
INVESTIGATION OF THE IMPACT OF COMPUTERIZED DYNAMIC ASSESSMENT ON LISTENING COMPREHENSION TESTS BY ARAB POSTGRADUATE EFL STUDENTS AT UUM

Lina Lafta Jassim

College of Art, University of Thi Qar

Iraq

Lnar83@gmail.com

Abstract

The current study aimed to answer the following questions: Is there any statistically significant difference between the test-takers actual and mediated scores in the listening test? Can computerized dynamAssessmentent reveal the test-takers LP in the listening question types? Is there any statistically significant difference between the number of hints in the Question types across the levels of the listening ability?.48 Arab postgraduate EFL students who attended an intensive English course at UUM were participants in this study? To mediate the test-takers with hints to process the listening questions, computer software was developed, and it produced three types of scores: Actual, negotiated, and learning potential. The current study's findings emphasized the significant differences between the actual and mediated scores with various listening ability levels in almost all the question types. Generally, the results indicated that computerized dynamAssessmentent had a significant and positive effect on the improvement of proficiency of Arab EFL learners on the dialogue and monologue tasks. Using dynamAssessmentent was recommended by teachers, incredibly computerized dynamAssessmentent in particular, as the

information gained from this assessment mode empowers teachers to provide learners with more individualized and accordingly more effective teaching and assessment techniques and strategies in a sociocultural context.

Keywords: computerised, impact, assessment arab, postgraduate etc

Introduction

This study put forward an alternative assessment approach and its relationships with learning based on Vygotsky's Socio-Cultural Theory (SCT) of mind (1978), as well as the concepts of computerized dynamAssessmentent (C-DA) in the TOEFL iBT listening test. DynamAssessmentent (DA) questions the classical perceptions about teaching aAssessmentent by mergiAssessmentent and instruction into a seamless task where various forms of support and mediation are undertaken to unveil the cognitive and metacognitive strategies of the test-takers (Lidz&Gindis, 2003). The other challenges that affect the association between teaching and aAssessmentent lie within the mediators' lack of familiarity with DA

conceptions and a myriad of theories underpinning how DA and C-DA are carried out (Poehner, 2008). Therefore, teachers most often find themselves at a loss to take the challenges of developing suitable assessment tools, implementing processes, and inferring outcomes (Torrance & Pryor, 1998). Instead, they employ several practices and tasks such as cloze tests, group assignments, and tests but with a theoretical understanding of how such assessment methods are implemented to inform the test-takers ability. Such traditional approaches to Assessment are more likely to yield wrong inferences about the test-takers ability, especially when they emanate from a fuzzy overview of the construct, such as listening comprehension.

Although the washback EffectEffect explores the influence of Assessment on teaching, DA practitioners reverse this association by allocating more importance to teaching. That is, to connect Assessment to education, the assessment processes should emerge from a well-established inquiry of the instructional activities and educational performances as carried out in the classroom (Poehner, 2008) to enable teachers to be engaged in a more active role in ascertaining the relevant assessment practices to the learners' potential. Teachers should not limit their testing methods, nor should they test the test-takers performance with a single final achievement test.

Recently, there has been a lot of research on online Assessment and how it is used to promote and facilitate learning. For instance, (Wang, 2008) recorded the

online leading role learners can have without the mediators' presence. From a Vygotskian perspective, this sociocultural and active e-learning context is given its due momentum. Both learning and Assessment are seamlessly merged, using software to put forward hints based on the learners' answers. Vygotsky's theory promotes the process rather than the product to understand learning and development where "the potential development varies independently of actual development, meaning that the latter, in and of itself, cannot be used to predict the former" (Lantolf & Thorne, 2006, p. 328).

Review of the Literature

DA applies Vygotsky's SCT of mind Assessment in language learning, offering new language classroom insights and information to improve interventions. For Grigorenko (2002), DA is changing the assessment procedures and shifting toward a new assessment philosophy that focuses on the role of intervention in helping individuals develop. Lidz (1995) praised the DA test-intervention-retest format for altering the test-takers learning behavior. For Poehner (2007), DA is an ongoing and contextualized activity that engages the learners to unveil their underlying potential to change their learning behavior.

Vygotsky's SCT of mind perceives learning as a social process, and that human intelligence generates in a sociocultural context where mutual interactions become fundamental in

cognitive development. He believed everything is learned by interacting with others and integrating into the individuals' mental structure. The Zone of Proximal Development (ZPD) is determined by the learners' ability to benefit from mediation; wheassessmentent assesses this size and describes the learners' changing ability to learn with assistance. The collaboratiAssessmentent of the learners' abilities is a predictor of their functioning rather than a measure of independent performance. Therefore, the ZPD is assessed within this shared activity. The assisted version represents the learners' maturity of psychological functions where mediation applies traditional artifacts, notions, and accomplishments (Lantolf & Thorne, 2006) to regulate the individuals' mental and social activity. However, when the learning experience is not mediated, learners might face tremendous difficulties coming to grips with the learning reality (Feuerstein & Feuerstein, 1991).

The C-DA form used in this study was conducted through a computer where learners received online mediation through sets of prefabricated hints. According to Poehner and Lantolf (2013), the potential learning score (LPS) considers the difference between the learners' actual and mediated scores, using this formula: " $LPS = (2 * mediated\ score - actual\ score) / maximum\ score$ " (Kouzlin & Grab, 2002). The mediated score is the culmination of the score emanating from mediation, a software program, or a human mediator (Poehner & Lantolf, 2013). In the current study, the mediated score is computed by deducting the hints' frequencies from the total score of each item. For example, if

the item's total score was three (3), and the learner used two hints to get the correct answer, the mediated score would be one (1). The hints and prompts were gradually organized from implicit to explicit to make the learning ability as malleable as possible.

Similar to DA, C-DA is grounded in the theoretical framework of Vygotsky (1978/1986). With the development of various technological tools, Dixon-Krauss (1996) recommended their application to realize Vygotsky's vision of designing lessons in a way that simplifies teaching. In interventionist DA, "a prefabricated and fixed set of clues and hints is determined in advance and offered to learners as they move through a test item by item" (Poehner & Lantolf 2010, p. 318). C-DA provides a more in-depth discovery of the learner's capabilities (Tzuriel & Shamir, 2002). It can additionally act as a classroom teacher to mediate the learners in their ZPD (Crook, 1991).

