

The Noticing Function of Classroom Pop Quizzes and Formative Tests in the Uptake of Lexical Items of EFL Intermediate Learners

Hassan Soleimani¹ & Leila Najafi²

¹Department of Applied Linguistics, Payame Noor University, Tehran, Iran

²Payame Noor University, Tehran, Iran

Correspondence: Hassan Soleimani, Department of Applied Linguistics, Payame Noor University, Tehran, PO box 19395-3697, Iran. Tel: 98-912-253-5368. E-mail: arshia.soleimani@gmail.com

Received: May 18, 2012 Accepted: June 11, 2012 Online Published: July 10, 2012

doi:10.5539/ijel.v2n4p73

URL: <http://dx.doi.org/10.5539/ijel.v2n4p73>

Abstract

The main concern of the present research was to examine the effect of classroom testing in bringing up the noticing of learners using pop quizzes and formative tests. Following the noticing hypothesis by Schmidt (1990, 2001) and backwash effect of testing on teaching, we specifically tested whether noticing through testing might result in better uptake of lexical items by Iranian EFL learners. Following MacKey and Gass (2005), a comparison design study was conducted using three groups of learners with intermediate level of language proficiency. Based on the results of Oxford Placement Test, 77 female EFL students in Iran were selected and randomly assigned into three groups. The first group took pop quizzes, the second group took formative tests, and the third group was incidental learning. Since the data were not normally distributed, non-parametric tests were applied to test the hypotheses formulated for the purpose of the study. The results of the statistical tests revealed the probable positive effect of *noticing function of testing* on the acquisition of lexical items. The research results might be helpful for teachers to include cognitive tasks to bring up noticing and awareness of learners to facilitate input-to-intake process.

Keywords: noticing, testing, pop quiz, formative test, lexical items

1. Introduction

The conscious versus unconscious controversy has always been a core issue in the field of second and foreign language pedagogy, and different approaches regarding the role of consciousness have appeared. Regarding consciousness in language learning, Baddeley (1976, 1997) argues that “a continuum of consciousness mediates our selection of input and that it ranges from being a largely unconscious process to a highly conscious one” (cited in Combs, 2004, p. 5). According to Krashen’s (1981) acquisition-learning hypothesis, language acquisition is a subconscious process. On the other hand, Schmidt’s (1990) Noticing Hypothesis maintains that the subjective experience of noticing is the necessary and sufficient condition for the conversion of input to intake.

To Cook (2008), raising awareness of language in general facilitates second language learning. Learning a second language and developing effective communication requires the mastery of all four skills: speaking, writing, listening, and reading. Vocabulary and grammatical structure are central to mastery in all the skills as well. Vocabulary helps language learners continually improve in four skills at all levels of language learning. According to Nagy (1988), increasing vocabulary knowledge is believed to be central to the process of education both as means and as an end. An area problematic for many students now seems to be lack of enough vocabulary knowledge, and the number of these students is expected to grow. In the literature concerned, research results reveal contradictory issues, and no clear direction regarding the most effective way to teach vocabulary is agreed upon.

Schmidt (2001) claims that “language learners who take a totally passive approach to learning, waiting patiently and depending on involuntary attentional processes to trigger automatic noticing, are likely to be slow and

unsuccessful learners” (p. 30). Hulstijn (2001) points out that most vocabulary is learned from context, but learning vocabulary through reading and listening alone is not very sufficient (cited in Schmidt, 2001).

In recent years, some second language researchers have stressed the effect of noticing on learning different aspects of language including vocabulary. According to McCarten (2007), an important vocabulary acquisition strategy, which Nation (2001) calls “noticing”, is seeing a word as something to be learned. In his view, know-ing what to learn is a necessary prerequisite to learning. Teachers can help learners get into the habit of noticing by making it clear in classroom instruction and homework assignments. Regarding the importance of noticing, McCarten (2007) suggests making personalized groups and teaching students to notice the new information for better learning. According to him, it is a useful step to take time to organize the new vocabulary in some way that allows students to “notice” and apply the target words as the basis for a communicative activity or to record them for review.

Considering the importance of learning vocabulary and challenges that exist in developing vocabulary knowledge, the present study seeks to examine how different test types might bring up noticing on the part of the learners, i.e., the beneficial effect of testing on learning vocabulary. It appears that testing might mediate the noticing of lexical items which may consequently lead to learning of those items.

2. Review of Literature

2.1 Noticing Hypothesis by Schmidt (1990)

Schmidt (1990) argues that learning cannot take place without “noticing”. To Ellis (1994), “Noticing is of considerable theoretical importance because it accounts for which features in the input are attended to and so become intake (information stored in temporary memory which may or may not be subsequently accommodated in the interlanguage system)” (p. 361). Truscott (1998) has also maintained that in the strong form of the hypothesis, favored by Schmidt (1990, 1993, 1994, & 1995), noticing is a necessary condition for learning. A weaker version also exists which is supported by some researchers as noticing is helpful but might not be necessary.

