85(3), 159-162. Retrieved September 20, 2008, from the MasterFILE Premier database, No. 5811751.

Kling, N., Miller, C., & Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business*, 81(2), 67-72. Retrieved September 0, 2008, from the MasterFILE Premier database, No. 19083578.

Marcell, M. (2008). Effectiveness of regular online quizzing in increasing class participation and preparation, *International Journal for the Scholarship of Teaching and Learning*, 2(1). Retrieved September, 2008, from <http://academics.georgiasouthern.edu>.

Shafiq, F., & Siddiquah, A. (2011). Effect of Classroom Quizzes on Graduate Students' Achievement. *Journal of Educational Research* (PP. 76-79).

Shirvani, H. (2009). Examining An Assessment Strategy On High School Mathematics Achievement: Daily Quizzes Vs. Weekly Tests. *American Secondary Education*, 38(1), 34-45.

Turney, A. H. (1931). The effect of frequent short objective tests upon the achievement of college students in educational psychology. *School and Society*, 33(858), 760-762.

Zraggen, F.D. (2009). *The Effects of Frequent Testing in Mathematics Classroom*, M.S. Dissertation, University of Wisconsin-Stout Menomonie, WI.