

The Effect of Standardized Testing on Motivation of Iranian EFL Learners

Jasem Mohammadi

M.A in TEFL

Sajjad Farokhipour (corresponding author)

PhD Candidate in TEFL

Abstract

Standardized tests can improve student motivation. When students are encouraged to set high goals for these tests, and are praised on their accomplishments afterwards, they feel like they are able to do better on tests and projects and assignments. It can make the entire school community better and a more positive environment. More and more schools are giving new forms of standardized tests in new ways based on the technology. They are also thinking of new ways to make students more motivated about taking standardized tests, as a lot of students can get stressed the week they take them. Importantly, students try more to achieve the best score by standardized testing that's why standardized testing can improve the students' motivation. In this study, we decided to assess their achievement in language learning in order to test the effect of standardized test on motivation of Iranian EFL learners because we believe that their achievement in learning can show whether they are motivated to learn or not.

Key Words: Testing standardized testing, motivation.

1. Introduction

One of the reasons that tests are unappealing to some students is that tests fairly reveal what we do and do not know. This feedback can violate the positive feelings we hold about ourselves and our abilities, which are often inappropriately optimistic, especially in the classroom (Hacker, Bol, Horgan, & Rakow, 2000). This violation causes students to rate instructors more poorly (Isley & Singh, 2005) and to generate complicated but unsupported theories about supposed learning styles that their classrooms are failing to support (Pashler, McDaniel, Rohrer, & Bjork, 2009). Well-designed tests can be part of the solution to improved educational outcomes. The literature reviewed here suggests some ways in which traditional standardized tests can be modified to take greater advantages of these qualities. SBAC and PARCC are both using computer-adaptive systems for testing, a quality that enables certain interventions that will boost the ways in which tests serve as learning events. Knowing that tests serve a role in learning, and not merely assessment, might allay some of the major concerns that students (Strauss, 2015), school administrators (Perez & Rado, 2015), and governments (Harris, 2015) have with standardized tests as they are currently implemented.

2. Review of Literature

We believe that standardized and qualified tests can improve the EFL learners. Standardized tests have other benefits in EFL learners. There are several studies about the effect of standardized tests on EFL learners. Testing also increases the effectiveness of the way in which we choose to access and organize the tested information. For example, when studying a list of categorized materials, quizzes increase both the number of categories that are reported on a final test and the number of items from each of those categories (Zaromb & Roediger, 2010). These beneficial effects are probably due to the fact that testing promotes clustering of similar items during the test, a retrieval strategy that is very effective (Mulligan, 2005). Most of the benefits come from the first few tests, indicating that it does not require much compromise in the allocation of class time to administer periodic tests. In addition, students of all abilities appear to benefit from the opportunity to take tests (Pan, Pashler, Potter, & Rickard, 2015). As we will see below, these benefits are not limited to enhanced memory for the tested material.

3. 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 motivation of the Iranian EFL learners?
2. Is there any difference in motivation of the Iranian EFL learners taking weekly quizzes and those taking biweekly ones?
3. Is there any difference in motivation of the Iranian EFL learners taking biweekly quizzes and those taking no quizzes?
4. Is there any difference motivation of the Iranian EFL 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 motivation of Iranian EFL learners.
2. There is no difference in motivation of the Iranian EFL learners taking weekly quizzes and those taking biweekly quizzes.
3. There is no difference in motivation of the Iranian EFL learners taking biweekly quizzes and those given no quizzes.
4. There is no difference in motivation of the Iranian EFL learners taking weekly quizzes and those given no quizzes.

4. Methodology

In order to test the effect of standardized test on motivation of Iranian EFL learners, we decided to assess their achievement in language learning because we believe that their achievement in learning can show whether they are motivated to learn or not.

4-1. Participants

The participants of this study were EFL learners at pre-intermediate level in Shokouh institute in Tehran. 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 4-1. 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.

4-2. 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 4-2. 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 4.2 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).

4-3. 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.

5. 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 5-1 and 5-2 show the statistical analysis of data as follows:

Table 5-1. 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 5-2. 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 has no significant effects on learners’ language achievement”. As Table 4.1 displays, the mean score and standard deviation for experimental 1 are 17.4 and 7.89 respectively, and for experimental 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 5-3. 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 5-4. 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.

6. Discussion and Conclusion

6-1. 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 5-1 and 5-2, 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; Shafiq, et.al., 2011; Shirvani, 2009), which indicates that learners given frequent testing outperformed learners not given infrequent testing, our results also support such advantage.

6-2. 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 5-1 and 5-2 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 Kika et, 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 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.

6-3. 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 $.000 < .05$. Thus the third null hypothesis is rejected because learners of the experimental 1 significantly outperformed learners of the control group.

6-4. Discussion related to the fourth null hypothesis

As the fourth 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).

6-5. 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 wash back 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 wash back 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.

References

- Başol, G., Johanson, G. (2009). Effectiveness of frequent testing over achievement: meta-analysis study *International Journal of Human Sciences* [Online].6:2. Available: <http://www.insanbilimleri.com/en>
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92, 160-170.
- Harris, E. A. (2015). As common core testing is ushered in, parents and students opt out. *The New York Times*. Retrieved from http://www.nytimes.com/2015/03/02/nyregion/as-common-core-testing-is-ushered-in-parents-and-students-opt-out.html?hp&action=click&pgtype=Homepage&module=second-column-region®ion=top-news&WT.nav=top-news&_r=1
- Isley, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36, 29-42.
- Mulligan, N. W. (2005). Total retrieval time and hypermnnesia: Investigating the benefits of multiple recall tests. *Psychological Research*, 69, 272-284.
- Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, 83, 53-61.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. A. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, 9, 105-119.
- Perez, J., & Rado, D. (2015). CPS reverses course, will administer controversial test. *Chicago Tribune*. Retrieved from <http://www.chicagotribune.com/news/chi-cps-common-core-test-20150302-story.html#page=1>.
- Strauss, V. (2015, March 7). A 16-year-old takes the new PARCC exam. Here's her disturbing report. *Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/answer-sheet/wp/2015/03/07/a-16-year-old-takes-the-new-parcc-exam-her-disturbing-report/>
- 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.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38, 995-1008.
- Zraggen, F.D. (2009). The Effects of Frequent Testing in Mathematics Classroom, M.S. Dissertation, University of Wisconsin-Stout Menomonie, WI.