Schmidt and Frota (1986) have argued that two kinds of noticing are necessary conditions for acquisition. The first view holds that learners must attend to linguistic features of the input that they are exposed to, without which input cannot become intake. And the second one postulates that learners must “notice the gap”, i.e., compare the current state of their developing linguistic system, as realized in their output, with the target language system, available as input (cited in Thornbury, 1997). Thornbury (1997) states that in the classroom, the first kind of noticing is usually promoted through activities and procedures involving input enhancement (Sharwood Smith, 1993), whereby targeted features of the input are made salient to facilitate their becoming intake. The second kind of noticing is traditionally mediated through corrective feedback.

Schmidt (1990) postulates that language learners are not free to notice whatever they want whenever they want and that several factors influence noticeability. To him, good reasons exist for supposing that there is a close relationship between availability for noticing and stages of second language development, and that this can be partly explained by formal linguistic considerations but not completely. He explains some factors that influence noticeability: 1. Expectations are most importantly known as determinants of perceptibility and noticeability of linguistic features in the psychological literature as they facilitate the activation of particular psychological pathways, 2. Frequency also increases the probability of an item being noticed in input, 3. Perceptual salience, 4. Skill level also may be a factor influencing noticeability (This includes the automaticity of processing ability), and 5. Task demands also strongly influence noticeability and support the argument that what is learned is what is noticed.

Schmidt (2001) mentions some other factors that determine noticeability: “L2 learners process target language input in ways that are determined by general cognitive factors including perceptual salience, frequency, the continuity of elements, and other factors that determine whether or not attention is drawn to them” (p.7). According to Schmidt (2001), it can be argued that “task requirements, task instructions, and input enhancement techniques affect what is attended to and noticed in on-line processing, therefore causing their effects” (p. 11). Also one issue that is recently linked to vocabulary is Focus on Form (FoF). For example, Laufer (2005) in her paper examined the need for Focus on Form and of Focus on Forms from the vocabulary learning perspective. She argues that comprehensible input is not sufficient for acquiring vocabulary, and consequently Focus on Form is an indispensable component of instruction.

For operationalizing “noticing”, Robinson (2003) suggests on-line and off-line methods. According to Robinson (2003), methodologies for studying the role of awareness and noticing in learning (in a variety of linguistic

domains, across a variety of L2s) have included “both off-line verbal report measures, such as diary entries, questionnaire responses, and immediate and delayed retrospection, and on-line measures such as protocols” (p. 639).

2.2 The Relationship between Testing and Teaching

In this study, the attempt was to put forth the idea that there exists a relationship between teaching, learning, and testing, and that testing can influence learning. As Heaton (1975) states, in the past, the tendency of a large number of examinations was to encourage separation of testing from teaching, but to him, there is a close interrelation between testing and teaching and it is not impossible in practice to work in either field without the consideration of the other. According to Paulston and Bruder (1976) “there should be frequent testing of vocabulary in combination with the reading program. We see little sense in giving formal test on the reading itself.... But the learning of vocabulary is effectively reinforced by frequent testing” (p. 199).

Heaton (1988) has argued that “tests may be constructed primarily as devices to reinforce learning and to motivate the student or primarily as a means of assessing the student’s performance in the language” (p. 5). Regarding the relationship between testing and teaching, Hughes (2003) states that “the proper relationship between teaching and testing is that of partnership....There may be occasions when teaching is poor or inappropriate and when testing is able to exert a beneficial influence”(p. 2).

In this study, two types of tests including quiz and formative tests were used to serve the purpose of enhancing vocabulary learning through noticing. So they take the function of teaching device ignoring the scores that participants may gain in these tests. “Quiz is a compromise between short- term, subjective evaluation based on daily work and the longer term achievement test” (Chastain, 1988, p. 381). According to Chastain (1988), the quiz is the most common classroom test which is favored by many language teachers. Often, the threat of a quiz is applied as a technique by teachers to stimulate students to prepare their daily lesson. A more positive approach is the situation in which students get ready to be able to participate in classroom activities. The quiz may be announced or unannounced.

The other type of test applied in this study was the formative test. Hedge (2000) has stated that assessment is undertaken for different purposes, and one purpose is formative assessment which is “pedagogically motivated” where the teacher gains information from assessments about a learner’s progress and uses it as the basis for further classroom work. According to Hedge (2000), the purpose of formative assessment is keeping track of the learners’ progress as it happens and identifying ways of helping it during the course. The focus of formative assessment is on the process of learning. According to Brown (2004), “Informal assessment can take a number of forms, starting with incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student” (p. 5). He states that a good deal of a teacher’s informal assessment happens during classroom tasks designed by the teacher to make students perform and, then record results and make steady judgments about a student's competence. Paulston and Bruder (1976) suggest classroom tests as an efficient technique to improve vocabulary knowledge and maintain that “tests should be frequent (once a week) and short (10 minutes) with, if convenient, a longer midterm and final exam” (p. 199).

