



Applying Interventionist and Interactionist Approaches to Dynamic Assessment for L2 Listening Comprehension

Elyas Barabadi

Asst. Prof., Language Education, University of Bojnord, Iran, elyas.ba1364@gmail.com

Gholam Hassan Khajavy

Asst. Prof., Language Education, University of Bojnord, Iran, hkhajavy@gmail.com

Ali Mehri Kamrood

PhD Student of TEFL, Hakim Sabzevari University, Iran, alimehrikamrood@gmail.com

The current study reports the results of a project aimed at assessing L2 listening comprehension by drawing on two approaches to dynamic assessment: interventionist and interactionist. The former approach was actualized by providing two graduated hints which were fixed and standardized for all test takers while the latter was actualized by asking test takers to enter into dialogue with the tester to receive more flexible mediation. By taking into account how much mediation (how many hints) each test taker uses, the computerized listening test automatically provides three scores: traditional score, dynamic score, and learning potential score. 91 grade nine students at Iranian junior high schools took the computerized dynamic test. The findings of the study indicated that while having validity, the test could improve test takers' listening comprehension ability. Moreover, the test made it possible to obtain information regarding test takers' potentiality for learning which goes beyond and over their initial performance level. The pattern of mediation use by test takers across different test items indicated that more challenging listening questions are in need of more flexible and open-ended mediation. Using three scores for each test taker, language teachers can more effectively tailor their instruction to their learners' needs.

Keywords: computerized dynamic assessment, listening, ZPD, LPS, listening comprehension, assessment

INTRODUCTION

Rooted in Vygotsky' socio-cultural theory especially the notion of zone of proximal development (ZPD), DA posits that potential development differs from actual development, meaning that the latter cannot be a predictor of the former (Poehner & Lantolf, 2005). In other words, instead of focusing on students' current knowledge and

Citation: Barabadi, E., Khajavy, G. H., & Kamrood, A. M. (2018). Applying Interventionist and Interactionist Approaches to Dynamic Assessment for L2 Listening Comprehension. *International Journal of Instruction*, 11(3), 681-700. <https://doi.org/10.12973/iji.2018.11346a>

skills, dynamic procedures stress students' potential for learning and explore ways to realize those potentials (Kozulin & Garb, 2004). Proponents of DA (Elliot, 2003; Kozulin & Garb, 2004; Lidz & Elliot, 2006; Poehner & Lantolf, 2013; Shabani, 2014; Wang, 2015) claim that DA procedures are both feasible and effective for obtaining information about students' learning potential. DA procedures can better differentiate students concerning their gain of new text comprehension strategies. However, more research is needed to account for the informative role of DA in the second or foreign language (L2) context.

So far, the DA researchers have opted for either of the two approaches to DA, that is, the interventionist approach or the interactionist approach in conducting their DA studies. The former favors a standardized procedure to delivering mediation where a set of pre-planned and predetermined sets of hints are given to test takers in case they cannot give the right answer to a test question. The interactionist approach takes a more radical attitude towards providing mediation by postulating that mediation should be sensitive to each individual's ZPD. In other words, interactionist DA is not in line with any standardization in DA procedures. Though more effective than pre-determined form of mediation, this kind of individualized form of mediation is difficult to apply to a large number of test takers (Poehner, 2008). Nonetheless, one of the most significant current discussions in designing and implementing DA procedures is associated with the choice of the most appropriate approach of DA.

While the interventionist computerized version of DA can overcome one of the major shortcomings of the interactionist DA which is related to its application to a large number of students simultaneously (Lantolf & Poehner, 2014) regarding different language constructs such as reading comprehension (Lantolf & Poehner, 2014; Pishghadam & Barabadi, 2012), L2 grammar (Ahmadi & Barabadi, 2014) and listening comprehension (Poehner et al., 2015), the interactionist DA proponents have raised serious concerns over the credibility of the findings of such studies claiming that interventionist DA sacrifices the individuality of the learners for quantifying and standardizing the mediation process. Accordingly, the primary purpose of this study is to find a tradeoff between the two approaches via designing and implementing a Computerized Dynamic Assessment (C-DA) in L2 listening comprehension including two pre-determined hints for each item (interventionist), and a third hint asking test takers to ask for direct help from their teachers (interactionist) in case the first two hints did not prove useful for getting to the correct response. In other words, this approach to computerized dynamic assessment can simultaneously overcome two limitations of DA: its applicability to a large number of students by following a standardized format, and the effectiveness of mediation by making room for individualized form of mediation as is common in interactionist models of DA by including a third hint asking the test takers to ask for direct help from their teachers. More specifically, the following research questions are dealt with in this study:

1. Is there any significant difference between test takers' actual scores and mediated scores? In other words, can students obtain remarkable gain scores after receiving mediation?

2. Does the computerized dynamic listening test have reliability and validity?
3. To what extent can the Learning Potential Score (LPS) differentiate among the learners with the same actual scores?
4. How is open ended mediation (Hint 3) used across different listening questions?

