Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

An Investigation into the Effectiveness of the Adopted Strategies through Dynamic Assessment and Their Comparison with Reference to Washback Effect

Somayeh NOROOZI[1], Ali KAZEMI[2,*]

[1](M. A. Student of TEFL), Dept. of English, College of Humanities, Yasouj University

[2](Ph. D.), Academic Faculty Member of Dept. of English, College of Humanities, Yasouj University, Yasouj, Iran;

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. In the current climate of testing practices, language assessment is expected to serve the purposes of testing and teaching alike. This ambition is most pronounced in Dynamic Assessment (DA), which consists in using a number of strategies during assessment, ranging from simple hints and prompts to detailed explanations. The present study aimed to specify, besides the effectiveness of the strategies, whether they affect learning equally and which one has the most positive effect on learning comparing others. To this end, using a pre-experimental research design, and applying a test–teach–retest paradigm, consistent with the interventionist approach to DA, intervention (teaching strategies) was performed in six phases in a group of 30 elementary students, who were randomly selected, to specify the effectiveness of the strategies. In the seventh session, employing a true-experimental design, 15 students received detailed score reporting as one of the strategies supposed to have positive washback effect while the other 15 students were provided with other strategies (simple hints and different prompts). The analysis of t-tests indicated that all learners made score gains in posttests in all phases, however, the experimental group, who received detailed description of the test, benefited significantly compared to the control group, who received other strategies. As the strategies used during dynamic assessment and consequences of detailed score reporting can be beneficial for both testers and learners, this type of assessment and the given strategy is recommended. This study further discusses the significance of the findings in the context of testing and learning.

Keywords: Dynamic Assessment, Language learning, Language testing, strategies, washback effect, interventionist approach

1. INTRODUCTION

Language teaching and testing are interrelated and play fundamental roles in providing an appropriate situation for L2 learners to be competent enough with regard to proficiency level. Definitely teaching influence testing which, in turn, has effect on learning. Without good teaching good testing will not work. Through language testing the abilities of the students will be discovered; due to the results of the tests, educational decision will be made by the teachers, therefore based on tests results, teaching methodology may change.

Good testing happens when competent testers use instructive approaches to language testing and learning. Language learning is a product of transmission and the teacher and/or the tester is responsible for transmitting all of the information to the students in a way that is in accord with students' needs.

Student achievement problems are often highlighted in academic literature and therefore, it is incumbent upon educators to study the factors related to student achievements. Static assessment, which is a short-run, product-based assessment without elaboration of learning

*Corresponding author. *Email address: akazemi@yu.ac.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

http://dergi.cumhuriyet.edu.tr/cumuscij

potentials, has been a concern in testing situations. It does not mirror teaching process, but DA, which is a powerful intervention procedure, is a new approach to teaching, learning and testing that makes the learner a problem solver-thinker. Haywood and Tzuriel [20] define dynamic assessment (DA) as "a subset of interactive assessment that includes deliberate and planned mediational teaching and the assessment of the effects of that teaching on subsequent performance" (p. 40). In such an assessment what learners know is not so noticeable; how they learn which refers to the teaching phase is a question which has garnered many researchers' attention. The goal of DA is to monitor students' learning and provide feedback in regular intervals in order to improve learning and teaching and testing.

Cognitive developmental theories of Vygotsky and Feuerstein, who believe that human beings' abilities are developmental and not static, have led to the creation of the concept of dynamic assessment. Vygotsky proposed the concept of zone of proximal development (ZPD), which is a sociocultural approach. Likewise, Feuerstein proposed mediated learning experience (MLE) theory. According to Poehner [28] 'zone of proximal domain' (ZPD), with regard to Vygotsky's approach, is a distance between the actual developmental level of the learner and the potential development through adult's support. In such a domain, a cooperative situation is established [34, 35]. Also with regard to Feuerstein's Mediated Learning Experience (MLE) theory, mediation is used by adults in order to modify the children's learning [22, 33]. With respect to these two views (Vygotskian view and Feuerstein's view) as two bases for dynamic assessment, mediation in the form of cooperative instruction work in learning situation and support teaching.

Decisions about general education has caused exploration of an accurate procedure of assessment, which is a powerful intervention for an effective instruction in order to provide information about students' learning ability; that is to say, students must benefit from instruction and intervention during assessment. Employing the paradigm of test-teach-test, DA is an integration of assessment and intervention. Many researchers believe that DA, which can be provided for different groups of learners with different goals and needs, distinguishes between individual differences and makes the teachers aware of how different learners can learn differently. According to Carlson and Wiedl [13], through training in DA, regarding individuals' performance, relevant modifications and intervention will be applied into the test situation for the learners. Tzuriel [33] states simply that DA is not just an effective evaluation, but it is proved to be an appropriate intervention process. He adds DA is a distinct and independent theory in psychology of assessment on the basis of its goal and methodology. Following Tzuriel, Haywood and Lidz [19] comment that DA identifies the optimal learning and teaching conditions, appropriate interventions, and potential performance.

