Intl. J. Humanities (2016) Vol. 23 (2): (29-42)

The Durability of the Effect of the Frequent Quizzes on Iranian High School Students' Vocabulary Learning

Reza Nejati[1]

Received:2015/9/27 Accepted: 2016/6/23

Abstract

Frequent quizzes followed by feedback on students' strengths and weaknesses may enhance learning. Hence, the possible effect of weekly quizzes on language learning might be of interest in Iranian EFL context, especially in high schools. In order to understand the importance of weekly short quizzes, 88 students from four classes of high school second graders were selected through available sampling to form the experimental and control groups. The treatment went for twelve weeks during which the experimental group received 10 weekly quizzes while the control group did not have any quizzes. The study was carried out through a pretest- posttest- delayed posttest control- experimental intact groups. The results, analysed through Mixed Models ANOVA, revealed that both within and between group's differences were in favour of the experimental group. It is safe to claim that weekly quizzes improve students' performance on vocabulary and the effect of the treatment is durable.

Keywords: Short quizzes; Feedback; Formative assessment; Vocabulary learning; High school students

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

1 . Assistant Professor Shahid Rajaee Teacher Training University Shabanloo St. Lavizan, Tehran, Iran.reza.nejati@srttu.edu

1. Introduction

Previous research has proved that lexical knowledge plays an important role in EFL learners' language development because vocabulary has the central role in speech and without words one is actually speechless. For example, Taylor (2003) states vocabulary plays a key role in second language learning and academic achievement. Lewis (1993) argued that vocabulary should be at the centre of language teaching since "language consists of grammaticalised lexis, not lexicalized grammar".

Iranian students' vocabulary learning seems to be a matter of serious concern for the students, their parents, teachers and education officials. Teachers and educators put their efforts in finding ways to enhance learning. Tests and or quizzes may fulfil such a function. According to Madsen (1983) testing is an important part of every teaching and learning experience. Similarly, Bachman (1990) considers testing an important issue in foreign language teaching and learning. Farhady, Jafarpur, & Birjandi (1994, p.4) assert "testing is an important part of every language teaching and language learning experience". These facts may highlight the importance of testing as the facilitator of learning which seems to be neglected in our high schools.

Weekly short quizzes developed by teachers are possibly one of the appropriate remedies for weaknesses in learning. It is assumed that both teachers and learners can benefit from these tests because such tests may be helpful not only in providing feedback but also in providing learners with valuable practice and learning opportunities. As Madsen (1983) explains, quizzes can help students in their studies by allowing them to find out the areas of emphasis in the course. Quizzes also can help teachers to recognize students' areas of weakness and improve them.

Evaluation of students' performance in our high schools has probably not received due attention at least for two reasons; first, developing well-structured tests is what the teachers consider a difficult task to do, second, most teachers find test development and its administration a very time-consuming activity. If the results of this study support the effectiveness of using short quizzes in classrooms then it would be pedagogically reasonable to persuade our teachers to use classroom quizzes. This also draws attention to the role of quizzes in encouraging students to organize their knowledge, to assimilate large chunks of materials, and to learn from repeating the materials (e.g., Harlen, & Crick, 2003; Wilkerson, Stevens, & Krasne, 2009).

Teachers are presumably responsible for providing their students with appropriate learning opportunities and help them develop their vocabulary. Therefore, the objective of the present study is to examine the strength of the effect of weekly short quizzes on EFL students' vocabulary learning.

2. Literature Review

According to Cheng and Watanabe (2004) we all live in a testing world and our education system is packed with various high stake testing. A lot of EFL teachers have been concerned with assessing and evaluating student's progress during a course as well as their language achievement at the end of course. Heaton (1990) stated that tests are normally constructed as a device to reinforce learning, to motivate students, and to assess their performance in language tasks. Davies (1990) believes that language testing is

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

central to language teaching. Tests can provide feedback on teaching and be used to motivate students to study, to review the materials which have been covered in the classroom, and to evaluate learners' degree of learning of the materials (Bachman, 1990, Farhady, et al, 1994).

Andrade and Cizek (2010) maintain that assessment is the bridge between learning and teaching. Tomlinson (2005) holds that the main purpose of language testing is to provide opportunities for learning, both for the learners and for administrators of tests. Azorlosa and Renner (2006) indicated either announced or unannounced quizzes served pedagogical objectives.