Theoretical Underpinning of DA

DA has its origin in the Vygotsky's socio-cultural theory of mind. Vygotsky believed that it is possible to figure out individuals' fully-formed as well as emerging abilities by providing mediation, and analyzing how responsive and open they are to mediation (Vygotsky, 1998). A true measure of cognitive ability should provide not only a diagnosis but also a good prognosis concerning future performance. Vygotsky (1978) rejects the assumption that what individuals can do independently be considered as the only indicator of their mental abilities. Instead, adopting a quite non-orthodox stance concerning cognitive abilities, Vygotsky (1978) observes that "what children can do with the assistance of others might be in some sense even more indicative of their mental development than what they can do alone" (p. 85). From this perspective, learners' actual developmental level only defines the end product of development or what is known as zone of actual development (ZAD). On the other hand, ZPD defines the emerging functions which have not yet been matured but are in the process of maturation. According to Vygotsky (1987), these processes are "in an embryonic state" (p. 86).

Nonetheless, in Vygotsky's (1987) view, the mainstream testing systems primarily focus on individuals' independent activity not their imitative activity as indicator of their level of mental development. In other words, in evaluating mental development, only solutions to test problems which the test takers can arrive at without assistance from others, without leading questions, and without demonstrations are taken into account. In sum, traditional tests only aim at revealing test takers' ZAD not their ZPD. According to Vygotsky (1978), the ZPD is "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86). The actual level of development defines development "retrospectively". Moreover, in his view, ZPD can never accurately be predicted on the basis of ZAD. Accordingly, he strongly recommends the inclusion of external assistance in the form of leading questions, prompts, or demonstration in testing situations in order to accurately infer test takers' ZPD based on their openness and responsiveness to that mediation or external help. Inspired by this insight, many researchers (e.g. Ableeva, 2008; Ahmadi & Barabadi, 2014; Lantolf & Poehner, 2014; Pishghadam & Barabadi, 2012) have investigated the effectiveness of DA in L2 contexts. However, they have conducted their DA procedures through two opposing DA approaches, that is, interventionist which is favored by those who were in search of standardization and quantification; and interactionist whose adherents believed that presenting a one-size fits-all scheme of mediation for a large number of the participants is against the original ideas of Vygotsky and Feuerstein as the theoretical founders of DA. Hence, this study

aimed at conducting a mixed approach of C-DA including both the interventionist and interactionist approaches.

Literature Review

Pishghadam and Barabadi (2012) reported the construction and validation of computerized DA in the context of L2 learning. Particularly, they examined the effectiveness of electronically delivering mediation in improving L2 learners' reading comprehension. To achieve this aim, a software package named Computerized Dynamic Reading Test (CDRT) was developed by the researchers. The software was capable of providing test takers with strategy-based mediation. Each student was assigned two scores: a dynamic score and a non-dynamic score. The findings of the study indicated that the computerized dynamic test not only improved students' reading comprehension ability, but also it provided information about test takers' potentiality for learning. The psychometric features of testing, namely reliability and validity were reported in this computerized dynamic test. The researchers argued that computerized DA used in their study can enable L2 educators to use DA for a large number of test takers. In fact, the applicability of DA which was one of its major drawbacks was taken care of. Besides, by taking into account how much mediation each learner receives, such a computerized dynamic test can help L2 educators provide students with more individualized and consequently more effective instruction.

Poehner and Lantolf (2013) investigated the use of computerized DA in the context of L2 listening and reading comprehension in which mediation was delivered through an online format. By incorporating a system of graduated prompting for each item as well as inclusion of some transfer items, the computerized dynamic tests could act not only as a diagnosis tool but also as a means for promoting learners' abilities in reading and listening. Transfer items refer to those assessment tasks that have the same basic principles as regular items on the test but are more complex and demanding. Test takers' response to such items is indicative of whether or not they have internalized mediation that was offered previously. Moreover, learning potential score was an effective method of showing test takers' potential for learning by taking the difference between their mediated and unmediated performance as the basis for anticipating their future responsiveness to mediation.

Poehner, Zhang and Lu (2015) investigated the effectiveness of electronically delivering online mediation in two skills of reading and listening comprehension. The main premise of their online multiple-choice tests was that mediation is necessary for diagnosing development. Though the online provision of mediation was an innovation in their study, the mediation followed a standardized and fixed format which was similar to previously-mentioned studies. In other words, mediation was in the form of a set of graduated prompts arranged from the most implicit to the most explicit. Diagnosis in their study was made based on two sets of scores: test takers' actual score and mediated score. Based on the findings of their study, the researchers argued that these two scores along with the breakdown of test takers' performance on particular items "...provided a fine-grained diagnosis of their L2 development while also offering information relevant to subsequent teaching and learning" (p. 337). Though it generated different scores for a

large number of the participants, the standardization of the mediation process overlooked the individuality of the participants.

More recently, Ebadi and Saeedian (2016) examined Iranian EFL University students' development in L2 reading comprehension in more difficult and novel situations called transcendence tasks (TR). The main purpose of the study was to see whether test takers could sustain their improved performance in more challenging and novel tasks. The design of their study consisted of the following stages: pretest, enrichment program, posttest, and transfer assessments. Based on the feedback obtained during the pre-test, the researchers could provide contingent mediation during the enrichment sessions whose group format enabled the test takers to receive flexible mediation in an interactionist way. After the enrichment program, the students took three computerized dynamic tests using three software programs developed by the researchers. The first computerized dynamic test which was used as the posttest contained simpler passages than the two computerized dynamic tests that were used during the last stage of the study: transfer assessments. The mediation provided by these three computerized tests was in the form of graduated hints arranged from the most implicit to the most explicit. The results of the study indicated that test takers could transfer their learning during the enrichment sessions and posttest to transcendence tasks (e.g. the last two computerized dynamic tests). The researchers in this study could somehow combine interactionist and interventionist types of mediation in a single study. Though, these two types of mediation were provided in different time intervals.