Defining Listening

Listening is extensively used (Scarcella & Oxford, 1992) in second language learning, yet it is the least explicit skill (Vandergrift, 2004). It holds a key role in communication (Mendelsohn, 1994) as "an active and conscious process where the listener constructs meaning by using cues from contextual information and existing knowledge while relying upon multiple strategic resources to fulfill the task requirement" (O'Malley, Chamot Kupper, 1989, p. 19). Interpreting the listening input is contingent upon "the cognitive environment of the listener" (Buck, 2001,

p. 29). This makes listening to an active and inferential process (Buck, 2001; Rost, 2002), and for the listening message to be decoded, it has to rely on the prior and linguistic knowledge (Underwood, 1989; Vandergrift, 2007). Vandergrift (2007) stated that listening is at the heart of learning any language. It is a problematical active process where the test-taker must distinguish sounds, understand stress and intonation, recognize vocabulary and grammatical structure, and relate them to a particular context. Two cognitive processes are involved in listening: Top-down and bottom-up. While the former is about interpreting meaning using background knowledge or schemata, the latter linearly involves generating sense from the smallest to the largest spoken unit (Nunan, 1998). In the current study, listening is operationally defined as the test-takers ability to get the correct answer using the predetermined hints in an online environment. The mediated score only represents this ability.

The impact of DA on language skills, such as listening, has been investigated throughout the last two decades. For instance, Ableeva (2008) implemented DA and concluded that the difficulties in learning French uncovered the learners' unique ZPD, unlike the case with the non-dynamic pre-test. Likewise, Alavi, Kaivanpanah, and Shabani (2011) contended that group DA prepared for collaboration and interaction by highlighting standard practices. Additionally, the Author (2014) developed a DA listening test using static and DA approaches and concluded that DA was

conducive to understanding the cognitive and metacognitive listening processes. Ghahremani (2013) tackled the Effect of summative, formative, and DA on learning listening and recorded that the learners in the dynamic group outperformed the other groups. In addition, Emadi and Arabmofrad (2015) showcased a comprehensive account of interactive listening and found that DA instructions boosted the test-takers to initiate developmental changes. HashemiShahraki et al. (2015) conducted a study to estimate test-takers listening conversational implicatures of pragmatic knowledge and conveyed that the mediational support improved their practical grasp of conversational implicatures. For Wang (2015), DA of listening enhances the amalgamation of Assessment and instruction.

Several investigations have identified the possible Effect of C-DA on language skills. For example, Jacobs's (1998) Kidtalk software used a sequence of computerized tasks to assess pre-school learners' ability. Additionally, Birjandi and Ebadi (2012) implemented a similar automated context to gauge the developmental levels of the verbal command and found a significant correlation between the more advanced ZPD and the fewer time learners spent interacting with the mediator. Also, Poehner et al. (2015) designed online multiple-choice (MC) tests of L2 listening in which each test item is tagged along with implicit to explicit graduated prompts. The results indicated significant differences between the actual and mediated scores. In the same vein, Heidari

and Afgari (2015) addressed a web-based investigation on EFL learners' socio-cognitive progress through DA of listening and stressed the actual learners' ability and the diagnosis and Assessment potential of the developmental listening level.

Rationale and problem

The only interventionist study on C-DA of listening was carried out by Poehner et al. (2015). They addressed one of the main issues in applied linguistics, whethAssessmentent and instruction affect each other. What relationship can there be between education and an aAssessmentent? Concerns have arisen over "teaching to the test" and the "power" that tests have gained to control instruction and narrow down the curriculum where teaching aAssessmentent are perceived as mutually exclusive (McNamara, 2001; Shohamy, 2001). It can then be deducted that testing and teaching have been treated as two separate specializations, each having its well-recognized professional journals and conferences. The first step is needed to debate how "interventions based upon the results of dynamic testing provide superior gains" (Elliott, 2003, p. 189). Therefore, this study aimed to show a significant difference between the test-takers actual and mediated scores in the listening test. The other purpose was to indicate how Can computerized dynamAssessmentent reveals the test-takers LP in the listening question types. In addition, it aimed to show if there is any statistically significant difference between the number of hints in the Question types across the levels of the listening ability.

Method

The participants of this study were 48Arab postgraduate EFL learners attending an English intensive course at UUM. They had to take the English language course because of their low performance in English. The test-takers age varied from 20 to 36, and they were mainly selected through the convenience sampling procedure. A placement test was administered to a larger pool of test-takers only, of which 40, who scored IELTS band 5 to 6, were selected to attend the English language course. The instrument used was an adapted C-DA TOEFL listening test where the test-takers listened to two lectures and dialogues and then received 16 MC questions to answer in 70 minutes. Data collection consisted of five phases: Test preparation, piloting, hints development, software preparation, description, and test administration.

The listening test consisted of three classroom lectures, two dialogues, and 34 test items in test preparation. For practical reasons, the listening test was trimmed. Five question types, main idea, attitude, function, detail, and inference, were utilized. In phase two, test piloting, the test was piloted on a group of 30 EFL learners with almost the exact characteristics of the target group. Score reliability, Cronbach's alpha, was estimated at 0.823, and the results of item analysis retained all items since no item was found to be faulty. In phase three, hints' preparation, the test-takers were given two academic lectures and two dialogues, and 16 questions to answer. Three implicit to explicit indications were prepared for each of the listening items. Eight TOEFL teachers checked, adjusted,

and validated the predetermined clues for validity purposes.