Frequent testing refers to testing within shorter periods than commonly used two or three midterms and final exam type evaluations (Basol & Johanson, 2009). There exist different definitions of frequent testing; for example, Dineen, Taylor, and Stephens (1989) defined frequent testing as a type of examination which is carried out daily. McKenzie (1972) defined it as a weekly based activity, and finally, Kling, Miller, and Reardon (2005) defined it on a monthly basis.

Classroom quizzes make students better aware of their course objectives and areas of emphasis, as well as their weaknesses and language learning needs, thus enable them to make up for such weaknesses (Cohen, 1980; Madson, 1983; Heaton, 1990). Previous studies have shown that quizzes increase class attendance, students' reading time and confidence (Azorlosa & Renner, 2006; Clump, Bauer, & Whitleather, 2003; Marchant, 2002; Ruscio, 2001; Sporer, 2001; Wilder, Flood & Stromsnes, 2001). Frequent testing is beneficial because it can provide the school, teachers, parents, and students with useful feedback on student performance through class room discussions after quizzes (Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C., 1991, Stanlee & Popham, 1960).

Moreover, frequent quizzes provide opportunities for the learners to recall the material for longer period of time and also make them ready for final exams (Johnsom & Kiviniemi, 2009). Since students have to study harder to get good marks on the short quizzes, it creates extrinsic motivation for the students (Dustin, 1971). This has a double effect in that students prepare more for the quizzes since the test itself is a good source of motivation (Zarei, 2010).

Short quizzes cover small amounts of materials and as a result materials are processed more deeply (Stanlee & Popham, 1960). Moreover, Selakovich (1962) believes that frequent testing provides opportunity for classroom discussion on the materials covered in the class. Roediger and Karpicke (2006) investigated the effect of frequent tests on some under graduate students' retention of materials. The results of the study showed that those students who were tested frequently during the course remembered information better than those who were not given tests frequently. Faez (1999) investigated the wash back effect of frequent testing on students' reading comprehension ability. She found that quizzes had positive wash back effect on students' learning and teachers' instruction. These facts provide evidence for the effectiveness of short quizzes on language learning.

There seems to be little agreement among scholars on the issue of using frequent quizzes in classrooms. Some scholars argue that classroom testing would increase instructional effectiveness and would encourage students to study harder and review the covered materials more

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

(Farhady, et al, 1994; Narloch, Garbin & Turnage, 2006; Madsen, 1983; Padilla-Walker, 2006). They claim that these additional testing devices provide opportunities for teachers to correct students' errors, to reward good performance, to emphasize on the areas of weakness, to discuss the materials covered, and to give students a good indication of what they are expected to learn. However, some other scholars note that the process of quiz construction, delivery, and scoring cause time away from instruction and as a result it is not a good idea to use them for instructional purposes (Haberyan 2003; Marshall, 2007). Some educators maintain that giving tests might become tedious for students and reduce learners' enthusiasm for learning (Bangert-Drowns, Kulik, & Kulik, 1991). They stated that quizzes have no positive effect on learners' performance on final exam (Azorlosa & Renner, 2006; Beaulieu & Utecht, 1987; Lumsden, 1976).

As the review of literature shows frequent quizzes have had positive effects on students' achievement. However, to our best knowledge, little research has been done on the 'strength' of the effectiveness of frequent tests in Iranian high schools. Therefore, it is worthwhile to study the strength of the effect of the frequent quizzes on Iranian students' vocabulary learning. This study intends to address the following research question.

Do frequent quizzes have durable significant effects on high school students' vocabulary learning?

3. Method

3.1 Participants

The participants in this study were 88 second grade high school students in Qazvin. These participants were to begin the first semester of the national school year and they were majoring in biology and mathematics. Participants' age range was between 15 to16. All the participants shared the same L1 and cultural background. They were selected from students of two different schools and were divided into two independent groups. Each group consisted of two classes, including one male and one female class to provide a sufficient number of subjects for the study. Two classes with a total of 42 students served as the experimental group, and two classes with a total of 46 students served as the control group. All classes received the same instruction except for classroom testing. The treatment was given in students' classroom during the regular 90 minutes English language teaching period for 12 sessions.

3.2 Instrumentation

The following is the list of materials and instruments used in this study.