Finally, Mehri, Davoudi, Amirian, and Ghaniabadi (2018) investigated the transcendence of learning among 43 Iranian EFL University students who took the online computerized dynamic test of listening skill. The main purpose of this study was to examine and account for test takers' development in the context of English listening skill by analysing their performance on transcendence/transfer (TR) tasks. These listening tasks were more demanding than the original listening tasks used in their study. The analysis of TR scores revealed some instances of regression in the case of vocabulary questions, sustenance in the case of idiomatic and purpose questions, and progress in the case of inferential and actual questions. Besides, significant instances of regression, sustenance, and progress were observed across individuals. Although reporting TR scores in this study could provide useful information regarding learners' development across different language constructs, the mediation used in this study was in the form of only predetermined hints arranged from the most implicit to the most explicit.

A brief look at computerized dynamic tests reviewed above except Ebadi and Saeedian (2016) indicate that they only drew on predetermined and fixed type of mediation which is favoured by interventionist approach to dynamic assessment. In the case of Ebadi and Saeedian' (2016) study, it was also shown that the two types of mediation were offered during two distinct sessions. The present study, however, aims at exploring the feasibility and effectiveness of both types of mediation, that is, conducting an amalgam of both interventionist and interactionist approaches of providing mediation in a single DA procedure by inclusion of two graduated hints and a third hint asking test takers to

start a dialogue with the tester in order to receive more contingent mediation which is more responsive to their ZPD. To the best of the researchers' knowledge, no DA researcher has embarked on such a study so far.

METHOD

Participants

In the piloting stage, 63 ninth grade students took the traditional listening test in order to examine the difficulty level of the items. After the C-DA listening test was designed, another group of 20 ninth grade students took the test to make sure about the efficiency of the hints and to prepare some probing questions so that teachers can better and more effectively mediate test takers in the third hint. Finally, the modified computerized listening test was administered to 91 ninth grade students who were in different junior high schools. The participants consisted of 55 male and 36 female students and the mean age of the participants was 16. The test was administered by 13 language teachers in 13 junior high schools in North Khorsan and Khorasan Razavi in April and May 2016. It should be noted that determining the precise proficiency level of test takers in dynamic assessment is not an issue since what matters in such tests is to determine to what extent the test takers can benefit from mediation depending on their proficiency levels. Nonetheless, the fact that all test takers were selected from grade nine makes it possible to claim that the content of the test and overall difficulty level of the test match test takers' proficiency level.

Instrument

The only instrument used in this study was computerized dynamic test of listening which is capable of generating two scores; an actual score based on test takers' independent performance on the test, and a mediated score by taking into account the amount of mediation used by each test taker. The software consists of three parts: introduction, the main body of the test consisting of 12 items, and the scoring file. The introduction serves two purposes: test takers fill out the form with their personal information about them such as name and major. It also gives test takers a short description of DA. The main part of the software involves 12 listening items. Each item is composed of a short conversation along with a completion item to be filled out by a word or phrase from the conversation. For each item, four hints were prepared. There is no time limit for answering the questions. Finally, upon completion of the test, a scoring file containing actual score, mediated score, and the number of hints for each item is generated.

Test construction procedure

Item Construction

As mentioned earlier, the purpose of this study was to develop a computerized dynamic test of L2 English listening comprehension for Iranian junior high school students. The first step in designing the test was to prepare test items. Since the current listening test is criterion-referenced, identifying the distinguishing language characteristics of the target language use (TLU) domain was not difficult (Wagner, 2015). Indeed, the criteria which are in line with the curricular goals of the class enabled us to make the test task

characteristics similar to and representative of TLU domain. Each test item included a conversation as language input to use Bachman and Palmer's (1996) terminology. The 12 conversations used in this study were taken from Khate Sefid which is a private Iranian publisher producing supplementary materials for public school courses including English. Initially, there was a large pool of listening items in Khate Sefid package designed for grade nine students. Having examined all these items, two experts in EFL testing chose 20 listening items for the purpose of the current study. Further examination of these 20 items by the researchers of the study led to the omission of further eight items, and ultimately 12 items were used for our study. Although such a procedure is not appropriate from a traditional psychometric perspective to assessment, for dynamic assessment whose overriding purpose is test takers' development, what matters most is the provision of appropriate mediation that can help not only students' development but also can reveal their potentiality for learning.

After each conversation is played, test takers listen and at the same time can read a question in English to which they need to answer by writing the correct word or phrase in the blank. First, the test takers were presented with six easy items and then with six difficult items. The difficulty level was taken into account by considering the length of conversation, new words, and the rate of delivery (Wagner, 2015). The grade nine textbook has six lessons dealing with six different language functions including personality, travel, festivals, services, media, and health. For each function of the textbook, one easy and one difficult item was designed. On average, the conversations of the easy items contained five turns. Besides, each turn was rather short containing only the key words related to the function of the lesson. As for the six conversations used in difficult items, it was expected that these conversations were more difficult since they were double the length of the first six conversations. Besides, these conversations included many new vocabularies unrelated to the function of the lesson. Lantolf and Poehner (2005) also used somehow the same method for developing more difficult items for their study.

As for the characteristics of the expected response which is another important step in designing a listening comprehension test (Wagner, 2014), we considered limited production response as the best kind of response since giving hints to test takers would not give away the correct response. In other words, item formats such as multiple choices and matching were not used because when the test takers are given a hint, the guessing chance would increase.