2.3 Washback as Noticing Facilitator

Regarding the interaction between teaching and testing, the purpose of this study was to bring testing to the class and highlight its role in learning. In the present study, testing was used to bring up noticing. Testing was intended to make the students notice the words that were unknown to them in the texts. This was believed to realize the noticing hypothesis of Schmidt (1990) and tap students thought processes and consciousness to pay focused attention to the new words for better learning.

Bachman and Palmer (1996) state that the test takers may improve their language knowledge while taking the test or by receiving feedback after the test. Also, the characteristics of the test task may affect the strategies test takers use. According to Fulcher and Davidson (2007), washback is sometimes known as backwash, and it is “the effect of a test on learning and teaching. They also maintain that washback studies focus on practices or behaviors which would not be present if it were not for the test” (p. 377).

According to Pan (2009), systemic validity is another term introduced by Fredericksen and Collins (1989) which refers to the effects that the introduction of the test into an educational system brings about in instructional methodology. It is stated that tests cause instructional changes in the education system and these changes encourage cognitive skills that the test is designed to measure. Brown (2004) states students achieve washback when they recognize their areas of success and challenge through the testing experience. According to him, a test achieves washback when it becomes a learning experience.

Shih (2009) points out that researchers have paid most of their attention on the washback of tests on four dimensions of teaching practice including “(1) content of teaching (2) teaching methods (3) assessment methods (4) overall teaching style, classroom atmosphere and teachers’ feelings toward the test” (p.188). Hughes (1993) distinguishes between testing effects on participants, processes, and outcomes. The effect that a test has on participants including the teacher, learners, and materials writers preparing for a test and their perceptions and attitudes to the task encourages them to modify their teaching and learning processes and these flowingly effect on products that are the learning outcomes including knowledge of target skills and test scores (cited in Green, 2007). In the present study, considering the necessity of noticing as a process to internalize vocabulary and regarding the dearth of studies concerning the role of testing in helping learners noticing the input, the beneficial role of noticing through pop quizzes and formative tests and their probable effect on learners’ uptake of vocabulary was examined.

3. Method

3.1 Design of the Study

The present study follows a comparison design (Mackey & Gass, 2005). Since the purpose was to compare the effects of different kinds of testing on the noticing of lexical items and then compare the effect of testing with another approach to vocabulary learning namely incidental learning, three groups were established to help us attain this purpose. The study also benefits a pre-test post-test design among the first two groups. The third group to be compared with these two groups was the post-test only group. Of course, the design of pre-tests in the first group was different from that of the second group regarding the order of presentation of tests and treatments.

3.2 Participants

The participants in the study were 77 female intermediate English as a foreign language (EFL) learners sampled from a language institute in Qom, Iran. A version of Oxford Placement Test was applied to make sure that the participants were all from the intermediate level and homogeneous. The analyses carried out with the data collected from the placement test allowed the researchers to confirm this initial equivalence. Based on the results of the placement test, the 77 EFL students were randomly assigned into three groups: 27, 24, and 26 participants in different groups. The three groups had the same background and were taught the same books. Their age range was from 17 to 30 years old.

3.3 Instruments

To attain the purposes of this study, some materials were used for the treatments, and some tests were designed by the researchers as instruments.

3.3.1 Oxford Placement Test (OPT)

OPT test was applied to select homogeneous intermediate groups. The test contains 50 multiple choice questions which assess students’ knowledge of key grammar and vocabulary from elementary to intermediate levels, a reading text with 10 graded comprehension questions, and an optional writing task.

3.3.2 Active Skills for Reading (Book 2: Intermediate)

The words from *Active Skills for Reading (Book 2)* by Anderson (2007) were the targeted items for this study. This book was chosen because the level matched the level of participants and also the context, and the subjects were common and interesting.

3.3.3 Pre-Tests

As it was explained in the design of the study, two groups out of the three groups in this study took pre-tests. The pre-tests were designed by one of the researchers to perform two functions of Pop-Quiz in group one and Formative Test in group two. The purpose was to have students engage in thought processes and focus on the new words consciously to notice them for better uptake.

The pre-tests were conducted during three different sessions. Each test was composed of twenty questions. So the total number of the words in the pre-tests for every group was sixty. Tests were multiple-choice tests of vocabulary, and the words were chosen from *Active Skills for Reading (Book 2)* for intermediates. The new words were tested in the context of sentences derived from a version of Cambridge dictionary.

3.3.4 Post-Test

Based on the results of pre-tests, some words were excluded from the list of words as known words out of the 20 items in every pre-test. Only 45 words were recognized as unknown words. So the post-test included 45 words out of 60.