In teaching phase of DA, mediations employed consist of various strategies which may affect learning. Comparing these strategies, the notion of 'washback' may be posed. Many researchers have defined washback as the effects of tests on teaching and learning which may lead to change and innovation. Amengual-Pizarro [3] puts it in these words: "The effects of tests on teaching and learning referred to as 'washback' or 'backwash' in applied linguistics" (p. 582). Andrews, Fullilove, & Wong [4] believe that "washback is used to refer to the effects of tests on teaching and learning, the educational system and the various stakeholders in the education process" (p. 208). According to Hamp-Lyons [18] tests play an important role in social lives, so being concerned about the results of the tests is natural. Consequences of tests are very critical and applicable in educational system particularly for the teachers who decide about educational plan and teaching methodology on the basis of the results extracted from the tests. Also Spolsky [31] states that tests include some unavoidable problems which lead to

unreliable results; decisions are made on the basis of these results, thus it is essential to be cautious about the interpretation of the tests results.

Generally what happens in the classroom is oriented toward examination materials; therefore, the quality of the tests has, automatically, impact on teaching and the methods teachers employ for an effective instruction. Alderson & Wall [2] maintain that the tests and the assessments are considered necessary for encouraging the learners and lead to effective change in teaching methodology and curriculum design along with learners' motivation to study more, they are believed to have positive effect on education.

The nature of washback has not been discovered yet. What really causes the effects has not become clear thus far. But what is obvious is that teachers based on the results of the tests make decision about classrooms activities and make efforts to provide appropriate instructions regarding the learners' needs to satisfy educational requirements.

Bailey [5] suggests language learning goals, authenticity, learner autonomy and self-assessment, and detailed score reporting as four factors which are thought to promote beneficial washback according to the literature. But she points out that more studies are needed to prove these claims. In another study in 1999 she puts emphasis on score reporting for attaining beneficial washback and believes that the result must be described in non- technical terms to be interpretable and understandable by everybody. Following Bailey, Gattullo [15] believes previous papers play an important role in developing subsequent tests, thus providing detailed results about the test is more beneficial than providing just a single score on learners' performance by merely correcting the papers. There is little empirical research about detailed score report by teachers or testers to improve language skills by learners, thus as Bailey [6] clarifies, further investigation is necessary and more attention is needed to be paid to this factor than what it is available with current tests.

Preparation for the test is making the students ready for the exam in any way that familiarizes the learners with the test format and materials through teaching. Detailed score reporting as one of the way of the strategies employed by the teachers and/ testers to prepare the students for the test is assumed to promote beneficial washback effect, but this claim has little available empirical research in the literature.

According to Marquardt and Gillam [23], in dynamic assessment what is common is that students are provided with explanations about the expected responses in the exam, so the very crucial and detailed points about the questions of the exam will be clarified for the students. Mediations in dynamic assessment can range from simple hints to detailed explanations; therefore which one of these strategies may lead to positive consequences on learning is one of the salient issues of discussion.

Although detailed score reporting and preparation before the examination is assumed to have positive consequences, it is a very strong claim to believe that this testing strategy definitely promotes washback potential. Due to the dearth of studies observed on DA and its various strategies employed in the mediation between the tests in teaching phase which may lead to positive washback effect, the current study is conducted to examine this noteworthy phenomenon. The present study combines, for the first time, two areas of testing (dynamic assessment & washback) and explores their influence on learning a foreign language. Applying three stages, pre-teaching, teaching and post-teaching, the aim is to specify the effectiveness of DA to instruction and learning in assessment situations and then to determine whether detailed score reporting as one of the strategies employed in teaching phase of this type of assessment, has positive effect on learning.

The basic objective of this research study is to find answers to the following questions:

1) Is intervention (teaching strategies) employed in teaching phase effective in terms of learning to result from Dynamic Assessment?

2) Does detailed score reporting as one of the strategies employed in teaching phase associated with DA have positive washback effect on learning?

2. METHOD

The research design adopted in the study is experimental and, given the nature of the study, the test–teach–retest paradigm has been adopted. This paradigm is consistent with the intervention procedure, which is characteristic of DA and is required for an effective instruction, which is, in turn, based on mediated learning experience (MLE) having to do with providing students with information about their learning. Regarding the second question of the study the very design used was true-experimental: the pretest-post-test control and experimental group design to specify the effect of the strategies used in the given intervention procedure.

3. PARTICIPANTS AND SAMPLING PROCEDURE

For the purposes of the study, out of language institutes in Shiraz, one institute was selected randomly. From among the language classes which were believed to be at the intended level of proficiency, based on the placement tests administered by the institute, two classes were chosen. In order to ensure that the participants were indeed at the proficiency level intended in the study, and that they were homogenous in terms of their proficiency, Oxford Quick Placement Test, which has been pre-tested and validated by 19,000 students, or so in 60 countries and is extensively used throughout the world, was administered. Based on the results of the test, the outliers were left out and 30 students were chosen as the participants of the study.