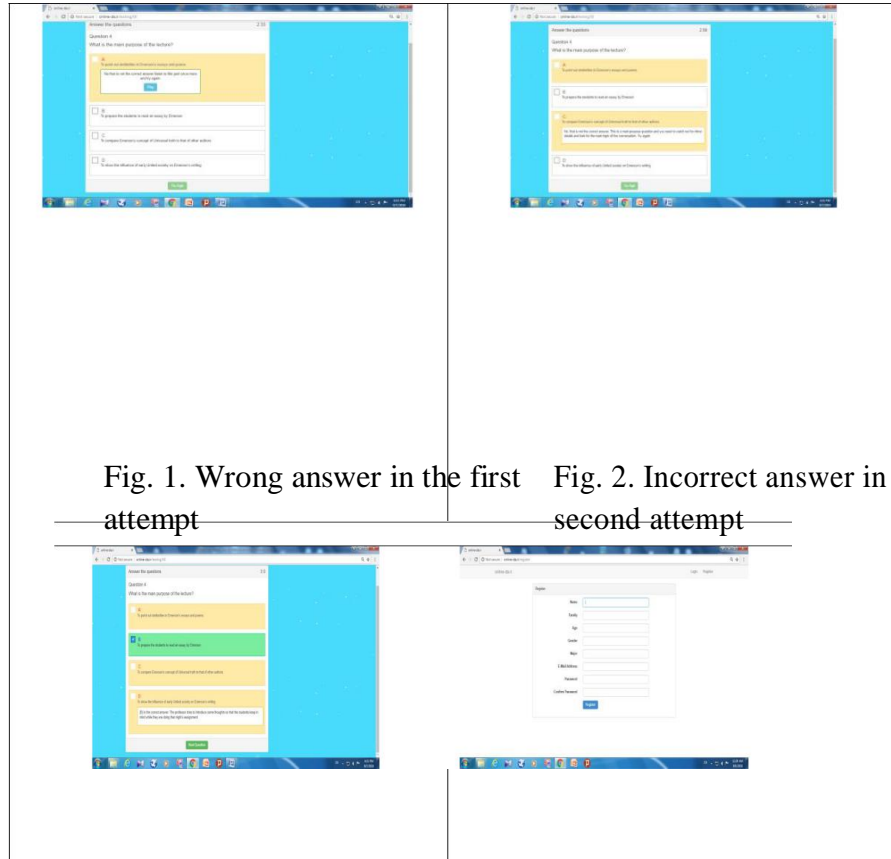


Fig. 1. Wrong answer in the first attempt

Fig. 2. Incorrect answer in the second attempt

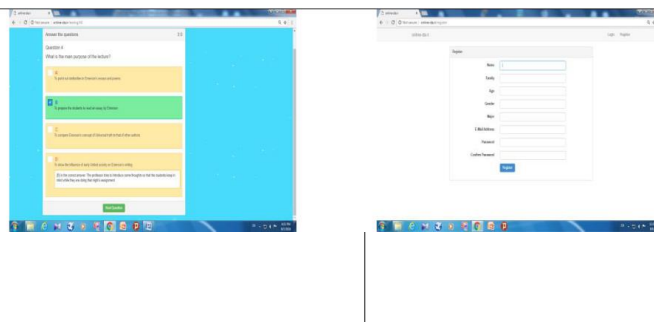


Fig. 3. Wrong answer in the third attempt

Fig. 4. Opening page

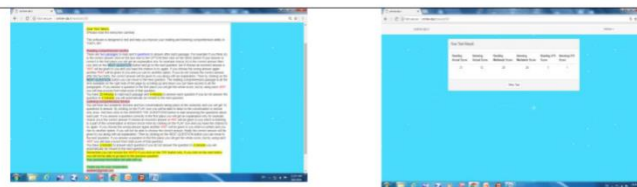


Fig. 5. Instruction page of the online test

Fig. 6. Actual and mediated scores

In phase four, software preparation and description, a computer software expert-designed software to test listening dynamically by offering the test-takers some prepared hints. The software is comprised of three parts: The opening page (Figure 4), the test (a sample of some answers is presented in Figures 1-3), and the scoring file (Figure 6). On the opening page of the software (Figure 4), test-takers were asked to insert their details. The following section describes the software (Figure 5, see Appendix A for a detailed description of the instructions). In the first section, the test-takers listened to a three-minute dialogue, followed by the first Question. If the first attempt was correct, the test-takers could view an explanation for the right option before tackling the next item. If their first attempt was incorrect, a hint was provided, and it asked them to listen again to a part of the conversation.

In the case of the correct answers, an explanation was displayed before proceeding to the next item. If the second trial was incorrect, a more explicit hint was displayed. However, if the test-takers failed to select one answer within three minutes, they were systematically moved to the next Question. The following is an example of the hints:

What is the primary purpose of the lecture?

- a. To point out similarities in Emerson's essays and poems.*
- b. To prepare the students to read an essay by Emerson*

- c. To compare Emerson's concept of Universal truth to that of other authors,*
- d. To show the influence of early united society on Emerson's writing*

Hint One: No, that is not the correct answer. Listen to this part of the lecture once more and try again.

Hint Two: No, that is not the correct answer. This is a central purpose question, and you need to watch out for minor details and look for the main topic of the conversation. Try again.

Hint Three: No, (b) is the correct answer. The professor tries to introduce some thoughts so that the students keep them in mind while they are doing that night's assignment.

If the correct answer is chosen in the first place:

Yes, (b) is the right answer. The professor tries to introduce some thoughts so that the students keep them in mind while they are doing that night's assignment.

In phase five, the software's scoring procedure yielded three sets of scores: Actual, mediated, and LP upon finishing the test. The actual score received three (3) marks if the test-takers answered correctly; if not, it would be zero (0). However, the mediated score was weighted so that a test-taker received each mediating prompt; one mark was deducted. Therefore, for any given item, the test-takers actual score was zero (0) to three, but their mediated score could extend from zero (0) to three (3), depending upon the used number of hints. Finally, an LPS was calculated for each test-taker using Kozulin and Garb formula

(2002). In the last phase, test administration, the test was administered after a formal approval had been received from the ethical committee. Before administering the test, the instructions were explained in Persian, where the test-takers were informed about the purpose of the study and the test procedure using the software.

Results and Discussion

The reliability coefficient analysis (Table 1) displayed a good Cronbach's alpha value ($\alpha = .752$) in the actual and mediated types of scores, totaling 32. To examine the construct validity of these scores, the PCA¹ of the rotated component matrix of the actual and mediated scores designated that the factors loaded at higher values ranging from .90

to .95. In table 2, the researcher presented the descriptive statistics of the actual and mediated scores. The actual scores had values of 7.35 (SD=9.07), 7.63 (SD=4.17), 7.81 (SD=9.22), and 7.44 (SD=4.73), and 9.02 (SD=4.77) for attitude, detail, function, inference and main idea, respectively. The SD values in attitude and function were higher than the other question types, indicating that the test-takers scores in detail, inference, and main idea items were relatively closer. In contrast, the test-takers scores in the attitude and function items were more spread. The mean of the mediated scores in attitude (M=10.41; SD=7.91), detail (M=10.68; SD=4.52), inference (M=10.16; SD=5.12), and main idea (M= 10.51; SD=5.45) items were lower than the function items (M=11.74; SD=7.73).