Test Piloting and Preparing Hints

Once the items were developed, the test was piloted in its traditional format without providing any mediation. The purpose of this piloting stage was to make sure about the difficulty level of the items and do some other modifications. Having revised the items based on the information obtained from the piloting, the next step was to prepare appropriate hints by making some modifications in two important input characteristics: input format and characteristics of the language of input. Regarding the former, many researchers (Lynch, 2009; Shin, 1998; Wagner, 2010) believe that accompanying oral input with visual input in many cases enhances test takers' comprehension by providing

more contextual information. Regarding the latter, correspondence with Wagner through email convinced us to use partial transcription of the conversations as the first hint. Moreover, the inclusion of some pictures related to the content of the conversations as well as a very short explanation of the content acted as the second hint. As mentioned before, the third hint asks the learners to start a dialogue with the tester in order to receive online mediation.

After the computerized dynamic test of listening comprehension was prepared, it was administered with few test takers to make further modifications to items instruction, test rubric, and the expected response. Additionally, the open-ended dialogue that occurred between the researchers as assessors and test takers in the third hint enabled the former group to prepare some leading questions and prompts so that language teachers could later use them when they gave the test to their own students. This format of DA in which two standardized forms of mediation (the first two hints) and one open-ended form of mediation (seeking out direct help from teacher) were integrated allowed us to use the benefits associated with both interactionist and interventionist approaches to DA.

Based on Aljafraah and Lantolf (1994), the ZPD-based hints should have two important characteristics. First they should be contingent and second they should be graduated. In order to meet the former, the hints were presented to the test takers only when they needed them, that is, they did not have access to all of them at once. To meet the second characteristic, the hints were arranged from the most implicit to the most explicit. In the first hint, only the partial transcription of the conversations is provided while in the second hint, the test takers are provided with some pictorial information plus some explanation in Persian, and in the third hint, they can get into an open-ended spoken dialogue in Persian with their teacher concerning the correct response. It should be noted that at this stage, the assessors were instructed to provide mediation (scaffolding) to the point where the test takers were on the brink of getting to the right answer. However, they were told not to reveal the correct answer. So, in the case of some learners, a very implicit probing question was enough to help them find the correct answer while in the case of some other test takers, the assessors had to be more explicit. In sum, a total of three hints for each item are included and in case the test taker fails to get to the correct answer after receiving the third hint, the correct answer together with a short explanation about it is provided, and then the next test item is presented.

A Sample Item Along With Three Hints

To better understand the process of mediation provision in the course of three hints, item number one is presented below. As mentioned before, each item consists of a conversation as the language input, an aural (as well as a written) question, and finally a blank space to be filled out with the correct response which is usually a word or a short phrase.

The topic of the first conversation is about *personality*. Here the test takers listen to a conversation between two friends. This item like other items in the test aims at examining test takers' global comprehension of the conversation, as this is the main goal of ninth grade textbook. Here is the conversation:

A: Who is that man?
 B: He is my uncle.
 A: What's he like?
 B: Oh, he's very funny and clever
 A: Is he nervous too?
 B: No, he isn't.

The question that follows the first conversation is as follows:

Q1: His uncle is funny but he is not.....

Upon giving a wrong answer, the hints are presented as follows:

Hint 1: Your answer is wrong; please read the partial transcription of the conversation, and then listen to the conversation once again.

Here is a screenshot of the item number 1 together with the first hint in Persian



Figure1

A screenshot of the item number 1 together with the first hint in Persian

Hint 2: Your answer is wrong. Please pay attention to the pictures which are related to the conversation and also to the short explanation about the conversation, and then listen to the conversation.



Figure2

A screenshot of the item number 1 together with the second hint in Persian

Hint 3: Your answer is still wrong. Ask your teacher for help and then listen to the conversation one more time.



Figure3

A screenshot of the item number 1 together with the third hint in Persian

Below is a dialogue that happened between Ahmad, one test taker and his teacher when Ahmad refers to his teacher for direct help:

Ahmad: *what's wrong here? I typed in three personality traits but the computer says my answer is wrong!*

Teacher: *you have only tried positive traits but in this item, you need to put a negative personality trait like "angry", "careless"...*

Ahmad: *Aha, thank you so much*

At this junction, the student got to the right answer simply by being reminded of what specifically he needed to focus on while listening to the conversation for the last time; especially when the teacher told him to expect a negative personality trait because the sentence containing the blank is negative: *Her uncle is funny but he is not...* .If the third hint which is in the form of direct help from the teacher does not lead to the correct response, test takers are given the correct response along with a short explanation to make the right answer clear for the test takers. At this point, test takers are presented with the next question.

Data Collection Procedure

Since one purpose of the study was to include participants from different public schools, it was not possible to have all the test takers take the test at the same time. Accordingly, the software was sent to 13 language teachers via email along with some instructions concerning its installation. All 13 teachers opted to install the software on their laptops because the schools where they were teaching at the time of data collection did not have computer rooms except three schools. Although these schools were equipped with computer rooms, there were other problems such as lack of speakers or headphones. Therefore, against our expectation, the computerized dynamic test which

was originally intended to be administered with large number of students at the same time was administered individually. On average, every teacher could give the test to seven students. Apart from the description that the learners received in the second page of the software, we asked the teachers to present a five-minute description on computerized dynamic test of L2 listening and the way the students should answer the test. Having administered the test to their students, teachers were asked to send the scoring files generated by the software to the e-mail address through which the software had been sent. The data of the study which consists of computer-generated scoring files were gathered in April and May, 2017.