4. INSTRUMENTS

The instruments used to collect the data were versions of the Speaking Test of Pearson Test of English General (PTEG), formerly known as the London Tests of English, designed to assess and accredit general English language proficiency. They are especially developed to tap the non-English speakers' oral ability in English. Altogether, tests (pre and post-tests) included three sections: monologues (to assess ability to speak about matters of personal information and interest), picture description and role play. The test takers were scored on how well they performed over the three speaking sections against the following criteria: fluency, accuracy, interaction, range (a very basic repertoire of words and simple phrases) and phonological control.

Although the reliability and validity indices of the tests have been reportedly confirmed in Paltridge and Starfield [25] and Taylor and Geranpayeh [32], it was decided to establish the reliability of the tests for the purposes of the current study. In order to accomplish this, the tests were piloted on two different groups representative of the target population of interest and their reliability indexes were established through Cronbach Alpha. The piloting was carried out on two groups each comprising of 15 students. The reason for piloting on two different groups was that it was difficult for the test takers to sit all the tests for a long time. Moreover, this could have negatively affected their performance. The piloting in all lasted about 150 minutes; it took each test taker 5 minutes to answer the given questions. The reliability report of the tests is given in the following table.

Table 1. Reliability report of the tests used in the study.

Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
0.84	0.77	0.78	0.80	0.80	0.87

According to the table above, the reliability index ranges from 0.77 to 0.87, which, according to Nunnally [24], represents acceptable indices of reliability.

5. PROCEDURE

A test was first administered to determine the proficiency level of the participants. Based on the results, 30 students were chosen as the participants of the study. The participants were given the test individually, their speech recorded and scored later; the highest possible score was 25. The problems which the participants had in their speaking were noted down by the examiner and were later discussed with them in a group session. In order not to disrupt their flow of speech it was not indicated to them what their problems were at the time of exam. During the instruction, explanations provided for the test-takers were provided in abstraction from the real test materials and outside the testing situation. In so doing, after individual testing, the examiner reviewed the problems which they had in a group session with all the test-takers attending. Specific references were made to the problems in their oral performance. They were also briefed on ways, or rather strategies (to use a technical term in literature on dynamic assessment) of tackling the problems. The strategies provided ranged from simple and standardized hints to complete explanations of linguistic rules. Consistent with the approach adopted and the relevant theoretical background, the mediation provided was in the form of prompts, hints, and leading questions arranged in a hierarchical manner, from implicit to explicit. Individual oral testing took place 5 minutes and the group session (teaching phase or mediation) 10 minutes.

In the post-tests, just the tests were administered to the participants and their speech was recorded and later scored. No mediation or help was provided whatsoever prior to, during or following the post-tests for them.

They were thus given a series of pre- and post-tests in terms of which they were repeatedly compared. The materials used between successive tests were different, however. The reason for repeated pre-tests, post-tests and comparisons was two-fold. The first reason was that in testing practices consistent with Dynamic Assessment, teaching and testing go hand-in-hand and one cannot say where one should stop and the other begins. The second reason lies in the conceptualization of learning in this situation. Among other things, in DA, the eventual attainment is not of immediate interest. Rather, the mediating stages of learning are more important. Therefore, we were interested in the gradual development, as revealed through successive post-tests. It took 3 weeks to provide the instructions through DA.

Regarding the second question of the study, following the above mentioned experiment, the participants were randomly assigned to two groups (i.e., control and experimental). A pre-test of speaking was administered to both groups to ensure that they began the experiment on an equal footing. The time allotted to the pre-test was 5 minutes and the highest possible score was 25. Having established that there was no statistically significant difference between the two groups in terms of their oral ability, through the respective t-test, we could begin the experiment. The control and experimental groups received different instruction strategies. A few days later, preparing a taxonomy of the errors, the scores of the pre-test were reported in detail to half of the examinees (15 students) as experimental group. In this stage of the study, the control group received other strategies other than detailed explanations. For the experimental group information of individual parts of the tests was interpreted so that test takers can be informed about their very mistakes. The other half, control group, received no detailed report on scores. After several days, based on the materials provided and in consistent with the pre-test, a post-

test was administered to both control and experimental groups in order to find the effect of this description of the scores on learning.

In both phases of the study, following the instructions, the learners were tested. Their oral performance was recorded and later scored. In order to ensure the consistency of scoring the oral performance of the participants, five recordings were first randomly chosen and independently scored by the examiner and a research assistant, carefully basing the scoring on the yardsticks to be applied to scoring. In the initial scorings, there was 83% agreement between the two scorers. Given that the level of agreement was not high enough, the cases of disagreement were discussed and agreed upon, followed by a second round of scoring of five different randomly chosen recordings. Given that the level of agreement reached 94%, the examiner decided to score the rest of the recordings.