Proficiency Test

Oxford Placement Test (Allan, 2000), a standardized and validated test, was administered in order to find out whether participants were homogeneous or not. The test had forty multiple-choice items of Language functions, grammar, vocabulary and cloze test. In order to check the reliability of this test, KR-21 formula was used. It was found that the KR-21 reliability index of the Oxford Quick Placement Test in the Iranian context was .78 which was an acceptable index of reliability

Researcher- made vocabulary test

Since this study aimed at investigating the effects of frequent short quizzes on students' vocabulary learning, a vocabulary achievement test had to be developed. This test was first administered as a pre-test in order to examine students' knowledge of

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

vocabulary before the treatment. Also it was used as the post-test at the end of the course to check the effectiveness of the treatment i.e., how much participants' vocabulary learning improved during the course. This test consisted of forty multiple-choice type items. The reliability of the test was estimated through KR-21 formula and reliability index turned out to be .81.

Short Quizzes

A series of short quizzes were carefully prepared for the purpose of the study. In the process of the preparation of quizzes certain important points were taken into account:

a) Each quiz contained 10 multiple-choice items.
b) 15 minutes at the beginning of each class session was allocated to each quiz.
c) The difficulty level of the words used in the quizzes was the same as that of vocabulary used in their text books.

Students' Text Book

The textbook of junior students of Iranian high schools (English book 3) written by Birjandi, Nourouzi, and Mahmoudi (2013, 1392 Hejeria) was used as the instructional material. Choosing participants' own textbook could possibly have some advantages over other sources as:

a) Students paid more attention to their own textbook rather than some other materials outside their school syllabus.
b) The difficulty level of the lexical items in the textbook is geared to the students' level.
c) Not using students' own textbook was tantamount to holding extracurricular classes, which was unauthorized.
d)

3.3 Procedure

On the first session of the semester, the proficiency test was administered to both groups. On the second session they received the researcher- made vocabulary test. Each correct answer received one point and there was no penalty for incorrect responses. Since repeated testing was to serve as the independent variable, the experimental group received a quiz every session for 12 weeks while the participants in the control group did not receive any quiz during the course. At the beginning of the next session students were provided with feedback on the previous quizzes. One week after the end of the treatment the researcher- made vocabulary test was administered to all four classes as the post-test to compare the experimental and control groups' performance. Finally, three weeks later the participants were given the same test as a delayed post-test in order to check the stability of their learning.

4. Results

In this section, first, the results are detailed and then the discussion of the findings is presented. The research question was:

Do frequent quizzes have durable significant effects on high school students' vocabulary learning?

For the purpose of this study, a series of short quizzes were administered to the students in the experimental group. After the treatment two tests, namely posttest and delayed posttests were given to both control and experimental groups. In order to answer the research question above the data were analysed through the mixed models repeated measures ANOVA.

Before proceeding with ANOVA report, a brief description of the results is presented. The descriptive data is presented in Table 1 below. The mean of both groups in the pretest is very close to each other. However, in other test conditions, the experimental group stands above the control group.

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

Table 1 *Descriptive Statistics for Pretest, Posttest and Delayed Posttest*

Tests	Group	Mean	Std. Deviation	N
Pretest	Control	21.0217	1.65313	46
	experimental	24.9048	4.83286	42
	Total	22.8750	4.02810	88
Posttest	Control	23.2826	2.49143	46
	experimental	34.0238	3.92918	42
	Total	28.4091	6.29292	88
Delayed Posttest	Control	22.9348	2.55953	46
	experimental	33.2619	4.44511	42
	Total	27.8636	6.29392	88

The standard deviations of means in all test conditions including the pretest, posttest and delayed posttest are somewhat close to each other indicating the groups are relatively homogeneous.

Next, the assumption of sphericity was checked. Sphericity refers to the equality of variances of the differences between treatment levels. As Table 2 suggests the assumption of sphericity is violated. In such cases, the researcher has different alternatives such as MANOVA, adjustment of the significance tests of Greenhouse-Geisser or Huynh-Feldt or multilevel modeling (Tabachnik & Fidell, 2007). Moreover, readers may agree that students in a given class would be more similar to each other than to students in another class. In some sense, students exposed to treatment A may be more similar to each other and those exposed to treatment B may be more similar to each other. The behavior of students in class A, for example, will be similar and different from students in class B. The classroom is seen as a contextual variable. In other words, students are clustered within classes. Additionally, the data gathered for the purpose of this study is both of within and between people type. As such, the person acts as a context within which knowledge and or ability is assessed. It follows that the measurements done on such classes are dependent. Hence, the assumption of independence of measurement behind ANOVA is not met (Field, 2003). Following Tabachnik & Fidell, and Field, multilevel model of repeated measures ANOVA is used here. It should be added that when the assumption of equal variances does not hold, SPSS produces Welch correction as in tables 4,5 and 6 below.