Table 1 Reliability Statistics of Actual and Mediated Scores (n=185)

Cronbach's Alpha	Number of Items
.752	32

¹Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Table 2 Total Actual and Mediated Scores (n=185)

Total actual scores					Total mediated scores				
Min.	Max.	Mean	SE	SD	Min.	Max.	Mean	SE	SD

Attitude	.00	20.00	7.35	.66	9.07	.00	20.00	10.41	.58	7.91
Detail	.00	17.78	7.63	.30	4.17	.00	17.78	10.68	.33	4.52
Function	.00	20.00	7.81	.67	9.22	.00	20.00	11.74	.56	7.73
Inference	.00	18.33	7.44	.34	4.73	.00	20.00	10.16	.37	5.12
Main idea	.00	20.00	9.02	.35	4.77	.00	20.00	10.51	.40	5.45

The paired t-test, Table 3, shows a statistically significant difference in the actual and mediated scores (column 8) and that the mediated mean scores outperformed the actual ones. The actual means ran from 7.35 (attitude) to 9.02 (main idea), and from 10.16 (inference) to 11.74 (function) for the mediated means. However, the homogeneity of test-takers scores in actual and mediated responses was relatively the same.

Table 3 Actual and Mediated Scores (n=185)

		Mean	SD	SEM	t-value	df	Sig.
Pair 1	main. Idea.actual	9.02	4.77	.35	-3.90	184	.00
	main.idea.mediated	10.51	5.45	.40			
Pair 2	function.actual	7.81	9.22	.67	-5.25	184	.00
	function.mediated	11.74	7.73	.56			
Pair 3	attitude.actual	7.35	9.07	.66	-4.53	184	.00
	Attitude. mediated	10.41	7.91	.58			
Pair 4	inference. actual	7.44	4.73	.34	-5.94	184	.00
	Inference. mediated	10.16	5.12	.37			
Pair 5	detail.actual	7.63	4.17	.30	-6.09	184	.00
	Detail. mediated	10.68	4.52	.33			

Figure 7 introduces the descriptive statistics of the actual and mediated scores of the dialogue and monologue listening input, moving from 6.81 (column 2) and 6.49 to 12.45 and 13.22, respectively. Moreover, the mediated scores in both contexts were more homogenous than the actual ones, thus leading to better improvements and more homogeneity of scores.

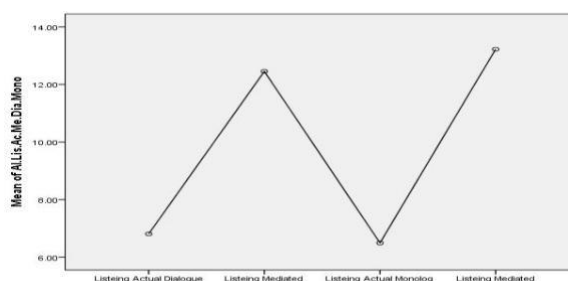


Fig. 7. Dialogue and monologue actual and mediated scores

Concerning the multiple comparisons of the actual and mediated scores in dialogue and monologue tasks, Table 4 demonstrated no significant difference (column 5, with values of .80 and .10) between the natural and mediated scores in dialogue and monologue contexts. At the same time, it was not the case between the other pairs of comparisons with a significant level of $p < .000$.

Table 4 Dialogue and Monologue Comparisons of Actual and Mediated Scores (n=185)

(I)	(J)	95% Confidence Interval				
Fac. AI.Actual.	Factor.Analysis	(I-J)				
Mediated. Dialogue.		Mean			Lower Bound	Upper
Actual.Mediated.Dialogue.		Difference	SE	Sig.	Bound	
Monologue	Monologue					
Actual dialogue	Mediated dialogue	-5.64*	.31	.00	-6.51	-4.76
	Actual monologue	.312	.31	.80	-.56	1.18
	Mediated monologue	-6.41*	.31	.00	-7.28	-5.53
Mediated dialogue	Actual monologue	5.95	.31	.00	5.08	6.82
Actual monologue	Mediated monologue	-.76	.31	.10	-1.64	.10
	Mediated monologue	-6.72*	.31	.00	-7.59	-5.85

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

The descriptive statistics of the LPSs in the various item types, Table 5, demonstrated that the LPS in function item types (=.39) was the highest, followed by attitude (=.30), detail (=.30), inference (=.27), and main idea (=0.14) item types. The median scores of detail item types were the highest (.55), followed by inference and main idea item types (.23 and .22). The median of function and attitude item types was zero (0), implying the test-takers lacked other

hints to answer correctly. Another challenging issue in the minimum LP of all item types was negative with -1.33, -2.00, -1.50, and -1.22 due to other latent variables that distracted the test-takers from answering correctly.

Table 5 Descriptive Statistics of LPSs (n=185)

	Main Idea	Function	Attitude	Inference	Detail
Mean	.14	.392	.30	.27	.30
Median	.22	.00	.00	.33	.55
Mode	.00	.00	.00	.50	.89
SD	.51	1.01	.919	.62	.679
Min.	-1.33	-2.00	-2.00	-1.50	-1.22
Max.	.89	2.00	2.00	1.33	1.33

Table 6 gives data on the LP in the five-question types. The LP of the main idea items went from -1.33 to .89 with 16.2% of the test-takers who had an LP of 0.00. The function LP items varied from -2 to .2, and the mode of function question types was 1.33, which represented a high LP and positive EffectEffect of mediation. The function LP item type indicated that the mediated practices were influential in reaching the correct answers. The LP of attitude items differed from -2 to .2 with a mode of 0.00. The attitude LP specified that 54.1% of the test-takers had an LP below 0.00. The LP of inference items extended from -1.33 to 1.33 with a mode of .50. The detail LP items ranged from -1.22 to 1.33 with 0.89. The wide variety of the LPSs in the detail part meant that the items were more accessible to the test-takers.