6. DATA ANALYSIS

For the purpose of analyzing the data, in addition to descriptive statistics, a series of t-tests were run. First, paired sample t-tests were run to monitor the progress of the participants through successive tests. Moreover, independent sample t-tests were run to compare the two groups in terms of learning to result from the form of the mediation. Finally, Effect Size, which is simply a way of quantifying the possible differences between two groups, was also determined. Specifically, Effect Size (ES) analysis was carried out to compare the effectiveness of the strategies used in the teaching phase of dynamic assessment.

7. RESULTS

In the current study, we were primarily concerned with investigating, besides the contribution of dynamic assessment to learning, the effectiveness of different strategies in DA. The major interest was to find out whether there is different learning to result from applying different strategies in Dynamic Assessment. In order to do that, it was necessary to monitor the progress of the participants through successive tests administered. Given the nature of the correlations to be made, paired sample t-tests were run to see how the participants did over the successive tests administered.

The analysis of the data indicates that overall, there has been a trend of progress from one test to another for the group in question. To be more specific, the trend is such that the higher the pre-test score, the higher the respective post-test score. In order to see whether the differences between the pre-test scores and the corresponding post-test scores are statistically significant, a number of t-test were run. It could be seen that in all the cases, the differences are significant. This is clearly indicative of the contribution of teaching (mediation) in dynamic assessment to learning. Table 2 is the representative of the correlations between pre- and post-test 1.

Table 2. Correlations between Pre- and Post-test 1.

N	Correlation	Sig.
30	0.830	0.000

Correlation analyses showed significant correlations. That is to say, pretests and posttests scores were highly correlated. (r> .800). Table 3 provides the descriptive statistics after the first intervention (mediation).

Table 3. Descriptive statistics for the performance on Pre- and Post-test 1.

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	17.4667	2.88556	0.52683
Post-test	30	21.5000	2.88556	0.52683

As could be seen, the mean score of the post-test (21.5000) is higher than that of the pre-test (17.4667). The participants outperformed in the first post-test. To see whether the difference is statistically significant, a paired sample t-test was run.

Table 4. Paired samples t-test for the performance on Pre- and Post-test 1.

Paired Sample Test							
Paired Differences					T	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-4.03333	1.71169	0.31251	-4.67249	-3.39418	-12.906*	29	0.000

* The difference is significant.

The table above clearly indicates that the difference between the two tests is statistically significant

(t=-12.906, p<.05). Given the mean values, it could be seen that the participants outperformed in the first post-test. I
second phase of instruction for them. Then, they were given a post-test.

Table 5. Correlations between Pre- and Post-test 2.

N	Correlation	Sig.
30	0.903	0.000

Like the correlation analysis for pre and post-test 1, correlation analyses in this phase of the study, also, showed significant correlations (r> .900).

Table 6. Descriptive statistics for the performance on Pre- and Post-test 2

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	18.0333	3.11264	.56829
Post-test	30	21.2000	2.70886	.49457

As it is observed from the table, the mean score of the post-test (21.2000) is higher than that of the pre-test (18.0333). A t-test was also run to see whether the difference between the mean scores is statistically significant.

Table 7. Paired samples t-test for the performance on Pre- and Post-test 2.

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-3.16667	1.34121	.24487	-3.66748	-2.66585	-12.932*	29	.000
* The difference is significant.							

Given the results of the table above, enhancement could be seen in terms of the participants' performance on the second post-test and the difference is statistically significant ($t=-12.932$, $p<.05$).

This phase of the study was followed by four more phases. The general procedure was the same, though the materials differed from one phase to another. The participants received instructions consistent with dynamic assessment procedure. Four more phases of the study were followed by four more post-tests, the results of which appear, for reasons of space, in Appendix A. Relatively speaking, in all phases of the study, the participants outperformed after the teaching phase and the differences were statistically significant. Moreover, given the mean values, a trend of steady progress could be seen.

The results of the successive post-tests reported were indicative of the steady progress and better performance of the participants who were provided with the materials used for the purpose of instruction in line with dynamic assessment procedures.

The findings suggest that dynamic assessment could indeed benefit language learning. This has been apparent in the steady progress of the participants over the period in which the current experiment was in progress.

As noted earlier, almost all the students benefited from the mediation and were able to apply the acquired strategies to the new speaking tests. The strategies ranged from simple hints and prompts to detailed explanation the employment of which depends on learning potential of the learners. The main question was that, in general, 'is there any difference between these strategies in providing a beneficial condition for the learner to benefit from the mediation and obtain the optimal information'? To this end, the learners were divided into two groups. Experimental group received detailed score reporting as the strategies to the mediation. The analysis of the data indicates that there has been a progress from pre-test to post-test in both groups. In order to see whether the differences between the pre-test scores and the corresponding post-test scores are statistically significant, paired sample t-tests were run. It could be seen that in both cases, the differences are significant. Mean, standard deviation and correlation of the tests before and after the treatment in both groups, also the results of the Paired Samples Test are indicated in the Appendix B.