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

Table 2 Mauchly's Test of Sphericity for the Tests

		Within Subjects Effect
		Time
Mauchly's W		.498
Approx. Chi-Square		59.244
Df		2
Sig.		**.000**
Epsilon	Greenhouse-Geisser	.666
	Huynh-Feldt	.680
	Lower-bound	.500

As mentioned in the preceding paragraph students exposed to a treatment may be more similar to each other and those exposed to another treatment may be more similar to each other. Sensibly, the measurements are not independent. Hence, the researchers cannot use Repeated Measures ANOVA; instead, they are advised to run Mixed Models ANOVA (MMA). Through intra-class correlation, MMA helps researchers overcome the problem of non-independence of measurements. The intra-class denotes the proportion of the total variability in the dependent variable that can be attributed to the treatments.

Table 3 Intra-class Correlation Coefficient for Posttest and Delayed Posttest

		Single Measures	Average Measures
Intra-class Correlation		.938	**.968**
95% Confidence Interval	Lower Bound	.910	.953
	Upper Bound	.957	.978
F Test with True Value 0	Value	31.220	31.220
	df1	105	105
	df2	105	105
	Sig	.000	**.000**

As Table 3 above suggests, the intra-class coefficient is significant. It is safe to note that variability in the outcome within classes is minimized. It follows that the students in each group have behaved similarly.

Next, the overall fit of a multilevel model is tested. To do so, multilevel models are compared in terms of -2 log-likelihood followed by a chi-square likelihood ratio test. First, the basic model, fixed effects (intercept, group (treatment), time (covariate) and interaction between group and time) were examined. The index of -2 log-likelihood turned out to be 1455.549 with 5 parameters. Then random factors were added to the equation to assess the slope of the 'time' and the treatment. There were 12 parameters including fixed effects (5 parameters) and random effects (7 parameters).The index of -2 log-likelihood turned out to be 1425.98. Therefore:

X^2 change= 1455.549-1425.98 =34.569

Df change= 12-5 =7

Basically, the smaller the value of the log-likelihood of the random effects model,

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

the better. In the present case the value of likelihood dropped from 1455.549 to 1425.98. If we look at the critical values for the chi-square statistic with 7 degrees of freedom, they are 14.07 ($p < .05$) and 18.48 ($p < .01$); therefore, this change is highly significant. All in all, the model fits the sample of the study. In other words, the treatment better justifies the results and the findings of this study can be generalized to the population.

As Table 4 indicates the interaction between the treatment and the covariate (group* time) is statistically significant: *F (1, 90.75) = 33.547, p = .000.* In such cases, researchers are advised to set a more conservative alpha level ($\alpha = .001$) for rejecting the null hypothesis. The difference between groups is statistically significant: *F (1, 86.55) = 47.262, p = .000* .The difference between posttest and delayed posttest (time) is significant: *F (1, 90.75) = 85.208, p = .000.* It can be maintained that the linear trend is significant. That is to say, students in both groups were affected by time.

Table 4 Type III Tests of Fixed Effects for Linear Trend

Source	Numerator df	Denominator df*	F	Sig.
Intercept	1	86.555	4196.030	.000
Group	1	86.555	47.262	.000
Time	1	90.751	85.208	.000
group * time	1	90.751	33.547	.000

*Decimal numbers in denominators df signifies Welch correction

Next, the quadratic trend is examined. To see whether this quadratic trend has improved the model we need to compare the −2LL for this new model, to the value when only the linear polynomial was included. The value of −2LL for the linear model is 1455.549 and that of the quadratic model is 1407.531. We can compute the change in −2LL :

X^2 change= 1455.549- 1407.531= 48.018
Df change= 12-11=1

If we look at the critical values for the chi-square statistic with 1 degree1 of freedom, they are 3.84 ($p < .05$) and 6.63 ($p < .01$); therefore, this change is highly significant.