To avoid practice effect, the sentences in pre-tests were not included in the post-test. The researcher put the unknown words in the context of other sentences from Cambridge or the examples in the glossary at the end of the source book. Also, the distracters of the multiple-choice items were changed.

3.4 Pilot Study

Before starting the main study, we needed to make sure of some issues related to the participants, design, and procedure of the study. To this end, a group of intermediate students were selected to carry out a pilot study. The pilot study was instrumental in deciding on the final design and procedure of the study. The following decisions were made based on the pilot study:

- 1) During the pilot test, we noticed that students had difficulty with some distracters because they were close in meaning and caused confusion and took the students' attention away from the questioned word, so they needed to be changed. The students had underlined some words in the distracters as unknown words. This meant that they were difficult and needed to be excluded or replaced by other easier equivalent words.
- 2) One issue that was not taken into account was the students' motivation. Because the pilot test was conducted as extracurricular plans in the class, the students were not motivated to participate and considered it as a burden upon them. So it was decided that the real procedure of the study be included in the actual syllabus of the class.
- 3) Almost 70 minutes was enough for the teacher to complete the treatments and for the students to take the test at the same session.
- 4) The subjects should be able to understand the contexts and know all of the words used in the stem and distracters except the words tested.

To test the reliability of the post-test, the post-test was piloted on another group who had studied the book *Active Skills for Reading* (Book 2) before. Internal consistency of the post-test as a measuring instrument was calculated for pilot group by using Kuder-Richardson 21(K-R 21) formula ($K-R 21 = 0.67$).

3.5 Procedure

3.5.1 Placement Test

In order to conduct the study with intermediate level, a placement test was needed to choose as homogeneous students as possible. To this end OPT was applied, and based on the results of the placement test three classes were identified as intermediates. The students of the selected classes were randomly assigned into three homogeneous groups of respective 26, 24, and 26.

3.5.2 Treatments and Pre-Tests during the Study

Each group in the study followed a different procedure to serve the purposes of the study. Three groups of EFL students participated in the study.

The participants in group one attended the class every other day, three times a week, 90 minutes each. The study conducted on this group was accomplished during four sessions (regardless of the session during which the students took the placement test). During the first three sessions, the study was pursued like the following: the pop quizzes that were prepared beforehand were given to the participants. To enhance the noticing of lexical items that the participants encountered in the quizzes, they were not told that they were going to receive such tests. The participants were asked to read every question carefully and choose the item that best corresponded the meaning of the questioned word. To draw the students' attention on the unknown words, the researchers asked them to underline the words that they judged as unknown. This was for the purpose of making sure that students notice the unknown words and focus on them. The participants were asked not to guess the meaning of unknown words, because we needed to know how many words were new to them and later compare the results of these pre-tests with their post-test.

After the pop quizzes, the participants received a treatment on the new words that they had just confronted in the pop quiz. The treatment included simple provision of the meaning of new words. No special technique for teaching the new words was used because the main purpose of the study was to test the effect of noticing of lexical items. This procedure continued for three sessions, and a total of 60 new words were taught during this period.

The procedure of the study for group two was the opposite: the participants were asked to read the passages and underline the unknown words. The purpose of underlining is the same as that of group one; in addition, noticing of lexical items in group two was compared to those of group one to find out whether students noticed more in any one of these conditions and if more noticing resulted in more learning.

After underlining the words, the participants received treatment. The treatment again was carried out in the same way as group one. It was the simple provision of the meaning of new words in context. The participants were given a formative test after the treatment. The tests taken by group two were the same as the tests taken by group one. The titles of the tests were different and that was because of the different function that we intended them to perform; as the pop quiz in group one and formative test in group two. Here in group two, the participants were not required to underline the words in formative tests because the words were taught beforehand in the treatment, and they were no longer new for the students.

The purpose of the formative test was to infer how much learning occurred due to noticing and if this formative test *would* lead to more learning measured by the post test. This procedure of the treatment was followed by the formative tests continued for three sessions.

The participants of group three were incidental learners. They were asked to study some passages from active readers (Book 2), on their own, out of the class in a time span determined by the researcher. In fact, they did not receive any treatments or tests in the class. Also, it was not specified in their assignment to underline the new words that they encountered in the passages since the main aim of the researcher was to test if incidental learning was effective compared to the learning which was supposed to take place in groups one and two.

The students were not informed that they would take a vocabulary test at the end, to make sure that they would not apply any kind of strategy or technique for learning the words, or to attempt to memorize the words for the sake of examinations.

3.5.3 Post Tests

The post test was the same for all of the three groups. It was held to measure how much of the new words the participants had learned. Of course, based on the results of the pre-tests in groups one and two, a total of 15 words were removed out of the 60 words because they were proved to be known by most of the students. Only 45 words were assumed to be unknown and were included in the post test.