Given the main focus of the study, which was differentiating between the teaching strategies in dynamic assessment, the two groups were compared through relevant descriptive and inferential statistics. Table 8 provides the descriptive statistics for experimental and control groups' performances before the experiment.

Table 8. Descriptive statistics for control and experimental groups' performance before the experiment.

Group statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Experimental group	15	17.6000	2.38447	.61567
Control group	15	17.1333	2.44560	.63145

As could be seen, the mean score of the experimental group and that of control group are approximately the same. To see whether the participants start the experiment on an equal footing, and the difference is not statistically significant an independent sample t-test was run.

Table 9. Independent sample t-test for control and experimental groups' performance on pre-test.

Independent Samples Test									
Levene's Test for Equality of Variances			t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	upper
Equal variances assumed	.014	.907	.529*	28	.601	.46667	.88192	-1.33986	2.27319
Equal variances not assumed			.529	27.982	.601	.46667	.88192	-1.33991	2.27324

* The difference is not significant.

Based on the table above, the difference between the two groups is not statistically significant (t=.529, p>.05). Table 10 provides the descriptive statistics for experimental and control groups' performances after the experiment.

Table 10. Descriptive statistics for control and experimental groups' performance after the experiment.

Group statistics				
	N	Mean	Std. Deviation	Std. Error
Experimental group	15	22.8000	1.47358	.38048
Control group	15	19.2667	2.28244	.58932

As could be seen, the mean score of the experimental group (22.8000) is higher than that of the Control group (19.2667). In spite of the fact that there was no major difference between the two groups prior to the mediation, the experimental group outperformed the control group in the post-test. To see whether the difference is statistically significant, an independent sample t-test was run.

Table 11. Independent sample t-test for control and experimental groups' performance on post-test.

Independent Samples Test									
Levene's Test for Equality of Variances			t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	upper
Equal variances assumed	4.275	.048	5.037*	28	.000	3.53333	.70147	2.09643	4.97023
Equal variances not assumed			5.037	23.943	.000	3.53333	.70147	2.08538	4.98128

* The difference is significant.

Based on the table above, the difference between the two groups is statistically significant (t=5.037, p<.05) suggesting the better performance of the experimental group compared with the control group (effect size= 1.54804).

8. DISCUSSION AND CONCLUSION

The goal of our research program was to examine how and to what extent DA works in EFL contexts and learning situations. In summary, this study indicated that selected dynamic testing procedures can be effective in obtaining information on students' learning potential.

The particular dynamic approach adopted in the current study is based on interventionist approach to DA and the theories of change and cognitive modifiability which necessarily underlie methods of assessment.

As to the findings, it was empirically verified that after each treatment in all phases of the study, a significant progress observed in students' performance level and higher scores were obtained. Paired t-tests were used to investigate whether learners had made score gains following the mediations. The results indicated that a significant gain in speaking scores had indeed occurred on all six posttests. These findings are indicative of the highly beneficial role of using strategies during dynamic assessment as far as the intervention approach is concerned and the clear advantage of this assessment over the static one. Some other studies such as Bethge, Carlson, & Wiedl [9], Berman and Graham [8], Guterman [17], Kozulin and Garb [22] and Tzuriel [33] found that dynamic assessment could noticeablly benefit EFL learning.

In order to examine the assumption of the positive washback effect of the detailed description of the scores employing paired t-test and independent samples t-test, the difference between the performance of experimental (learners received detailed score reporting) and control groups (learners received no detailed score reporting, but just simple hints) was indicated. According to the results, score gains were made in both groups in posttest. It was confirmed that this strategy, detailed description of the results, can work beneficial and help the learners to be aware of their very mistakes and concentrate on depth of their learning. Analysis of differences by t-test and the degree of effect size showed that the difference between two groups reached a significance level after the experiment. Greater change was made from pretest to posttest in experimental group.

The larger mean of the experimental group indicated that the given teaching strategy, which was washback effect-based treatment, has been more effective in improving the speaking performance of the learners. The findings of the study revealed that providing detailed results about the test has positive washback effect on learners than reporting just a single score without any information about the errors due to the fact that individual parts of the tests and very detailed points about the exam are clarified by the teacher.