Table(6)LP of the Questions Types (n=185)

Main Idea		Function		Attitude		Inference		Detail	
LP	%	LP	%	LP	%	LP	%	LP	%
-1.33	.5	-2.00	7.6	-2.00	5.4	-1.33	1.1	-1.22	1.1
-1.33	1.1	-1.33	1.1	-1.33	3.8	-1.17	.5	-1.11	2.2

-1.11	2.7	-.67	8.1	-.67	6.5	-1.00	3.8	-1.00	2.2
-.89	4.3	.00	34.1	.00	38.4	-.83	1.1	-.89	2.7
-.67	4.3	.67	10.3	.67	17.3	-.83	1.1	-.78	1.6
-.44	.5	.67	.5	1.33	24.9	-.67	3.8	-.78	2.2
-.44	3.8	1.33	31.9	1.33	2.2	-.67	1.6	-.67	1.1
-.22	2.2	1.33	2.2	2.00	1.6	-.50	3.8	-.67	2.2
-.22	1.1	2.00	4.3	Total	100.0	-.33	1.6	-.56	2.2
-.22	.5	Total	100.0			-.33	1.1	-.56	2.2
.00	16.2					-.17	2.2	-.44	3.8
.22	12.4					-.17	.5	-.33	1.6
.22	.5					.00	8.6	-.33	.5
.22	4.9					.17	1.6	-.33	.5
.44	3.2					.17	1.1	-.22	1.6
.44	2.2					.17	3.8	-.11	.5
.44	15.1					.33	10.8	.00	.5
.44	3.2					.33	1.6	.22	1.1
.67	11.9					.33	2.2	.33	4.3
.67	4.9					.50	13.0	.33	.5
.89	2.2					.67	4.9	.33	1.6
.89	2.2					.67	11.4	.44	7.6
Total	100.0					.83	3.8	.44	2.7
						.83	.5	.56	6.5
						1.00	8.6	.56	3.8
						1.17	1.1	.67	3.8
						1.33	4.3	.67	6.5

Total	100.0	.78	2.2
		.78	5.4
		.89	13.5
		1.00	3.8
		1.11	5.4
		1.22	1.6
		1.33	.5
Total	100.0		

To compare the number of hints across the ability levels, a series of Chi-Square analyses, Table 7, was performed. The test-takers were classified based on four relatively equal groups, while their actual scores were examined based on their language ability. The majority of the test-takers clustered around the moderate and moderate-high levels with 48.64% and a low percentage of 14.0 high achievers. (n=48).As shown in the table below:

Table(7)Test-takers' Listening Ability Levels

Ability level	Frequency	Percent
Low	34	18.37
Moderate low	34	18.37
Moderate high	56	30.27
High	26	14.05
Total		100

The results of a series of Chi-square analyses of the mediated scores, Table 8, revealed there were statistically significant differences ($p < .00$, $p < .01$) in using hints in the main idea dialogue, function dialogue, attitude dialogue, detail monologue, inference monologue, main idea dialogue, inference dialogue, detail monologue questions types. However, no statistically significant differences were found in detail monologue (19), inference dialogue (24), and central idea monologue (16) question types.

A paired t-test was carried out, Table 9, and it represented a statistically significant difference in the actual and mediated scores. Therefore, the null hypothesis that the hints could support the test-takers could be safely rejected. The descriptive statistics confirmed that the mediated scores' mean scores outperformed their actual scores.

Table 8 A Sample of the Significance Level Hints in Question Types Across Ability Levels (n=185)

	Mediated ability levels	Mean rank	Chi-square	df	sig	
1. main idea.dialogue	low	34	56.24	15.215	3	.00
	moderate low	34	67.74			
	Moderate high	56	85.73			
	High	26	88.81			
2.function.dialogue	low	34	49.29	20.705	3	.00
	moderate low	34	74.34			
	Moderate high	56	85.99			
	High	26	88.69			
3.attitude.dialogue	low	34	58.59	17.690	3	.00
	moderate low	34	62.82			
	Moderate high	56	85.07			
	High	26	93.58			
	Total	150				
4. main idea.monologue	low	34	63.53	5.035	3	.16
	moderate low	34	81.78			
	Moderate high	56	74.66			
	High	26	84.75			
	Total	150				
5.detail.monologue	low	34	59.74	10.598	3	.01
	moderate low	34	75.13			
	Moderate high	56	76.90			
	High	26	93.58			
6.detail.monologue	Total	150				

	low	34	53.29	21.475	3	.00
	moderate low	34	80.56			
	Moderate high	56	74.04			
	High	26	101.08			
	Total	150				
7. inference. monologue	low	34	59.15	16.217	3	.00
	moderate low	34	64.21			
	Moderate high	56	83.18			
	High	26	95.12			
	Total	150				
9.main.idea.dialogue	low	34	59.09	13.805	3	.00
	moderate	34	69.29			
	Moderate high	56	79.41			
	High	26	96.65			
	Total	150				
10.inference.dialogue	low	34	62.04	5.315	3	.15
	moderate low	34	77.54			
	Moderate high	56	78.15			
	High	26	84.71			
	Total	150				
11.inference.dialogue	low	34	52.68	17.422	3	.00
	moderate low	34	72.62			
	Moderate high	56	82.64			
	High	26	93.73			

	Total	150				
12.detail.dialogue	low	34	63.91	4.183	3	.24
	moderate	34	80.78			
	Moderate high	56	76.04			
	High	26	82.58			
	Total	150				
13. detail.monologue	low	34	59.50	12.170	3	.00
	moderate low	34	71.79			
	Moderate high	56	77.86			
	High	26	96.19			
	Total	150				
14. detail. monologue	low	34	66.90	7.699	3	.05
	moderate low	34	77.49			
	Moderate high	56	70.92			
	High	26	94.02			
	Total	150				
15.detail.monologue	low	34	62.34	4.704	3	.19
	moderate low	34	80.21			
	Moderate high	56	78.11			
	High	26	80.94			
	Total	150				

Table 9 A Sample of Test-takers’ Actual and Mediated Performance as per Question Types (n=185)

Question Types	Mean	SD	SEM	t-value	df	Sig.
----------------	------	----	-----	---------	----	------