Analysis

Immediately upon completion of the test by the test takers, two scores are generated by the software: an actual score which is based on the test takers' first try of each item and a mediated score which is based on the average number of hints used in each item. The actual score indicates test taker's independent performance without receiving mediation. In the case of the actual score, test takers obtain either 4 points or zero depending upon whether they gave a correct or incorrect response to each item. Thus, a test taker who answers all the items correctly will receive the maximum score of forty eight. However, the mediated score is not the matter of either and or. Instead, test takers' assisted performance which is based on the number of hints he/she has used is taken into account. For each test item, test takers may receive 1, 2, 3, 4 and less probably zero points depending upon the number of hints he/she has used. Put it another way, for each item, if no hint is used, the maximum of 4 points is awarded, 1 hint, 3 points; 2 hints, two points; 3 hints, 1 point; and finally the correct response is provided by the software and no point is given.

In addition to actual and mediated scores, learning potential score (LPS) was also calculated using Kozulin and Garb's (2002) formula whose purpose was to infer test takers' responsiveness to mediation. LPS obtained through this formula makes it possible to account for the change in scores (e.g. the difference between actual and mediated scores) relative to the maximum possible score on the test when test takers receive mediation. The formula is as follows:

$$LPS = \frac{(spost - spre)}{MaxS} + \frac{Spost}{MaxS}$$

Where

Spost = mediated scores

Spre = actual scores

MaxS = the highest mediated score gained in this test

As Poehner et al. (2015) argue, learning potential as represented through LPS should not be considered as a fixed trait like traditional conceptualization of intelligence and aptitude. Instead, LPS should be considered only as an indicator of test takers' responsiveness to mediation. Thus, LPS can provide insights into how much effort a test taker needs to move to the next level of development. Moreover, according to Poehner and Lantolf (2013), LPS is related to test takers' performance on transfer items. In other

words, test takers with higher LPS are expected to do better on transfer items than those with lower LPS.

FINDINGS AND DISCUSSION

In Line with Vygotsky's notion of ZPD, the major concern of DA is to ascertain whether test takers can improve their performance (e.g. listening ability) when provided with mediation. In fact, this concern is addressed in our first research question: can students obtain remarkable gain scores after receiving mediation?

Table 1 presents the descriptive statistics for the four types of scores obtained on the dynamic listening test. The comparison between the actual and mediated mean scores indicates that test takers could improve their scores by 13.54 points (i.e. mean gain score). As indicated by standard deviations of actual and mediated scores, test takers' performance became more homogeneous in dynamic test.

Table 1

Descriptive statistics of four types of scores on the listening test

	Actual Score	Mediated Score	LPS	Gain score
N	91	91	91	91
Mean	21.18	34.69	1.00	13.54
Std. Deviation	10.74	6.41	.10	5.41
Minimum	.00	18.00	.75	.00
Maximum	48.00	48.00	1.20	24.00

The results of paired-samples t test indicated that the difference between test takers' independent and mediated performance is significant ($t = 23.74$, $p < .001$, Cohen's $d = 1.52$). That is, test takers could substantially increase their scores after they received mediation. However, as Poehner et al. (2015) put it, "the point of DA is not to improve student test scores but to attempt a diagnosis of actual and potential, or proximal, development" (p. 347). This is in line with Vygostky's contention that true diagnosis should take into account both ZAD and ZPD.

The second research question deals with reliability and validity of the test as conceptualized in DA literature which is fundamentally different from traditional conceptualization of validity and reliability. Reliability coefficient was only calculated for test takers' scores on the traditional test since DA seeks change and development in test takers' performance not stability. The items had a high level of internal consistency, as estimated by Cronbach' Alpha of .79. As Poehner (2008) cogently puts it "While reliability may be a desirable characteristic in NDA [non-dynamic assessment], it is a highly undesirable outcome of a DA procedure, which seeks to bring about change" (p. 73). This idea reflects Vygotsky' (1998) contention that we should understand individuals not to measure them.

With regard to validity of a listening test, Wagner (2014) argues that Bachman and Palmer's (1996) framework of language task characteristics especially characteristics of the input, and characteristics of the expected response should be taken into consideration in order to make the test task characteristics similar to and representative

of the TLU domain. In the case of criterion-referenced listening testing, “the criteria to be assessed will dictate the characteristics of the test task” (Wagner, 2014, p. 52). In the context of the present study, conversations as well as the functions and notions of each lesson of grade nine textbook were considered as the TLU domain, and attempt was made to make the test task characteristics as similar to and as representative of the TLU domain by including the same functions, key words and expressions, and grammatical structures that existed in grade nine textbook. According to Wagner (2014), this matching will minimize threats to the construct validity of the listening test. With DA, however, the concept of validity ought to take into account another construct as well; that is, development. In other words, the construct validity of a dynamic test is understood as the extent to which it is capable of helping the test takers develop. In this regard, Poehner (2008) states “...the validity of a DA procedure is best understood as the extent to which it promotes development” (p. 76). Drawing on Messick’s (1988) conceptualization of consequential validity, Poehner (2008) argues that the opportunities that are awarded or denied from learners should be taken as evidence of construct validity. That said, if DA is capable of bringing about significant change in test takers’ performance through provision of mediation, it is valid. In the case of the present study, the results of the paired samples *t* test indicated that the difference between the mean mediated and mean actual scores (as indicated in the mean gain score) is statistically significant. The effect size was calculated using the automatic calculator for the effect size available at <http://www.socscistatistics.com/effectsize/Default3.aspx> (accessed November 10, 2016). The effect size (1.52) was large for the difference between two groups of scores and cannot be considered trivial.