Language teaching and testing are integrated in educational system [11, 14, 19, 22, 27, 28, 29]. As mentioned before, DA relates teaching to testing and through providing interventions and modifications helps the teachers and the testers to recognize the ability of the individuals and compensate their learning problems. Instruction and assessment work simultaneously; teachers discover where learners problems lie and with regard to that provide support to overcome such problems. DA is integrated with instruction and mediation, the instruction that is incorporated into the dynamic assessment. Results obtained in the present study thus reflect not only the students' learning potentials but also the quality of mediation provided during the assessments. Thus, it is clearly evident that the same results may not be obtained when

mediation is provided by different mediators. The quality of mediation definitely influences the extent to which progress is made in performance level.

According to Bethge, Carlson and Wiedl [9], DA, due to its teaching phase, reduces test anxiety and brings about better performance. Other researchers [7, 8, 10, 12, 13, 17, 29, 33], also, put emphasis on elaborated mediation in DA.

Language is not an abstract system, but it is a social and inter-mental activity which takes place in ZPD, in other words, learning language is an ongoing process and takes place in collaborative dialogues not outside it. We can combine the information from Vygotskian approach with feuerstein's views to examine how DA works. According to research supporting the utility of dynamic assessment, it has a better perspective value than static tests. Dynamic assessment (DA) offers a conceptual framework for teaching and assessment, not just assessment. Its teaching phase aims at understanding individuals' abilities and promoting their development. DA, applying mediated intervention, better predicts the precise learning potential than static tests. The comparison based on the results obtained from aided and independent performance has clarified the point.

It takes on greater importance when incorporated to help identify explanatory variables accounting for washback. Indeed, the clearest evidence of washback on student performance emerges from the second question of the current study, where it appears to be linked to the mediation of exam using a strategy consisted of a series of very specific points. The issue of detailed score reporting which is one of the related strategies of dynamic assessment is widely held in the available literature, to promote beneficial washback. And yet, relatively little empirical research has been conducted to substantiate this claim. According to Bailey [5, 6], it may be one of the factors which bring about positive washback effect. This study aimed at proving this assumption.

The consequences of detailed score reporting can be beneficial for both learners and teachers; as the former try to make use of the points and be more prepared for the later tests, the latter employ the given strategy and are more informed about the students' abilities and their learning potential, therefore, make effort to take into consideration all these precise abilities on the basis of the errors made by the learners on the test.

According to many linguists, there is a one to one correspondence between teaching and testing. Currently testing comes before teaching and learning process. The influence of tests on teaching and learning is an issue of consideration. Results must be described in non- technical terms to be interpretable and understandable by everybody to be able to make use of them and show a better performance after taking care of them and concentrating on them precisely.

Saif [30] puts emphasis on the process before test design to attain a considerable benefit from the test. Teachers may find close relationship between test preparation and preferred results. Alderson and Hamp-Lyons [1] state that, teachers' preferred teaching methods and activities before the test in order to bring about beneficial results help students to find a logic for their learning and lead to improved test performance. These authors believe that explanation will cause internalization. While in an analysis by Kolahi [21], classifying the mistakes, made by the students in writing examination, caused a more desirable instructional program which resulted in improvement in learning outcomes, lack of positive washback to the learners when

they are prepared for the test in advance was found in the study by Green [16]. In his analysis he finds no correlation between writing test preparation and better achievement. He contends that a number of factors such as age, social class, first language, personality, motivation and cognitive style interfere with the direct influence of other factors. Generally, according to the results gained from his investigation, teaching to the test is not preferred over teaching the targeted skills.

Interpretation of assessment result in detail possibly helps positive washback; therefore, instead of reporting a general score, interpreting assessment information of individual parts of the tests may be more beneficial. Quick interpretation of a test result does not provide details related to the learners' answers, hence no precise information about their errors in a test. Detailed reporting can have important consequences for teachers as well as learners; moreover, teachers prepare students for the test in future effectively with regard to their specific errors on the tests. According to Gattullo [15], past papers play a crucial role in producing tests, thus providing specific results about the test is superior than reporting just a single score on a unidimensional scale; he addresses this process interaction with the results, not merely correcting them.

Bailey [6] believes that the result must be described in non-technical terms to be interpretable and understandable by everybody.

Specifying test as positive or negative phenomenon is not a simple task and some researchers [1, 4, 16] put emphasis on unpredictability and indirectness of washback effects, therefore, more investigation is required. Alderson and Hamp-Lyons [1] make this point explicit when they state that a test by itself does not have positive or negative washback, but the amount and type of washback on teachers and learners varies with respect to the type of the test. As Bailey [5] notes “there are also concerns about what constitutes both positive and negative washback, as well as about how to promote the former and inhibit the latter” (p. 259). We suggest that we need to look closely at classroom events in particular, in order to see what exactly is involved in the students’ performance.

From the results of the two-phase study reported above, it seems clear that providing support after or during the administration of the test, no matter the form, definitely influences the students’ performance. Conclusion reached is that providing the more specific and detailed information about the test by the teachers or the testers along with an appropriate condition for concentration on information provided lead to better performance level on behalf of the learners; it was found out that it is the strategies used and the support provided by the teachers or testers in test procedure that causes beneficial effect on learning. In order to describe the results of the test in detail, formative/ diagnostic assessment, which is an indication of dynamic assessment, in principled ways is needed.