Table 5 Type III Tests of Fixed Effects for Quadratic Trend

Source	Numerator df	Denominator df*	F	Sig.
Intercept	1	116.431	4415.619	.000
Group	1	71.711	135.522	.000
Time	1	139.223	132.826	.000
time * time	1	107.505	77.233	.000

*Decimal numbers in denominators df signifies Welch correction

As Table 5 suggests, quadratic trend is significant: *F (1, 107.5) = 77.23, p = .00.* That is to say the difference between two groups over the course of time is significant. In order to find the difference between the two groups, pairwise comparison is provided.

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

Table 6 Pairwise Comparisons between control and experimental groups

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Df*	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
Experimental	Control	6.811	.585	71.711	.000	5.645	7.978
Control	Experimental	-6.811	.585	71.711	.000	-7.978	-5.645

*Decimal numbers in denominators df signifies Welch correction

If we look at Table 6, we notice that the mean difference of 6.811 between the two groups is statistically significant in favour of the experimental group: mean difference = 6.811 (df=71.71, p = .00, 95% CI [5.645, 7.978). CI may be interpreted as if repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.

Table 7 Estimates of Covariance Parameters for the groups across time

Parameter		Estimate	Std. Error	Wald Z	Sig	95% Confidence Interval	
						Lower Bound	Upper
Residual		7.019	.9573	7.332	.000	5.372	9.170
	Var: Intercept	.0000	.0000	.000	1.000	.0000	.
	Var: [group=1.00]	6.763	2.422	2.792	.005	3.352	13.65
Intercept+ group+	Var: [group=2.00]	.0000	.0000	.000	1.000	.0000	.
Time+ time *time	Var: time	3.080	.9505	3.240	.001	1.682	5.639
[Subject= id]	Var: time*time	.0000	.0000	.000	1.000	.0000	.
	ARH rho	.0000	.0000	.000	1.000	-.7147	.7147

Table 7 provides the random parameters in the model. First of all, the variance of the random intercepts was *Var* (intercept) = 0.00. This suggests that the researchers were correct to assume that students' knowledge at baseline did not vary significantly across groups. The variance of the people's slopes varied significantly *Var* (time) = 3.08. This suggests that the change in knowledge over time varied significantly across students. It holds true with the linear model. Finally, the covariance between the slopes and intercepts (0.00) suggests that there is no relationship between the intercepts and the slope. This amounts to saying that no matter what the means of the groups (intercept) are at the start of the experiment, both groups improve at the same rate (slope). To avoid the ceiling effects, it is desirable to carry out a profile examination. The profile plot of the means of the two groups is provided below. The profile helps us see that the slope of the distribution at the time of posttest is steeper than the pretest for the experimental group. It may follow that the treatments did make a difference among the groups. At the time of the delayed posttest, both control and

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

experimental groups level off signalling a curvilinear trend.