In order to test the concurrent validity of the listening test, the Pearson product moment correlation coefficient was run between mediated and actual scores. The results of the analysis indicated that there is a statistically significant correlation between two sets of scores ($r = .92, p < .001$) accounting for the concurrent validity of the test. This strong correlation indicates that test takers with higher actual scores generally obtained higher scores in dynamic test as well. On the other hand, test takers with lower traditional scores benefited more from mediation as indicated by their gain scores. This, however, is not surprising since “those with high actual scores had less room for improvement when mediation was offered” (Poehner et al., 2015, p. 348).

The third research question is concerned with test takers’ learning potential which is actualized through the concept of LPS. This question specifically asks whether test takers with the same actual scores benefit differently from mediation. Aside from the fact that mediation in general improved the overall performance of test takers, LPS scores indicated that different test takers benefited from mediation in different degrees. Kozulin and Garb (2004) classified LPS scores in three categories: high (≥ 1), mid (.71-.8) and low ($\leq .70$). The lowest and highest LPS scores in this study were .75 and 1.20, respectively. Those with higher LPS scores are more open and responsive to mediation than those with lower scores. The pedagogical implication of this variation in LPS scores is that different test takers require different degrees of instructional support for development (Poehner et al., 2015).

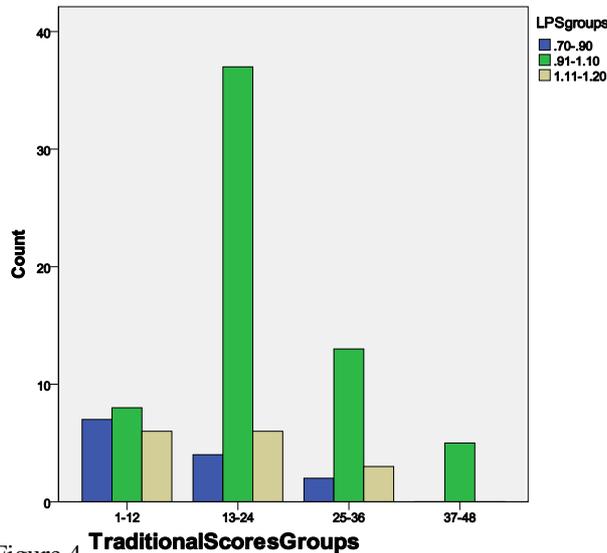


Figure 4
Test takers' LPS scores across different actual scores

As indicated in Figure 4, test takers' actual scores are grouped into four categories on the horizontal axis while test takers' LPS scores are grouped into three categories with three different colors. In order to understand how LPS works, take into consideration the first category on the horizontal axis. This category indicates the lowest actual scores obtained by test takers which range from 1 to 12. These low traditional scores, however, by no means mean that their learning potential scores are low as well. Rather, within this category of actual scores, there are test takers with high, medium, and low LPSs. The same pattern can be observed in other categories of actual scores. In other words, LPS scores and actual scores cross each other. However, it should be noted that LPS scores as well as mediated scores do not directly reflect learning, but they indicate the degree to which test takers are responsive to mediation. Indeed, responsiveness to mediation is a mechanism capable of creating a potential for microgenesis, or learning over a course of a single session (Poehner & Lantolf, 2013).

To illustrate the concept of LPS more clearly, consider two test takers in this study whose actual scores were the same; say, for example, 16. While one test taker could obtain a mediated score of 35, the other test taker could increase his mediated score only 10 points and obtain a mediated score of 26. The former could obtain an LPS of 1.12 while the latter gained an LPS of .75. Accordingly, it is likely that test takers with the same actual scores obtain different mediated and LPS scores. This indicates that test takers with the same level of independent performance might perform quite differently when mediation is available.

The last research question aimed at uncovering the relationship between the difficulty level of the test items and the kind of mediation used by test takers. Results of the study showed that there was a significant relationship between the difficulty level of the items and the type of mediation used by the test takers. The difficulty level of the items was marked by the length of the conversations. The second six conversations were double the length of the first six conversations. As indicated in Table 2., more than half of the test takers could answer the first six questions without receiving any mediation. Hint 1 which was in the form of partial transcription of the conversation ranked as the most useful since about thirty percent of the test takers could get to the correct answer using this hint. Indeed, the most implicit hint was proved the most useful type of mediation for the test takers in the first six questions.

Table 2

The average number and percentage of hint used by test takers on the first half of the test

	No hint	Hint1	Hint2	Hint3
Count	50.66	27.16	10	3.16
percentage	55.67%	29.84%	10.98%	3.47%

However, the pattern of mediation use was reversed in the second half of the test in which the conversations were about double the length of those of the first half. In this case (see Table 3.), the third hint asking the test takers to enter into dialogue with the testers proved the most useful type of mediation. It is obvious that the more challenging questions of the second half of the test made the test takers ask for direct help from the testers. In other words, the flexible nature of mediation of the third hint proved more useful than the first two hints which were fixed.