It is wise to bear in mind that that the goal of teaching and testing a foreign language is to prepare the learners for communicative purposes, therefore, there is a must for the teachers and the testers to look for innovative and useful ways to help the students find out their abilities, and not to neglect the fact that learners need to be oriented to such efficient ways and consequently make sure that the needs are met.

DA is one of the innovations used in educational situations to make optimal use of teaching and testing, however, thesuccess of this type of assessment requires a continual monitoring of learners' processes. From this perspective, effective instruction involves a continual dynamic assessment and subsequent modification of instructional strategies that can help students learn.

The findings might have implications for teaching and testing procedures to be consistent with the learners' needs and goals. As DA works in educational situations effectively and the teachers and the testers treated with respect to the approaches of DA benefit from the mediations and interventions, the application of this type of assessment is offered to the instructors and the test givers. The findings of this study may be a step to help the learners to achieve better in learning process and teachers and assessors to select more appropriate ways to modify their methodologies to better serve the requirements of the educational system.

Studies such as those described in this study convince us that students may benefit considerably from new approaches to instruction. We hope that this study has provided a beginning, and an incentive to explore further. There are numerous opportunities for future research on language teaching and testing and due to the lack of clear description of the nature of DA, more studies are required to identify the promise of this type of assessment.

Remarkably little research has traced the development of dynamic assessment and particularly its teaching strategies. Based on the available literature on DA, the researchers suggest that further research on it be needed. The findings lead to re-evaluating teaching and testing methodologies and making attempt to modify them to facilitate the developmental process.

REFERENCES

[1] Alderson, C.J., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. Language Testing, 13, 280-297.
[2] Alderson, C.J., & Wall, D. (1993). Does washback exist? Applied linguistics, 14, 115-129.
[3] Amengual-Pizarro, M. (2009). Does the English test in the Spanish university entrance examination influence the teaching of English? English Studies, 90, 582–598.
[4] Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. System, 30, 207–223.
[5] Bailey, K.M. (1996). Working for washback: A review of the washback concept in language testing. Language Testing, 13 (3), 257-279.
[6] Bailey, K.M. (1999). Washback in language testing. Princeton, New Jersey: Educational Testing Service.
[7] Barrera, M. (2003). Curriculum-based dynamic assessment for new- or second-language learners with learning disabilities in secondary education settings. Assessment for Effective Intervention, 29 (1), 69-84.
[8] Berman, J., & Graham, L. (2002). School counselor use of curriculum-based dynamic assessment. Australian Journal of Guidance and Counseling, 12(1), 21–40.
[9] Bethge, H. J., Carlson, J. & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven matrices performance, visual search behavior, test anxiety, and test orientation. Intelligence, 6, 89-97.
[10] Bolig, E. E. & Day, J. D. (1993). Dynamic assessment and giftedness: The promise of assessing training responsiveness, Roeper Review, 16 (2), 110-113.

[11] Boyle, J., & Fisher, S. (2007). Educational Testing. A competence-based approach. Oxford: BPS Blackwell.
[12] Carlson, J.S., & Wiedl, K. H. (1992a). Principles of dynamic assessment: The application of specific model. Learning and Individual Differences, 4, 153-166.
[13] Carlson, J. S., & Wiedl, K. H. (1992b). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), Interactive assessment (pp. 167–186). New York: Springer-Verlag.
[14] Delclos, V. R., Vye, N. J., Burns, M. S., Bransford, J. D., & Hasselbring, T. S. (1992). Improving the quality of instruction: Roles for dynamic assessment. In H. C. Haywood & D. Tzuriel (Eds.), Interactive assessment (pp. 317-331). New York: Springer.
[15] Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: An Italian case study. Language Testing, 17 (2), 278–288.
[16] Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. Assessment in Education, 14, 75–97.
[17] Guterman, E. (2002). Toward dynamic assessment of reading: Applying metacognitive awareness guidance to reading assessment tasks. Journal of Research in Reading, 25(3), 283–298.
[18] Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. System, 28, 579-591.
[19] Haywood, H.C. & Lidz, C.S. (2007). Dynamic assessment in practice. Cambridge: Cambridge University Press.
[20] Haywood, H. C., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. Peabody Journal of Education, 77(2), 40–63.
[21] Kolahi, S. (2007). Investigating the washback effects on improving the writing performance of Iranian EFL university student.In: The Second Biennial International Conference on Teaching and Learning of English in Asia : Exploring New Frontiers (TELiA2), 14-16 June 2007, Holiday Villa Beach & Spa Resort, Langkawi. Faculty of Communication and Modern Languages, Universiti Utara Malaysia, Sintok, pp. 1-13. ISBN 978-983-42061-2-3.
[22] Kozulin, A. & Garb, E. (2002). Dynamic assessment of EFL text comprehension of at risk students. School Psychology International, 23, 112-127.
[23] Marquardt, T.P., & Gillam, R.B. (1999). Assessment in communication disorders: Some observations on current issues. Language Testing, 16 (3), 249–269.
[24] Nunnally, J. C. (1978). Psychometric Theory. McGraw-Hill Book Company, pp. 86-113, 190-255.
[25] Paltridge, B., & Starfield, S. (2013).The Handbook of English for Specific Purposes. Oxford: Wiley Blackwell.
[26] Pearson Education Asia Limited. (2011). The official guide to PTE – Pearson test of English general. Hong Kong: Pearson Longman Asia ELT.
[27] Poehner, M. E. (2005). Dynamic assessment of oral proficiency among advanced L2 learners of French. Unpublished doctoral dissertation. The Pennsylvania State University.
[28] Poehner, M. E. (2008). Dynamic assessment: A Vygotskian approach to understanding and promoting second language development. Berlin: Springer.
[29] Poehner, M. E. (2009). Group dynamic assessment: mediation for the L2 classroom. TESOL Quarterly, 43, 471-49
[30] Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. Language Testing, 23, 1-34.
[31] Spolsky, B. (1997). The ethics of gatekeeping tests: What have we learned in a hundred years? Language Testing, 14, 242-247.
[32] Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalizing the test construct. Journal of English for Academic Purposes, 10, 89-101.
[33] Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. Educational Psychology Review, 12(4), 385-435.