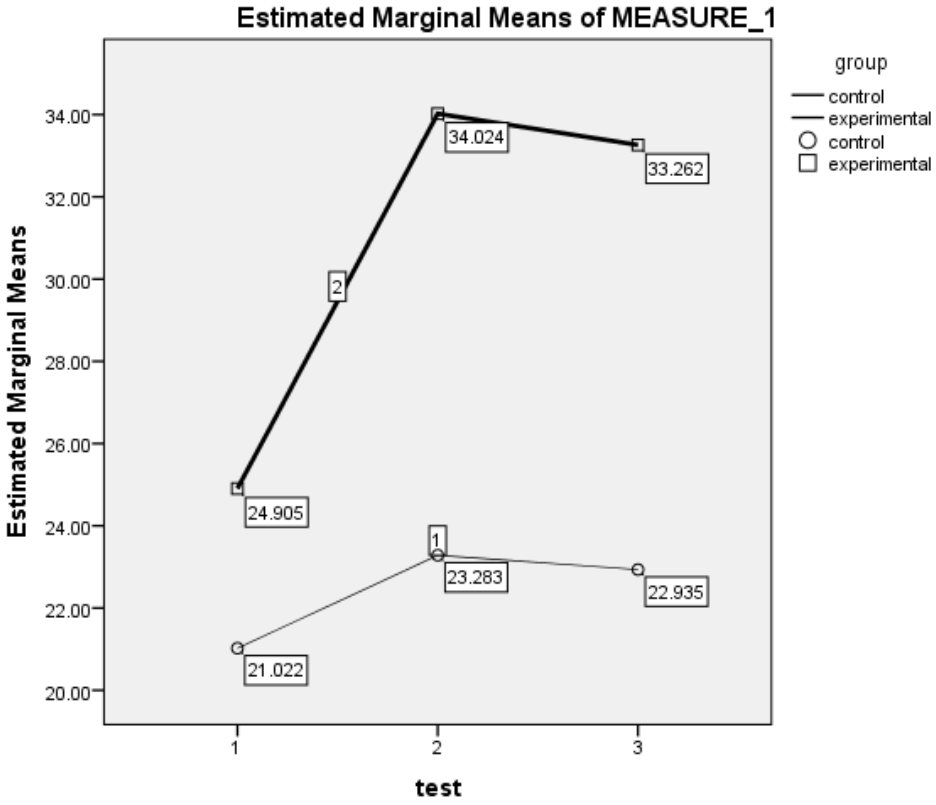

Figure 1 Means of Groups at Pretest, Posttest and Delayed Posttest

5. Discussion

The results of the present study showed that the performance of the students who received weekly quizzes was significantly better than that of the students who received no short quizzes during the course. It is revealed that frequent quizzes may have a positive wash back effect on participants' language achievement and also administering frequent quizzes has led to higher vocabulary achievement. The findings of the present study are in line with Soehren (1997), Graham (1999), and Kamuche (2005) studies which confirmed that weekly quizzes improve students' language performance. These findings can be attributed to certain factors explained below.

First, in order for the students in the experimental group to prepare themselves for the classroom quizzes, they had to review the covered materials more regularly and with more precision before the class (Dustin, 1971; Standlee & Popham, 1960) which presumably makes the long-term retention of the materials possible.

Second, in the present study, short quizzes were based on students' text book and course objectives. Therefore, the students in the experimental group had a better chance to become familiar with course objectives, areas of emphasis, and

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

instructional expectations of the teacher which may have improved their performance. Also frequent quizzes might have provided students in experimental group with extrinsic motivation; in order to get a good mark on short quizzes they had to study hard which might have led to their better performance on achievement test.

Moreover, in short quizzes small amounts of materials are tested and as a result these materials are processed more deeply which can lead to better learning (Standlee & Popham, 1960). Also, students can benefit from classroom discussions after quizzes and recognize their areas of weakness and try to improve them (Selakovich, 1962; Farhady et al., 1994).

Perhaps short quizzes have raised awareness among students in the experimental group and have made them testwise. Test wiseness , in turn, may have improved students' performance.

Typically, short quizzes are not used for evaluation purposes. Hence they do not assign any penalties or grades for the students. That is to say, short quizzes do not bring about debilitating anxiety among the students rather such tests are used to diagnose students' strength and weakness. Such information, in turn, may help teachers better plan their instruction and help students deepen their understanding and improve their learning.

As it was made clear earlier in the paper, the students were provided with specific corrective feedback thereby making them actively involved in the learning experience. When the learners are active, they may become self-regulated learners and better develop their capacity for learning.

6. Conclusion

The paper showed that frequent quizzes brought about durable effects among the students who were exposed to such quick tests. Based on the results obtained from the quizzes, the teacher had to modify his teaching and adapt it to the students' learning needs making the class student-centred. Readers may agree that student-centeredness may foster intrinsic motivation and facilitate learning.

Bearing these findings in mind, it may be safe to hold that classroom frequent quizzes can function effectively in EFL classes. Such tests can boost students' performance and also provide both teachers and students with information regarding the students' mastery of the materials covered in the course. Moreover, by administering weekly short quizzes teachers would motivate their students studying during the course rather than cramming the night before the exam. It is hoped that this important component of teaching and learning find its significance in our classroom syllabus.

The effect of short quizzes must not be over-generalized simply because education is not a flat enterprise rather it is a multivariate phenomenon. There are many factors involved in the process of education including the exigencies of the institutions, background characteristics of the teachers (their perception of pedagogy, teaching style and experience, dedication, etc.) and characteristics of the students (their age, learning style, resources, grade level, etc.).

References

[1] Allan, D. 2004. *Oxford Placement Test 2: Test Pack*. Oxford: Oxford University Press.