Pair 1	1.actual.mainidea.dialogue	1.02	1.42	.10	-13.9	184	.00
	1.mediated.mainidea.dialogue	1.99	.96	.07			
Pair 2	2.actual.function.dialogue	1.00	1.41	.10	-13.7	184	.00
	2.mediated.function.dialogue	2.01	.93	.06			
Pair 3	3.actual.attitude.dialogue	1.02	1.42	.10	-14.8	184	.00
	3.mediated.attitude.dialogue	2.05	.89	.06			
Pair 4	4.actual.mainidea.monologue	.92	1.38	.10	-13.7	184	.00
	4.mediated.mainidea.monologue	1.88	1.00	.07			
Pair 5	5.actual.detail.monologue	.77	1.31	.09	-16.8	184	.00
	5.mediated.detail.monologue	1.92	.89	.06			
Pair 7	7.actual.inference.monologue	.95	1.40	.10	-15.6	184	.00
	7.mediated.inference.monologue	2.01	.87	.06			
Pair 10	10.actual.inference.dialogue	.92	1.38	.10	-15.3	184	.00
	10.mediated.inference.dialogue	1.94	.90	.06			
Pair 12	12.actual.detail.dialogue	1.20	1.47	.10	-12.6	184	.00
	12.mediated.detail.dialogue	2.09	.91	.06			
	15.mediated.detail.monologue	2.08	.91	.06			

The sample comparison, Table 9 (8 pairs out of 16), made clear that there was a statistically significant difference between the test-takers' actual and mediated scores as to the various question types including the main idea dialogue, pair 1, column 6, (T (47)= -13.9, $p < .00$), function dialogue, pair 2, (T (48)= -13.7, $p < .00$), attitude monologue, pair 3, (T (47)= -14.8, $p < .00$), main idea monologue, pair 4, (T (47)= -13.7, $p < .00$), detail monologue, pair 5, (T (47)= -16.8, $p < .00$), inference monologue, pair 7, (T (47)= -15.6, $p < .00$), attitude dialogue (T (47)= -14.8, $p < .00$), inference dialogue, pair 10, (T (47)= -15.3, $p < .00$), detail dialogue, pair 12, (T (47)= -12.6, $p < .00$), and purpose monologue (T (184)= -13.5, $p < .00$).

Discussion

Results of the study proved that a complete understanding of the listening ability required active intervention in its development. Accordingly, this shifts the classical view of Assessment to a more interventionist one by accentuating the process rather than the product. The current study echoed the findings of other DA studies (e.g., Lantlof&Poehner, 2008; Poehner *et al.*, 2015), who argued that DA offers a diagnostic understanding of the test-takers' difficulties through the provision of particular predetermined hints and prompts during the assessment process. The results suggested that C-DA, i.e., the integration of teaching an assessment via software with predetermined suggestions, could serve as a mere diagnosis of the test-takers' ability both in the ZAD and ZPD. This

integration was efficient in probing the test-taker's LP. The findings could be supported because DA creates a supportive atmosphere to highlight the test-taker's further learning and improvement by judging their ZAD and ZPD (Ahmadi&Barabadi, 2014).

Moreover, the test-takers' significant improvement from the actual to mediated performance could typically be attributed to the aspects of C-DA that could eliminate any possible learning handicaps. DA procedures helped activate the metacognitive listening strategies, which was in line with other studies (e.g., Ajideh, &Nourdad, 2013; Alavi, KaivanpanahShabani, 2012; Haywood &Lidz, 2007; Pishghadam&Barabadi, 2012; Poehner, 2007). Concerning the question types' LP, the findings highlighted that the highest LP was in the function items, followed by the main idea, attitude, detail, and inference items. Therefore, gaining information about this potential could allow the language learners to have an accurate picture of their capabilities (e.g., Peña *et al.*, 2001). This aspect was reflected in the current study, especially when the C-DA hints spontaneously promoted the test-takers to self-assess their ability. Also, a statistically significant difference was found between high, moderate-high, moderate-low, and low achievers in the number of hints used in almost all question types, except for the inference and detail items. This showed that the test-takers in the diverse listening ability levels tended to have recourse to their multifarious traits to answer the main idea, attitude, and function items, while mainly

relying on the allocated hints to answer detail and inference items.

The results highlighted the significant and positive difference in the actual and mediated performances of the test-takers in the monologue and dialogue parts. However, like other studies (e.g., Ableeva, 2008; Poehner *et al.*, 2014; and Shrestha & Coffin, 2012), no significant difference was found in the actual and mediated scores in the dialogue and monologue contexts. This implicated that the test-takers brought into play their cognitive abilities in taking advantage of the hints in the monologue and dialogue types. Consequently, assessing listening in the dialogue and monologue contexts dynamically involved the test-takers in joint activities to overcome task difficulty and attain the level where they could construct meaning in an independent and self-governing way (Author, 2014, 2017). Further, DA attends to the development and learning and supports discovering the test-takers developing capabilities that are different from their actual skills (Shrestha & Coffin, 2012). However, Anton (2009) argued that educators could misrepresent the test-takers qualifications if they rely on only traditional assessments. The study's findings in this respect agree with other studies (e.g., Ahmadi & Barabadi, 2014; Anton, 2009; Haywood & Lidz, 2007; Author, 2014, 2017; Poehner & Lantolf, 2005).

A high LPS meant that the test-takers ZPD level was typically close to their own ZAD when the targeted capability was close to internalization (Kozulin & Garb, 2002). Conversely, a low LPS was evidenced by

the test-takers need for more predetermined hints and some external assistance to internalize the targeted LP in Question. Along with this conceptualization, it was reflected that the EFL learners with a low LPS in the study took advantage of much more mediation of predetermined hints than those EFL learners with high LPS. The findings of this study are congruent with Kozulin and Garb's study (2002). However, to the researchers' knowledge, regarding the difference between the numbers of hints applied for each Question type of listening, no study was found to scrutinize the difference between the variables above directly. Finally, the statistically significant difference between the test-taker's actual and mediated performances meant a significantly higher performance in the mediated performance than their actual performance one. This efficiency of C-DA for the test-takers language development in general and listening, in particular, is reported by other studies (e.g., Ahmadi Safa & Jafari, 2017; Alavi, Kaivanpanah & Shabani, 2012; Lantolf & Aljaafreh 1995; Sadeghi & Khan Ahmadi, 2011).