Table 3

The average number and percentage of hint use by test takers on the second half of the test

	No hint	Hint1	Hint2	Hint3
Count	29	13.83	20.5	27.66
percentage	31.86%	15.19%	22.52%	30.39%

Furthermore, as it is evident in Tables 2 and 3, the overall results of the comparison of the learners' response patterns in the first and second half of the test revealed that they have regressed in the second part which was more difficult and demanding than the first part. This fact indicated that they generally could not transfer their learning to the more challenging tasks. However when we compared the results of the two parts by the individual learners we found instances of progress. Though not many, there were some learners who performed better in the second part of the test. In order to show instances of progress across individuals we compared the response patterns of two learners with almost the same performance in terms of the overall number of the hints used in the first part of the test. As indicated in Table 4 learner 1 had regressed in answering the second part of the test requesting direct help from the mediator in half of the questions in the second part. Moreover, the overall number of hints used increased from 10 in the first part to 14 in the second part. However, learner 2 could answer 3 of the 6 questions in

the second part without using any hints. The number of hints used decreased from 11 in the first part to 6 in the second part indicating transfer of learning in the case of the second learner. This is in line with Poehner (2007) who reported both instances of progress and regression in the transfer section of his DA study.

Another explanation for students' use of more hints in the second part of the test is that the first six items were quite easy for them so that 55.67% of test takers answered these items correctly without any hint and 29.84% could answer these items correctly simply by being reminded that their first answer was not correct (first hint). Indeed, 85.51% of test takers did not receive any mediation in the real sense of the word. Put it another way, the first part of the test was rather within their ZAD. As such, no significant mediation was internalized in the first half of the test to be transferred to the second half of the test. We can argue that the real dynamic test begins with the second half of the test which includes difficult items. Introducing more difficult items is consistent with Vygotsky (1987) contention that when individuals are pushed to attempt more difficult problems, they will respond differently. In other words, such difficult problems can reveal learner's ZPD better than those problems or items which are easy. Therefore, we included more difficult items in order to understand test takers' responsiveness to mediation.

Table 4

The number of hints used by two learners across the questions

Question number		1	2	3	4	5	6	7	8	9	10	11	12
Number of hints used	Learner1	2	1	1	2	2	2	1	3	3	2	3	2
	Learner2	1	1	2	2	3	2	1	0	0	0	2	3

To illustrate the process of dialogic interaction (hint 3) between the assessors and the test takers, consider items 8 and 12 answered by students A and B respectively. The stem of item 8 is "The receptionist asks about Paul's wife to". The correct answer is "to check her passport". Here, the test taker A figures out the right answer using the first probing statement from the teacher.

Item 8

Student A: Sorry teacher. I typed the phrase "check in" but the computer says, "it is wrong".

Teacher: You are close to the answer. Look, when a tourist wants to check in a hotel, an ID document like "Shenasnameh" needs to be....

Student A: I think the answer has to do with "passport": to check the passport

Teacher: Exactly

In contrast to test taker A, test taker B attempting item 12 received three tokens of feedback before getting to the correct answer. The stem of item 12 is as follows: "Reza stayed home for two weeks last winter because he.....". The correct answer is "twisted his ankle".

Item 12

Student: My answer is wrong again. Maybe the problem is with my spelling

Teacher: No, the problem is that you are using an incorrect part of speech

Student: What do you mean?

Teacher: You need to use a “verb” here instead of a noun

Student: Ok, I think I should type the verb “broke his leg”

Teacher: No, this verb describes what happened to Reza’s friend. Please listen carefully this time to figure out what happened to Reza himself.

As these two examples indicate, test takers’ responsiveness to dialogic interaction (e.g. external mediation) can shed light on their emerging capabilities or what Vygotsky (1978) calls ZPD. While test taker A gets to the correct answer using the first token of feedback, the other one uses three tokens of feedback before getting to the correct answer. Thus, we can say that the first test taker is more responsive to external mediation and hence more prone to self-regulation than the second test taker who seems to be more dependent on other regulation.

As Poehner et al. (2015) argue, when reporting the outcomes of computerized dynamic test, it is necessary to take into account different types of scores. Sole reliance on one type of score to the exclusion of other scores might be misleading. Accordingly, actual scores can be used to understand abilities that have fully been developed at the time of assessment. However, mediated scores together with gain scores are related to ZPD or the fact that how responsive the test takers have been to mediation. Similarly, LPS scores indicate test takers’ responsiveness to mediation but in relation to the maximum mediated and actual scores. With LPS, even test takers who had high actual scores and consequently lower gain scores could obtain high and at times perfect LPS because this score takes into account the maximum score on the test as a basis for calculating learning potential score. It is important to know that how all these scores especially LPS and mediated scores can inform instruction. In other words, how can language teachers, learners and instruction in general benefit from these scores? By determining test takers’ approximate ZPD through mediated score and LPS, language teachers can figure out how much effort would be required by the learners to internalize and self-regulate the cognitive process in question (e.g. listening comprehension). For example, test takers with higher LPS scores are those who better respond to external mediation and therefore language teachers are expected to prepare and provide learners with useful mediation before they can move from other-regulation to self-regulation.