[2] Andrade, H.L., 2010. Summing up and moving forward. *Handbook of formative assessment*, pp.344-351.

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

[3] Azorlosa, J.L. and Renner, C.H., 2006. The effect of announced quizzes on exam performance. *Journal of Instructional Psychology*, *33*(4), pp.278-286.

[4] Bachman, L.F., 1990. *Fundamental considerations in language testing*. Oxford University Press.

[5] Bangert-Drowns, R.L., Kulik, J.A. and Kulik, C.L.C., 1991. Effects of frequent classroom testing. *The Journal of Educational Research*, *85*(2), pp.89-99.

[6] Başol, G. and Johanson, G., 2009. Effectiveness of frequent testing over achievement: A meta analysis study. *Journal of Human Sciences*, *6*(2), pp.99-121.

[7] Beaulieu, R.P. and Utyecht, K.M., 1987. Frequently administered formative tests and student achievement. *Journal of Instructional Psychology*, *14*(4), p.195.

[8] Birjandi, P., Nourozi, M., & Mahmoodi, G. 2013. *English Book 3*. Tehran: Ministry of Education.

[9] Cheng, L. and Watanabe, Y. eds., 2004. *Washback in language testing: Research contexts and methods*. Routledge.

[10] Clump, M.A., Bauer, H. and Whiteleather, A., 2003. To attend or not to attend: is that a good question?. *Journal of Instructional Psychology*, *30*(3), p.220.

[11] Cohen, A.D., 1980. Testing Language Ability in the Classroom.

[12] Davies, A. 1990. *Principles of language testing*. Oxford: Basil Blackwell.

[13] Dineen, P., Taylor, J. and Stephens, L., 1989. The effect of testing frequency upon the achievement of students in high school mathematics course. *School Science and mathematics*, *89*(3), pp.197-200.

[14] Dustin, D.S., 1971. Some Effect of Exam Frequency. *The Psychological Record*, *21*, p.409.

[15] Faez, F., 1999. The washback effect of frequent quizzes on the reading ability of EFL learners. *Unpublished master's thesis, University of Tehran, Tehran, Iran*.

[16] Farhady, H., Jafarpur, A. and Birjandi, P., 1994. Testing language skills: From theory to practice. *Tehran: SAMT Publications*.

[17] Field, J., 2003. *Psycholinguistics: A resource book for students*. Psychology Press.

[18] Graham, R.B., 1999. Unannounced quizzes raise test scores selectively for mid-range students. *Teaching of Psychology*, *26*(4), pp.271-273.

[19] Haberyan, K.A., 2003. Do weekly quizzes improve student performance on general biology exams?. *The American Biology Teacher*, *65*(2), pp.110-114.

[20] Harlen, W. and Deakin Crick, R., 2003. Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), pp.169-207.

[21] Heaton, J.B., 1990. *Classroom testing*.Longman

[22] Johnson, B.C. and Kiviniemi, M.T., 2009. The effect of online chapter quizzes on exam performance in an undergraduate social psychology course. *Teaching of Psychology*, *36*(1), pp.33-37.

[23] Kamuche, F.U., 2005. Do weekly quizzes improve student performance?. *Academic Exchange Quarterly*, *9*(3), pp.188-193.

[24] Kling, N., McCorkle, D., Miller, C. and Reardon, J., 2005. The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business*, *81*(2), pp.67-72.

[25] Lewis, M., 1993. The lexical approach (Vol. 1, p. 993).

[26] Lumsden, E.A., 1976. The effect of brief, pre-lecture quizzes on performance on regularly-scheduled tests, class participation, and student-evaluation of lectures. *Journal of Instructional Psychology*.

[27] Madsen, H.S., 1983. *Techniques in Testing*. Oxford University Press, 200 Madison Ave., New York, NY 10016 (ISBN-0-19-434132-1, $5.95).