Overall, this study supports C-DA's positive effect on EFL learners' listening ability. Applying various formats of DA procedure, encompassing C-DA in a sociocultural context, is regarded as an essential step to shifting the paradigm of "teaching to the test movement" (Shohamy, 2001) to a "testing to the teaching movement" whose true objective would be to serve and guide test-takers to learn independently. Mediation can potentially activate past knowledge, raise

consciousness, and help to boost active learning. And it is in this regard that the current study should be contextualized.

Implications, Limitations, and Recommendations

Several features of C-DA, namely improving the test-takers listening ability and providing information about their LPSs, could empower the language teachers and material and test designers to use such types of assessments interactively and productively. To this end, C-DA offered both EFL learners hints which engaged them with the appropriate tools to diagnose problems and find remedies to the listening problems, and language teachers to understand the test-takers LP in the Question types for which there is no need for a tutorial to reflect language test-takers ZAD and those for which a tutorial brings about correct responses to integrate language test-takers ZPD (Poehner&Lantolf, 2013). Additionally, the findings of this study bespoke the support role computers and software can give language teachers and test developers in implementing the main principles of DA, permitting educators to assess a large number of the test-takers dynamically simultaneously. Since C-DA allowed for learners' self-assessment and reassessment, such a procedure encourages and inspires language learners to join the language learning and assessment process essentially. Curriculum and material developers are recommended to use C-DA so that language test-takers are no longer dependent upon their teachers to be assessed and detect their progress. In other

words, with the availability of C-DA, language learners can consider and reassess themselves as many times as required, and such an opportunity naturally might pave the way to becoming autonomous language learners.

The first delimitation of this study rested on the use of the interventionist approach, and the main limitation of C-DA was that the chances of co-constructing the ZPD decreased (Poehner, 2008). Lidz and Gindis (2003), like the case with this study, stated that integration arises as the intervention is intertwined with the assessment process to explain the test-takers capabilities and assist them in reaching more challenging levels. Kozulin (2003) argues that test-takers cognitive development mainly depends on mastering these instruments; however, these tools might not function successfully without a mediator. Hence, the limited aspects of C-DA in making the test-takers reach their full potential.

Since the findings indicated C-DA's significant and positive impact on improving test-takers listening ability, EFL learners in similar contexts can take advantage of such a procedure to enhance their listening ability. Furthermore, EFL teachers should encourage their learners to participate in DA activities individually, in pairs, or in groups. C-DA had a significant and positive impact on listening monologues and dialogues of TOEFL, and perhaps EFL teachers are recommended to implement this model Assessment in various language courses. Educational decision-makers and teachers should initiate ways to use and apply C-DA and

traditional standardized tests. However, to garner more generalizable data regarding C-DA effectiveness, other language skills and even subskills such as vocabulary needed to be explored using the same type of Assessment. Curriculum and material developers are recommended to focus on DA and C-DA and suggest various materials and user-friendly software in class. This study can be replicated on other high-stakes tests by considering a proportionate number of males and females. This study was carried out among learners within the age span of 20-36 years old; the same research could be done among a different age group to check the probable EffectEffect of age on performance.

Conclusion

This study could be perceived as a timely contribution to the other works undertaken on

References:

- Ableeva, R. (2008). The effects of dynamAssessmentent on L2 listening comprehension. In J. P. Lantolf& M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 57-86). Equinox.
- AhmadiSafa, M., &Jafari, F. (2017). The washback EffectEffect of dynamAssessmentent on grammar learning of Iranian EFL learners. *Journal of Language Teaching and Learning*, 7(1), 55—68.
- Ahmadi, A., &Barabadi, E. (2014). Examining Iranian EFL learners' knowledge of grammar through a computerized dynamic test. *Issues in Language Teaching*, 3(2), 161-183.
- Alavi, S. M. (2014). Examining the role of dynamAssessmentent in the development aAssessmentent of listening comprehension. *English Language Teaching*, 1(2), 23—40.
- Alavi, S. M., Kaivanpanah, S., &Shabani, K. (2011). Group dynamic assessment: An inventory of mediational strategies for teaching listening. *Journal of TeachingLanguage Skills*, 30(4), 27-58.
- Anton, M. (2009). DynamAssessmentent of advanced second language learners. *Foreign language Annals*. 42(3). 576—598.

-
- Birjandi, P., &Ebadi, S. (2012). Microgenesis in dynamAssessmentent of L2 learners' socio-cognitive development via web 2.0. *Procedia-Social and Behavioral Sciences*, 32, 34—39.
 - Buck, G. (2001). *Assessing listening*. CUP.
 - Crook, C. (1991). Computers in the zone of proximal development: Implications for evaluation. *Computers and Education*, 17(1), 81–91.
 - Dixon-Krauss, L. (1996). *Vygotsky in the classroom: Mediated literacy instruction aAssessmentent*. New York, NY: Longman.
 - Elliott, J. (2003). DynamAssessmentent in educational settings: Realizing potential. *Educational Review*, 55(1), 15—32.
 - Emadi, M., &Arabmofrad, A. (2015). Individual dynamAssessmentent: An analysis of Iranian EFL learners' listening comprehension errors. *Theory and Practice in Language Studies*, 5(12), 25—39.
 - Feuerstein R. & Feuerstein S. (1991). *Mediated learning experience: A theoretical review*. In R. Feuerstein, P.S. Klein, & A. Tannenbaum (Eds.). *Mediated LearningExperience* (pp. 3—52). London: Freund.
 - Ghahremani, D. (2013). The effects of implementing summatiAssessmentent, formatiAssessmentent, and dynamAssessmentent on Iranian EFL learners' listening ability and listening strategy use. *Journal of language and translation*, 3(2), 5, 10—24.
 - Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. CUP.
 - HashemiShahraki, S., Ketabi, S., &Barati, H. (2015). Group dynamic assessment of EFL listening comprehension: Conversational implicatures in focus. *InternationalJournal of Research Studies in Language Learning*, 4(3), 73-89.
 - Haywood, H. C. &Lidz, C. S. (2007). *DynamAssessmentent in Practice: Clinical and educational applications*. CUP.
 - Heidari, D. M., &Afghari, A. (2015). The EffectEffect of dynamAssessmentent in synchronous computer-mediated communication on Iranian EFL learners' listening comprehension ability at an upper-intermediate level. *English Language Teaching*, 8(4), 14—30.
 - Jacobs, E. L. (1998). KID TALK: A computerized language screening test. *Journal of Computing in Childhood Education*, 9(2), 113—131.
 - Kozulin, A. (2003). Psychological tools and mediated learning. *Vygotsky's educational theory in cultural context*, 15-38.
 - Kozulin, A. & Garb, E. (2002). DynamAssessmentent of EFL text comprehension of at-risk students. *School Psychology International* 23, 112–127.
 - Lantolf, J. P. (2000). Second language learning is a mediated