[34] Vygotsky, L. S. (1986). Thought and language. Cambridge, MA: MIT Press.
[35] Vygotsky, L.S. (1994). Principles of social education for deaf and dumb children in Russia.
[36] In The Vygotsky Reader. R. Vander Veer & J. Valsiner (Eds.). Cambridge, MA:Blackwell.

APPENDICES

Appendix A

Table 1. Descriptive statistics for the performance on Pre- and Post-test 3

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	18.7333	2.77841	.50727
Post-test	30	21.7667	2.11209	.38561

Table 2. Correlations between Pre- and Post-test 3

N	Correlation	Sig.
30	.941	.000

Table 3. Paired samples t-test for the participants' performance on Pre- and Post-test 3

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-3.03333	1.06620	.19466	-3.43146	-2.63521	-15.583	29	.000

Table 4. Descriptive statistics for the performance on Pre- and Post-test 4.

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	18.5000	2.37443	.43351
Post-test	30	21.9667	2.04237	.37288

Table 5. Correlations between Pre- and Post-test 4.

N	Correlation	Sig.
30	.871	.000

Table 6. Paired samples t-test the participants' group's performance on Pre- and Post-test 4

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-3.46667	1.16658	.21299	-3.90228	-3.03106	-16.276	29	.000

Table 7. Descriptive statistics for the performance on Pre- and Post-test 5

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	18.6667	2.33907	.42705
Post-test	30	22.2667	1.77984	.32495

Table 8. Correlations between Pre- and Post-test 5.

N	Correlation	Sig.
30	.917	.000

Table 9. Paired samples t-test for the participants' performance on Pre- and Post-test 5

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-3.60000	1.00344	.18320	-3.97469	-3.22531	-19.650	29	.000

Table 10. Descriptive statistics for the performance on Pre- and Post-test 6.

	N	Mean	Std. Deviation	Std. Error
Pre-test	30	18.9000	2.02314	.36937
Post-test	30	22.6000	1.47625	.26952

Table 11. Correlations between Pre- and Post-test 6.

N	Correlation	Sig.
30	0.817	0.000

Table 12. Paired samples t-test for the participants' performance on Pre- and Post-test 6.

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-3.70000	1.17884	.21523	-4.14018	-3.25982	-17.191	29	.000

Appendix B

Table 1. Descriptive statistics for experimental group.

	N	Mean	Std. Deviation	Std. Error
Pre-test	15	17.6000	2.38447	.61567
Post-test	15	22.8000	1.47358	.38048

Table 2. Correlations between Pre- and Post-test for experimental group.

N	Correlation	Sig.
15	.850	.000

Table 3. Paired samples t-test for experimental group.

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-5.20000	1.37321	.35456	-5.96046	-4.43954	-14.666	14	0.000

Table 4. Descriptive statistics for control group.

	N	Mean	Std. Deviation	Std. Error
Pre-test	15	17.1333	2.44560	.63145
Post-test	15	19.2667	2.28244	.58932

Table 5. Correlations between Pre- and Post-test for control group.

N	Correlation	Sig.
15	.953	.000

Table 6. Paired samples t-test for the control group.

Paired Sample Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
-2.13333	.74322	.19190	-2.54492	-1.72175	-11.117	14	.000