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

[28] Marchant, G.J., 2002. Student reading of assigned articles: Will this be on the test?. *Teaching of Psychology*.
[29] McKenzie, G.R., 1972. Some effects of frequent quizzes on inferential thinking. *American Educational Research Journal*, *9*(2), pp.231-240.
[30] McKenzie, G.R., 1972. Some effects of frequent quizzes on inferential thinking. *American Educational Research Journal*, *9*(2), pp.231-240.
[31] Narloch, R., Garbin, C.P. and Turnage, K.D., 2006. Benefits of prelecture quizzes. *Teaching of Psychology*, *33*(2), pp.109-112.
[32] Padilla-Walker, L.M., 2006. The impact of daily extra credit quizzes on exam performance. *Teaching of Psychology*, *33*(4), pp.236-239.
[33] Ruscio, J., 2001. Administering quizzes at random to increase students' reading. *Teaching of Psychology*, *28*(3), pp.204-205.
[34] Roediger III, H.L. and Karpicke, J.D., 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), pp.249-255.
[35] Selakovich, D. (1961). An experiment attempting to determine the effectiveness of frequent testing as an aid to learning in beginning college courses in American government. *The Journal of Educational Research*, *55*(4), 178-180.
[36] Soehren, S. E. (1997). The effect of frequent quizzes on short-and long-term academic performance. *Journal of Dental Education*, *61*(4), 339-345.
[37] Sporer, R. (2001). The no-fault quiz**.** *College Teaching*, 49-61.
[38] Standlee, L.S. and Popham, W.J., 1960. Quizzes' contribution to learning. *Journal of Educational Psychology*, *51*(6), pp.322-325.
[39] Tabachnick, B.G. and Fidell, L.S., 2007. Using Multivariate Statistics, 5th Edn Boston: Pearson Education.
[40] Taylor, J.R., 2003. *Linguistic categorization*. Oxford University Press.
[41] Tomlinson, B., 2005. Testing to learn: A personal view of language testing. *ELT journal*, *59*(1), pp.39-46.
[42] Wilder, D.A., Flood, W.A. and Stromsnes, W., 2001. The use of random extra credit quizzes to increase student attendance. *Journal of Instructional Psychology*, *28*(2), pp.117-117.
[43] Wilkerson, L., Stevens, C.M. and Krasne, S., 2009. No content without context: integrating basic, clinical, and social sciences in a pre-clerkship curriculum. *Medical Teacher*, *31*(9), pp.812-821.
[44] Zarei, A.A .2010.Course performance: the effect of quiz frequency on course performance and its relationship with class attendance of Iranian university learners of English. *Journal of Persian Language and Literature, 26*, 52-69.

[DOR: 20.1001.1.25382640.2016.23.2.4.4]

ماندگاری تأثیر آزمون های کوتاه و مستمر بر واژه آموزی دانش آموزان ایرانی

رضا نجاتی[1]

تاریخ دریافت: ۹٤/۷/۵ تاریخ پذیرش: ۹۵/٤/۳

آزمون های کوتاه و مستمر به همراه بازخورد درباره‌ی نقاط قوت و ضعف دانش آموزان ممکن است یادگیری را تسهیل کند. بنا بر این، احتمالاً تأثیر این آزمون ها بر یادگیری زبان انگلیسی در مدارس ایران مهم است. برای پی بردن به اهمیت این آزمون های کوتاه هفتگی، ۸۸ دانش-آموز پایه دوم متوسطه در چهار کلاس، بصورت نمونه موجود، در ۲ گروه آزمایشی و گواه انتخاب شدند. این طرح ۱۲ هفته بطول انجامید که در آن از دانش آموزان گروه آزمایشی ۱۰ آزمون کوتاه بعمل آمد؛ اما از دانش آموزان گروه گواه چنین آزمونی بعمل نیامد. این تحقیق در قالب طرح « پیش‌آزمون- پس آزمون- پس آزمون موخر» با ۲ گروه آزمایشی و گواه انجام شد. نتایج تحقیق، با «تحلیل واریانس مدل‌های مرکب»، نشان می‌دهد که تغییرات درون و برون گروهی در گروه آزمایشی بیشتر از گروه گواه است. بنا بر این، می‌توان مدعی شد که آزمون های کوتاه و مستمر بر واژه آموزی موثرند و این اثر ماندگار است.

واژگان کلیدی: آزمون های کوتاه و مستمر؛بازخورد؛ارزشیابی تکوینی؛واژه آموزی؛ دانش آموزان دبیرستانی

[1]. استادیار دانشگاه شهید رجایی دانشگاه شانانلو سنت لوییزان، تهران، ایران reza.nejati@srttu.edu .

[DOR: 20.1001.1.25382640.2016.23.2.4.4]