-
- process. *Language Teaching*, 33(02), 79—96.
- Lantolf, J. P. (2007). Sociocultural source of thinking and its relevance for second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 31—33.
 - Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamAssessmentent. *Language Teaching Research*, 17(3), 323—342.
 - Lantolf, J. P., &Aljaafreh, A. (1995). A second language learning in proximal development: A revolutionary experience. *International Journal of educational research*, 23(7), 619—632.
 - Lantolf, J. P., &Poehner, M. E. (2011). DynamAssessmentent in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 11— 33.
 - Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. OUR.
 - Lidz, C. S. (1995). DynamAssessment ent and the legacy of L. S. Vygotsky. *SchoolPsychology International*, 16(2), 143—153.
 - Lidz, C. S., &Gindis, B. (2003). DynamAssessmentent of the evolving cognitive functions in children. *Vygotsky's educational theory in cultural context*, 99-116
 - McNamara, T. (2001). Language assessment as Social Practice: Challenges for research. *Language Testing*, 18(4), 333-349.
 - Mendelsohn, D. J. (1994). *Learning to listen: A strategy-based approach for the second language learner*. Dominic Press.
 - Nunan, D. (1998). Approaches to teaching listening in the language classroom. *Paper presented at the Korea TESOL Conference*, Seoul.
 - O'Malley, J. M., Chamot, A. U., &Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418—437.
 - Peña, E., Iglesias, A., &Lidz, C. S. (2001). Reducing test bias through dynamAssessmentent of children's word learning ability. *American Journal of Speech-LanguagePathology*, 10(2), 138—154.
 - Poehner, M. E. (2007). Beyond the test: L2 dynamAssessmentent and the transcendence of mediated learning. *Modern Language Journal*, 91(3), 323—40.
 - Poehner, M. E. (2008). *DynamAssessmentent: A Vygotskian approach to understanding and promoting second language development*. Springer.
-

-
- Poehner, M. E. (2009) Group dynamic assessment: Mediation for the L2 classroom. *Tesol quarterly*, 43(3), 471—491.
 - Poehner, M. E. &Lantolf, J. P. (2010). Vygotsky's teaching assessment dialectic and L2 education: The case for dynamAssessmentent. *Mind, Culture, and Activity: An International Journal*, 17, 312—330.
 - Poehner, M. E. &Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamAssessmentent. *Language teaching research*, 17(3) 323—342.
 - Poehner, M. E., &Lantolf, J. P. (2005). DynamAssessmentent in the language classroom. *Language Teaching Research*, 9(3), 233-265. Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamAssessmentent: Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337-357.
 - Rost, M. (2002). *Teaching and researching listening*. Longman.
 - Sadeghi, K., & Khan Ahmadi, F. (2011). DynamAssessmentent of L2 grammar of Iranian EFL learners: The role of the mediated learning experience. *International Journal of Academic Research*, 3(2), 931—936.
 - Scarcella, R. C., & Oxford, R. L. (1992). *The tapestry of language learning: the individual in the communicative classroom*. Heinle&Heinle.
 - Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman.
 - Shrestha, P., & Coffin, C. (2012). DynamAssessmentent, tutor mediation, and academic writing development. *Assessing Writing*, 17(1), 55—70.
 - Torrance, H. & J. Pryor. (1998). *Investigating formatiAssessmentent: Teaching, learning aAssessmentent in the classroom*. Open University Press.
 - Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer-assisted dynamAssessmentent. *Journal of Computer Assisted Learning*, 18(1), 21—32.
 - Underwood, M., (1989). *Teaching listening*. Longman.
 - Vandergrift, L. (2004). Listening to learn or learning to listen. *Annual Review of AppliedLinguistics*, 24, 3-25.
 - Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, 40(3), 191-210.
 - Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
 - Vygotsky, L. S. (1986). *Thought and language*. MIT Press.
 - Wang, P. (2015). The EffectEffect of dynamAssessmentent on the listening skills of lower-intermediate EFL learners in Chinese technical college: A pilot study. *Journal of Language*

Teaching and Research, 6(6), 1269—
1279.

- Wang, T. H. (2008). Web-based quiz-game-like formatiAssessmentent: Development and evaluation. *Computers & Education*, 51(3), 1247-1263

You have 4 minutes to answer each question. If you do not answer the question in 4 minutes, you will automatically be moved to the next Question.

Remember, you can receive the **HINTS** if you only click on the TRY button. If you click on the Next button, you will not be able to go back to the previous Question.

Your personal information will be safe with us.

Thank you for your cooperation.

Appendix A

Dear Test-Takers,

Please read the instruction carefully.

This software is designed to test and help you improve your listening comprehension ability in TOEFL IBT.

You will hear two academic lectures and two conversations at the university and get 18 questions to answer. By clicking on the PLAY icon, you will be able to listen to the conversation or lecture only once. And then click on the ANSWER THE QUESTIONS button to start answering the questions about each part. If you answer a question correctly in the first place, you will get an explanation why for example, choice (A) is the correct answer. If you choose an incorrect answer, a **HINT** will be given to you, which is listening to a part of the conversation or lecture by clicking on the PLAY icon, and you can try again. If you choose the wrong answer, another **HINT** will be given to you, which is written, and you can try another option. If you cannot select the correct answer, the correct answer will be given to you and an explanation. Then by clicking on the NEXT QUESTION button, you can move to the next Question. If you answer a question in the first place, you will get the total score, but by using each **HINT**, you will lose a score from the total score of that Question.