Table 1 below shows the statistics of *M*, *SD*, and *V* for each group's post-test. Based on this table, the reliability of the tests was calculated. Internal consistency of the test as a measuring instrument was calculated for the post-test of every group by using Kuder-Richardson 21 (K-R 21) formula. The obtained value of K-R 21 was 0.78 for group one, 0.72 for group two, and 0.81 for group three.

Table 1. Descriptive statistics of the post-tests

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>V</i>
Post Test of Group One	27	26.5185	6.95795	48.413
Post Test of Group Two	24	23.3333	6.18319	38.232
Post Test of Group Three	26	16.8462	7.08063	50.135
Valid N (listwise)	24			

3.6 Data Analysis

For data collection, two methods were applied. To test noticing of participants, on-line method of underlining (following Izumi & Bigelow, 2000; Soleimani, Ketabi & Talebnejad, 2008) was used, and to test the uptake of lexical items, post tests were accomplished. For scoring of tests, the number of correct responses was counted, and the marks were given out of the total number of test items. And for quantifying noticing, the number of underlined words by the participants was counted.

To decide whether to use parametric or non-parametric tests, we first conducted tests of normality. Following Soleimani (2009), among numerical methods, skewness, and among graphical methods, histogram were picked up. The analyses proved that the data were not normally distributed, so the non-parametric tests were applied to answer the three research questions. Wilcoxon Signed Ranks Test was performed to answer the first research question. Mann-Whitney Test was applied to compare the results of underlining by the two groups, and then the post tests of groups one and two were also compared using Mann-Whitney Test to answer the second research question. To answer the third research question the behavior of the three groups in terms of post tests was compared using non-parametric Kruskal-Wallis Test, and to find the differences between the post tests of groups one and three and the differences between groups two and three Mann-Whitney Test was conducted. As the study enjoyed comparison design, tests of comparison were needed to answer the research questions.

4. Results

First of all, normality tests were carried out to decide whether to use parametric or non-parametric tests for the data analysis. The numerical and graphical normality tests demonstrated that distributions of the scores of post tests of group one, group two, and group three were not normal. Regarding the distributions of the data which were not normal and taking this fact into account that the sizes of the groups in this study were small, non-parametric tests were chosen to be carried out in search of the answers to the research questions.

Three questions were set forth in this study. Considering the design of this study which was a comparison design, the researcher applied the Mann-Whitney Test to understand whether there was any difference between the behaviors of two samples in question. The non-parametric test of Wilcoxon Signed Ranks was applied to test the possible difference between the behaviors of two paired samples, namely, the underlining of lexical items in tests. For the comparison of three or more samples, the Kruskal-Wallis Test was applied to find out whether there were any differences between the behaviors of the samples at issue. The alpha level was set at 5%.

4.1 Research Question One

The first question in the study is whether testing has any influence on the noticing of lexical items in input. As it was mentioned before, two kinds of tests were applied in this study as pre-tests in groups one and two: pop-quizzes and formative tests.

The participants in group one were given a vocabulary pop-quiz before the treatment to test its effect on the noticing of lexical items and their uptake.

Wilcoxon Signed Ranks Test was applied to compare the results of pop-quizzes altogether with the results of post-test in group one to see whether there was any improvement in the participants' knowledge of the new words in their input. Table 2 shows the results of Wilcoxon Signed Ranks Test.

Table 2. Wilcoxon Signed Ranks Test for pop-quizzes and post-test in group one

Post Test of Group One (Pop-Quiz Group)	
Z	-4.546 ^a
Asymp. Sig. (2-tailed)	.000

In Table 2 we can see that the probability value of the Wilcoxon Signed Ranks Test is .00 which is less than the level of significance of .05. From the results, it can be concluded that there is a statistically significant difference between the scores of pop-quizzes and post-test among the participants of group one. So the null hypothesis that the pop-quiz does not influence noticing of lexical items is rejected.

The participants in group two were asked to underline the new lexical items in input to make sure that they noticed the new words in input. The participants then received treatment on the new lexical item. Then, they were given a formative test on the vocabulary they had just confronted in the texts.

Wilcoxon Signed Ranks test was applied to compare the results of formative tests with the results of post-test in group two, to see whether there was any improvement in the participants' knowledge of the new words in their input. Table 3 shows the results of Wilcoxon Signed Ranks Test.

Table 3. Wilcoxon Signed Ranks Test for formative tests and post-test in group two

Post Test of Group Two (Formative Test Group)	
Z	-4.290 ^a
Asymp. Sig. (2-tailed)	.000

In Table 3, we can see that the probability value of the Wilcoxon Signed Ranks Test is .00 which is less than the level of significance of .05. It can be concluded that there is a statistically significant difference between the scores of formative tests and post-test among the participants of group two. So the null hypothesis that formative tests do not influence noticing of lexical items is rejected.