CONCLUSION

In recent years, the use of computerized DA in L2 context has been on the rise (Ahmadi & Barabadi, 2014; Lantolf & Poehner, 2012; Pishghadam & Barabadi, 2012; Poehner et al., 2015) because it has the potential to increase the efficiency of DA especially with regard to its application to a large number of individuals. On the one hand, by taking into account both ZAD and ZPD, it gives us information not only about what learners have already acquired and accomplished but also what they would achieve if they are given mediation. Specifically, LPS can inform us of the amount of external mediation needed before a learner can reach independent performance. In this way, it is possible to prepare tasks with varying levels of difficulty for different learners; tasks which would neither be boring nor demanding but challenging enough to motivate them. On

the other hand, using standardized type of mediation in the form of graduated hints, DA can be applied to a large number of test takers. This, however, poses a serious challenge to computerized DA. The mediation may not prove useful for all test takers due to its standardized and fixed format. The current study was an attempt to examine the effectiveness of the integration of interventionist and interactionist approaches to DA in developing Iranian EFL learners' listening ability. To this aim, a computerized dynamic test was developed. The results of the study indicated that while having acceptable level of validity, the computerized dynamic test could improve test takers' listening ability significantly. By providing LPS scores, the test was also able to provide information about test takers' potentiality for learning. Besides, the pattern of mediation use across different test items indicated that with more difficult items, test takers tended to make more use of the third hint which was more flexible in nature. In the case of easier items, the pre-determined graduated hints were helpful enough for test takers to get to the right answer.

Given the fact that DA provides information regarding both test takers' actual development (i.e. past achievement) and their potential for learning by taking into account test takers' ZPD, language teachers are recommended to incorporate DA into their assessment repertoire in order to obtain a more representative picture of learners' abilities. This should be a picture that takes into account not only the current developed capabilities but also the emerging and maturing ones. Besides, C-DA allows teachers to track learners' errors and to determine the amount of mediation they received. Upon receiving such information which is generated automatically through computer, teachers can make effective instructional intervention for their learners.

One of the limitations of this study concerned the preparation of hints. They were prepared based on guidelines of some testing books. Further research is needed to prepare hints in response to the errors students would make on the original non-dynamic test. This would increase the chances of working within students' ZPD. Moreover, in this study, the effect of C-DA was examined on listening comprehension. Other studies can be conducted in relation to other language skills and sub-skills e.g. reading, writing, and grammar. Another shortcoming of this study is related to the lack of transcendence tasks (TR). In fact, students' performance on such tasks can uncover test takers' real development as intended by dynamic assessment. Finally, the items of the study were not constructed based on the psychometric features of assessment; rather, they were taken from a commercial publisher named Khate-Sefid. Other researchers can use more robust methods of constructing items before applying them in a dynamic assessment situation.

REFERENCES

Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension. In J. P. Lantolf & M. E. Poehner (Eds.). *Sociocultural theory and the teaching of second languages*. London: Equinox.

- Ahmadi, A., & Barabadi, E. (2014). Examining Iranian EFL Learners' Knowledge of Grammar through a Computerized Dynamic Test. *Issues in Language Teaching*, 3/2, 183-161.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Ebadi, S., & Saeedian, A. (2016). Exploring transcendence in EFL learners' reading comprehension through computerized dynamic assessment. *Iranian Journal of Language Teaching Research*, 4(1), 27-45.
- Elliott, J. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review*, 55/1, 15-32.
- Haywood, H. C. & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge University Press.
- Kamrood, A. M., Davoudi, M., Amirian, S. M. R., & Ghaniabadi, S. (2018). Transcendence of Learning in an Online Computerized Dynamic Test of English Listening. *CALL-EJ*, 19(1), 23-42.
- Kozulin, A. & Garb, E. (2004). Dynamic assessment of literacy: English as a third Language. *European Journal of Psychology of Education*, 19/1, 65-77.
- Kozulin, A. & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23/1, 112-127.
- Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education. Vygotskian praxis and the research/practice divide*. London: Routledge.
- Lidz, C. & Elliott, J. (2006). Use of dynamic assessment with gifted students. *Gifted Education International*, 21:151-161.
- Lynch, T. (2009). *Teaching second language listening*. Oxford, England: Oxford University Press.
- Pishghadam, R., & Barabadi, E. (2012). Constructing and validating computerized dynamic assessment of 12 reading comprehension. *Iranian Journal of Applied Linguistics (IJAL)*, 15/1, 73-95.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer.
- Poehner, M. E. & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9: 233-265.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, 17/3, 323-342.

- Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language testing*, 32(3), 337-357.
- Shabani, K. (2014). Dynamic assessment of L2 listening comprehension in transcendence tasks. *Procedia-Social and Behavioral Sciences*, 98:1729-1737.
- Shin, D. (1998). Using videotaped lectures for testing academic language. *International Journal of Listening*, 12: 56-79.
- Vygotsky, L. (1978). *Mind in society*. US: President and Fellows of Harvard College.
- Vygotsky, L. S. (1987). In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky*, vol. 1: *Problems of general psychology, including the volume thinking and speech*. New York: Plenum.
- Vygotsky, L. S. (1998). The problem of age. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky*, vol. 5: *Child psychology* (pp. 187-206). New York: Plenum.
- Wagner, E. (2014). Assessing listening. In A. J. Kunnan (Ed.) *The companion to language assessment* (pp. 1-17). John Wiley & Sons
- Wagner, E. (2010). The effect of the use of video texts on ESL listening testtaker performance. *Language Testing*, 27: 493-513.
- Wang, P. (2015). The Effect of Dynamic Assessment 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.