4.2 Research Question Two

The second question in this study is whether different types of tests such as pop quizzes and formative tests have different influences on the noticing of lexical items. The participants were asked to underline the unknown words. Later, the researcher counted the total number of underlined words in pop-quizzes by group one and the total number of underlined words in treatments by group two as measures of noticing lexical items by the two groups.

To answer this question, the researcher first used the non-parametric test for two independent samples which was Mann-Whitney Test and median test to compare the results of underlining by the two groups and to see whether there were any significant differences between the two groups in terms of noticing the lexical items. Then the post tests of the two groups were also compared using Mann-Whitney Test to see whether there were significant differences between the two groups in terms of post tests.

These comparisons sought to find the possible concordance between the patterns of post-test of group one and post-test of group two on the one hand and underlining of lexical items by group one and group two on the other hand.

Tables 4 and 5 show the results of Mann-Whitney Test conducted for comparing underlining by groups one and two.

Table 4. Mean ranks for underlining lexical items

	N	Mean Rank	Sum of Ranks
Underlining in Group One	81	78.89	6390.00
Underlining in Group Two	72	74.88	5391.00
Total	153		

Table 5. Mann-Whitney Test statistics for underlined lexical items

Underlining of Lexical Items in Groups One and Two	
Mann-Whitney U	2763.000
Wilcoxon W	5391.000
Z	-.569
Asymp. Sig. (2-tailed)	.570

Table 4 shows that the value of mean ranks for underlining in group one and underlining in group two are close to each other. And Table 5 shows that the probability level of Mann-Whitney Test is .570 which is much higher than .05. So these lead to the conclusion that there is no difference between the distributions of these data.

Tables 6 and 7 show the results of Mann-Whitney Test for the comparison of the post tests of groups one and two.

Table 6. Mean-Ranks for post-tests of groups one and two

Posts Tests	N	Mean Rank	Sum of Ranks
Group 1	27	29.52	797.00
Group 2	24	22.04	529.00
Total	51		

Table 7. Mann-Whitney Test for post-tests of groups one and two

Posts Tests	
Mann-Whitney U	229.000
Wilcoxon W	529.000
Z	-1.797
Asymp. Sig. (2-tailed)	.072

Table 6 shows a little higher mean rank for post test of group one compared to the post test of group two. However, the difference of 29.52 and 22.04 mean ranks is not very significant. And Table 7 shows that the probability level of this test is .072 which is higher than .05. These show that there is no significant difference between the two post tests. So the null hypothesis that there is no difference between the effects of pop-quiz and formative test on the noticing of lexical items is accepted.

4.3 Research Question Three

The third question of the study is whether there is any difference between the influence of testing and other approaches such as incidental learning on the noticing and uptake of lexical items. Here, the aim is to explore the

possible differences between post-tests of groups one, two, and three. Participants of group three did not receive any kind of tests or treatments. First, the behavior of the three groups in terms of post tests was compared using non-parametric Kruskal-Wallis Test. Tables 8 and 9 show the results of this test.

Table 8. Mean ranks for post-tests of the three groups

Posts Tests	N	Mean Rank
Group 1	27	50.96
Group 2	24	41.4
Group 3	26	24.37
Total	77	

Table 9. Kruskal-Wallis Test statistics for the post-tests of the three groups

	Posts Tests
Chi-Square	19.179
Df	2
Asymp. Sig.	.000

Table 8 shows the mean ranks of the post tests in descending order from group one to group three. It can be seen that the mean rank of group one is higher than the mean rank of group two and three, and that the mean rank of group two is higher than the mean rank of group three. Table 9 shows the Chi-Square value obtained for the post-tests and the probability value.

The critical value of Chi-Square at significance level of .05 and with the degree of freedom of 2 is 5.99, which is smaller than the observed Chi-Square which is 19.179. Also, the probability of the Kruskal-Wallis Test is .00 which is smaller than the level of significance of .05. Therefore, taking all these results into account, we can reject the null hypothesis that there is no difference between the post-tests of the three groups involved in this study.

To find the differences between the post tests of groups one and three and the differences between groups two and three, Mann-Whitney Test was conducted.

Tables 10 and 11 show the results of Mann-Whitney Test for groups one and three.

Table 10. Mean ranks for the post-tests of groups one and three

Posts Tests	N	Mean Rank	Sum of Ranks
Group 1	27	35.44	957.00
Group 3	26	18.23	474.00
Total	53		

Table 11. Mann-Whitney Test statistics for post-tests of groups one and three

	Posts Tests
Mann-Whitney U	123.000
Wilcoxon W	474.000
Z	-4.063
Asymp. Sig. (2-tailed)	.000

From Table 10, it can be inferred that the mean rank of post test of group one is higher than the mean rank of the post test of group three. Also, Table 11 shows that the probability level of this test is .00 which is less than the significance level of .05. So it can be inferred that the two groups' post tests differ significantly.

Tables 12 and 13 show the results of Mann-Whitney Test for the post tests of groups two and three.

Table 12. Mean ranks for the post-tests of groups two and three

Posts Tests	N	Mean Rank	Sum of Ranks
Group 2	24	31.85	764.50
Group 3	26	19.63	510.50
Total	50		

Table 13. Mann-Whitney Test for post-tests of groups two and three

	Posts Tests
Mann-Whitney U	159.500
Wilcoxon W	510.500
Z	-2.968
Asymp. Sig. (2-tailed)	.003

Comparison of the mean ranks of these groups shows that the mean rank of group two is higher than that of group three. By looking at Table 13, we can see that probability level of this test is .003 which is less than the .05. So the two groups' post tests results are significantly different.

5. Discussion

As the analysis of data obtained from this study revealed, the first null hypothesis in this study was rejected to confirm the opposite direction that pop-quizzes had positive effect on the noticing of lexical items. In the same vein, it was proved that formative tests affected the noticing of lexical items and their uptake. This was examined by comparing the results of pre-tests and post-tests taken by the same groups using. The results showed that students' performance had improved from pre-tests to post-tests due to the procedure followed in this study that was testing mediating noticing and learning of vocabulary items.

Students' noticing of new lexical items was tested using online measure of underlining the unknown words. The results of underlining by the participants of both groups were compared and proved to be the same. According to the statistics, also, the results of post-tests of group one and group two were compared and yielded no difference between groups in terms of their uptake of lexical items which was operationalized by post-test scores. With these results put together, it can be concluded that in both group one and group two, testing had the same effect on noticing which subsequently yielded equal results for the uptake of lexical items. Apparently, the second null hypothesis was accepted which maintained that there was no difference between the effects of pop-quizzes and formative tests on the noticing and uptake of lexical items.

Some studies in the literature of second language learning are in line with the findings of this study regarding washback effect, and some are in contrast with it. For example, the findings of the study by Jilani (2009), which described and evaluated the higher secondary school certificate exam in Pakistan whose primary form was maintained for more than thirty years, revealed that the exam had widespread washback effect in both individual and societal levels. A review study of the empirical studies of washback from external exams and tests that had been carried out in the field of English language teaching was conducted by Spratt (2005) to find the relationship between washback and the roles that teachers could play and the decisions that they could make. Her review study indicated that exams could not dictate what and how teachers taught and learners learnt and rejected washback effect

The beneficial effect of applying meta-cognitive strategies such as self-initiation and selective attention in learning vocabulary was also proved in a study by Gu and Johnson (1996) which aimed to establish the vocabulary learning strategies used by Chinese university learners of English and the relationship between their strategies and outcomes in learning English. The results of their study revealed that self-initiation and selective attention, two meta-cognitive strategies, were positively related to College English Test scores.

Regarding Noticing Hypothesis of Schmidt (1990), the findings of this study were in line with some other studies such as that of Mackey (2006). The results of the present study proved that there is a relationship between interactional feedback in the classroom, the learners noticing, and their learning of L2 question forms.

Another approach to vocabulary learning discussed in second language learning and acquisition literature is incidental learning. The third group in this study followed this approach. Results of the test given to this group was compared to the post-tests of groups one and group two. The results revealed the superiority of pop-quizzes

and formative tests effects on noticing and learning of vocabulary over incidental learning. The variable lacking in the third group was testing and the intended noticing thereof. The third null hypothesis was rejected. It can be inferred from this that noticing did not take place in group three and that participants did not pay conscious attention to the new words.

One study which supported the superiority of explicit instruction over implicit instruction was that of Radwan (2004). The results of the study by Radwan (2004) which explored the facilitative effects of various types of attention-drawing instructional conditions on the acquisition of English dative alternation showed that students who received explicit instruction performed better than those exposed to implicit instruction and a higher level of awareness correlates positively with language development

6. Conclusion

In the current study, the purpose was to examine testing as a possible way of bringing about noticing of lexical items in students. This would mix two approaches to learning in the literature: noticing hypothesis by Schmidt (1990) and the effect that testing might have on learning and teaching.

In the literature, many approaches for bringing up noticing have been tested. And according to the results of this study, tests of different type proved to be effective way for inducing noticing. Pop-quiz and formative tests were put into the study in two groups, and the results of post tests showed that these two types of tests might be efficient tools in inducing noticing and that there was no difference in their effects. On the other hand, incidental learning of lexical items was tested in group three, and it was proved by the post test scores that the pop-quiz and formative test groups outperformed the incidental learners in group three.

Considering the status of tests in schools held formally to assess students' achievement and assign their scores which provokes test anxiety and takes away students' consciousness and awareness and regarding the results of the present study that tests can serve pedagogical purposes and assessment purposes, it might be inferred that school teachers can apply different types of classroom tests and quizzes to their teaching methodology. This implies that testing can be a teaching device. Teachers can have students take some quizzes or tests during the term to teach them cognitive strategies of selective attention.

References

- Anderson, N. J. (2007). *Active Skills for Reading (Book 2)* (2nd ed.). Canada: Thomson ELT.
- Backman, L. F., & Palmer, A. S. (1996). *Language testing in Practice*. Oxford: Oxford university press.
- Batstone, R. (1996). Key concepts in ELT. *ELT Journal*, 50(3), 373. Retrieved October 27, 2011, from <http://eltj.oxfordjournals.org/>
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. Whiteman Plains: Pearson Education.
- Chastain, K. (1988). *Developing second language skills: theory and practice* (3rd ed.). US: Harcourt Brace Jovanovich Inc. publishers.
- Combs, C. (2004). What cognitive processes are triggered by input enhancement? *Columbia University Working Papers in TESOL & Applied Linguistics*, 4(Special Issue), 1-14. Retrieved September 12, 2011, from <http://journals.tc-library.org/index>.
- Cook, V. (2008). *Second Language learning and language teaching*. London: Hodder education.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Fredericksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researchers*, 18(9), 27-31.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment (an advanced Resource book)*. London and New York: Routledge.
- Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97. <http://dx.doi.org/10.1080/09695940701272880>
- Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643-679. <http://dx.doi.org/10.1111/j.1467-1770.1996.tb01355.x>
- Heaton, J. B. (1988). *Writing English language tests*. New York: Longman.
- Hedge, T. (2000). *Teaching and learning in the classroom*. Oxford: Oxford University Press.

- Hughes, A. (2003). *Testing for language teachers*. (2nd ed.). United Kingdom: Cambridge University Press.
- Izumi, Sh., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly*, 2(34), 239-273. <http://dx.doi.org/10.2307/3587952>
- Jilani, R. (2009). Problematizing high school certificate exam in Pakistan: A washback perspective. *The Reading Matrix*, 9(2), 175-183. Retrieved November 20, 2011, from <http://www.readingmatrix.com>
- Krashen, S. D. (1981). *Principles and Practice in Second Language Acquisition*. UK: Prentice-Hall International Ltd.
- Lauffer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook*, 5, 223-250. Retrieved January 12, 2011, from <http://www.ingentaconnect.com>
- Mackey, A., & Gass, S. M. (2005). *Second language research: methodology and design*. London: Lawrence Erlbaum Associates Publishers.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-430. <http://dx.doi.org/10.1093/applin/ami051>
- McCarten, J. (2007). *Teaching vocabulary: lessons from the corpus lessons for the classroom*. New York: Cambridge University Press.
- Nagy, W. E. (1988). *Teaching vocabulary to improve reading comprehension*. Newark, DE: International Reading Association.
- Pan, Y. C. (2009). A review of washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages*, 25(2009), 257-263. Retrieved January 12, from http://js.vnu.edu.vn/nn_4_09/b.8.pdf
- Paulston, C. B., & Bruder, M. N. (1976). *Teaching English as a second language: techniques and procedures*. Cambridge, Massachusetts: Winthrop Publishers Inc.
- Radwan, A. A. (2004). The effectiveness of explicit attention to form in language learning. *System*, 33(2005), 69-87.
- Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Australia: Blackwell Publishing Ltd. <http://dx.doi.org/10.1002/9780470756492.ch19>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158. <http://dx.doi.org/10.1093/applin/11.2.129>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press. Retrieved from www.nflrc.hawaii.edu/aboutus/schmidt/
- Schmitt, N. (Ed.). (2002). *An introduction to applied linguistics*. New York: Oxford University Press.
- Shih, Ch. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 188-206. Retrieved November 20, 2011, from <http://education.waikato.ac.nz/research/files/etpc/files/2009v8n2dial1.pdf>
- Soleimani, H., Ketabi, S., & Talebinejad, M. R. (2008). The noticing function of output in acquisition of rhetorical structure of comparison paragraphs of Iranian EFL students. *Linguistics Online*, 32(2), 59-74. Retrieved December 13, 2010, from <http://www.linguistik-online.de/>
- Soleimani, H. (2009). *Non-Parametric statistics for applied linguistics research*. Tehran: Rahnama Press.
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29. <http://dx.doi.org/10.1191/1362168805lr152oa>
- Thornbury, S. (1997). Reformulation and reconstruction: tasks that promote "noticing". *ELT Journal*, 51(4), 326-335. <http://dx.doi.org/10.1093/elt/51.4.326>
- Truscott, J. (1998). Noticing in second language acquisition: a critical review. *Second Language Research*, 14(2), 103-113. <http://dx.doi.org/10.1191/026765898674803209>