

Examining Iranian EFL Learners' Knowledge of Grammar through a Computerized Dynamic Test

Alireza Ahmadi

Associate professor of TEFL, Shiraz University, Iran

Elyas Barabadi

Ph.D. Candidate of TEFL, Shiraz University, Iran

Received: January 25, 2014; **Accepted:** October 27, 2014

Abstract

Dynamic assessment (DA) which is rooted in Vygotsky's (1978) sociocultural theory involves the integration of instruction and assessment in a dialectical way to achieve two main purposes: enhancing learners' development and understanding about their learning potential. However, the feasibility and appropriateness of mediation are two main concerns of DA. The former is concerned with the application of DA for a large number of students, while the latter is concerned with providing test takers with appropriate hints. The purpose of the current study was three-fold: to examine the difference between dynamic and nondynamic tests, to understand about test takers' potential for learning, and to find out how mediation works for high and low ability students. To achieve these aims, computer software was developed. The software is capable of both providing the test takers with graduated hints for each item automatically, and adapting the overall difficulty level of the test to the test takers' proficiency level. To test the efficiency of the software in employing dynamic assessment, 83 Iranian university students participated in the study. The results of the study indicated that the computerized dynamic test made significant contribution both to enhancing students' grammar ability and to obtaining information about their potential for learning. Based on the findings of the study, it can be concluded that the use of dynamic assessment can simultaneously lead to the development of the test takers' ability and provide a more comprehensive picture of learning potential. Accordingly, teachers are recommended to use dynamic assessment to make more informed decisions about their students.

Keywords: Vygotsky, sociocultural theory, dynamic assessment, computerized test, Iranian learners.

Authors' emails: arahmadi@shirazu.ac.ir; elyasbarabadi@yahoo.com

INTRODUCTION

Dynamic assessment (DA) is rooted in the innovative ideas of Russian psychologist, Vygotsky (1978) who held the belief that assessment and instruction should be merged into a unified activity. The integration of assessment and instruction not only promotes learners' development but also paints a more comprehensive picture of learners' abilities; namely, both their zone of actual development (ZAD) and zone of proximal development (ZPD). Given DA is not solely concerned with what students have acquired in the past and its main concern is with learners' potential for learning and their development through integration of assessment and instruction, it is a big advantage for learners. However, it has not been put into widespread use since most DA studies conducted so far have been case studies in which few participants could take the dynamic test (Ableeva, 2008; Birjandi & Ebadi, 2010; Lantolf & Poehner, 2008, 2004; Tajeddin & Tayebipour, 2012).

The computerized delivery of mediation in DA has been suggested as a solution for its narrowness of scope (Poehner, 2008). Pishghadam and Barabadi (2012) and Toe (2012) reported on the feasibility and effectiveness of computerized delivery of mediation in assessing test takers' reading comprehension. Targeting reading and listening skills through computerized dynamic assessment, Poehner, Zhang, and Lu's (2014) study also indicated that DA was capable of providing fine-grained diagnosis of test takers' development in two domains of reading and listening. To the best of the researchers' knowledge, test takers' grammatical knowledge has not been dealt with through computerized dynamic assessment (C-DA). Accordingly, this study was an attempt to dynamically assess and promote the grammatical knowledge of Iranian EFL learners via computer software in order to get around the major shortcoming of DA; that is, its narrowness of scope in terms of the number of participants. Nonetheless, C-DA poses another problem which is not existent in noncomputerized DA; namely, tailoring mediation to test takers' needs. In fact, electronically delivering mediation is not sensitive enough to test takers' ZPD in such a way that for some test takers, the test might be very easy while for others, the mediation might not be intelligible, and hence makes no contribution at all. Regarding this issue, Poehner (2008) observes "C-DA like other interventionist approaches has limitation on the kind and quality of mediation it offers. Indeed, mediation cannot be attuned to learner's

needs" (p. 177). Therefore, another main objective of this study was to address this problem by adjusting the overall difficulty of the test with test takers' proficiency level. In what follows, first the use of C-DA in L2 context is reviewed and then Kozulin and Garb's (2002) Learning Potential Score, which is used to assess test takers' potential for learning, is explained.

LITERATURE REVIEW

Computer-based DA

Before dealing with computer-based DA, it seems necessary to discuss some issues related to the application of DA. As mentioned before, though DA offers several advantages over traditional tests, it poses a number of acute problems. For instance, Hasson and Joffe (2007) note that DA approaches have been criticized for lack of inter-rater reliability. According to Haney and Evans (1999) other problems are related to lack of adequate knowledge base and expertise in this field and also time constraints. They conducted a survey to explore the issues related to the use of DA. The result of the survey showed that only half of the school psychologists were familiar with DA procedures and only half of them actually implemented DA. The result also indicated that school psychologists mostly used traditional assessment tools at schools. They did so due to the lack of adequate knowledge base about DA and time restraints. It has also been stated that DA practitioners must develop subjective judgment concerning what cognitive functions require mediation and to what extent (Haywood & Tzuriel, 2002). To sum it up, there are some problems with DA in general and interactionist DA in particular:

It is highly time consuming;

It requires a lot of expertise on the part of the test user (teachers);

It lacks inter-rater reliability.

In recent years, the use of C-DA has been considered a solution to overcome these shortcomings. In his discussion of advantages of C-DA, Poehner (2008, p.177) mentions the following points which are not achievable via noncomputerized forms of DA:

1. It can be simultaneously administered to a large number of learners.
2. Individuals may be reassessed as frequently as needed.

3. Report of each learner's performance is automatically generated.

In order to cope with the main shortcoming of DA; that is, its narrowness of scope, Pishghadam and Barabadi (2012) examined the effectiveness of conducting a computerized dynamic reading comprehension test (CDRT) on EFL learners. They designed software capable of providing predetermined hints in case test takers committed an error while answering reading comprehension questions. This computer program enabled them to test many university students by providing systematic and controlled mediation. Their sample consisted of 77 university students with moderate language proficiency. The results of their study in line with other DA studies in L2 context indicated that DA is useful not only in enhancing test takers' reading ability but also it can provide useful information regarding students' potential for learning.

Likewise, Teo (2012) developed a C-DA program that integrated mediation with assessment to support learners' inferential reading skills. 68 Taiwanese college EFL learners participated in her study. There were four levels of mediation in the C-DA program. The mediations progressed gradually from implicit to explicit. After reading each passage, the participants were asked one inferential question, and they had to choose one of the five given choices. In case they made a mistake, they were provided with mediation until they could answer the question correctly. The results of her study indicated that C-DA was a powerful tool in understanding about participants' potential for learning. Moreover, C-DA program became a valuable resource for her to create an effective one-on-one mediated learning environment facilitating individualized instruction.

Extending the use of C-DA to reading and listening, Poehner and Lantolf (2013) and Poehner, Zhang, and Lu (2014) delivered listening and reading comprehension tests in an online format. The researchers reported on the use of transfer items in order to emanine the effect of graduated propmts (mediation) on test takers' development of reading and listening comprehension. The three types of scores generated by the computerized dynamic tests helped the reseachers establish accuarate diagnosis of the test takers' L2 developemnt.

Learning Potential Score (LPS)

Kozulin & Garb (2002) carried out a study of dynamic assessment of text comprehension for adult EFL learners. The results of their study indicated that DA is capable of both assessing the current knowledge of students and their ability to benefit from mediation. However, the extent to which the test takers benefited from mediation varied from one test taker to another. In other words, some learners made more use of mediation than others. This was true for learners with different levels of proficiency. In order to account for the differing use of mediation by different learners in their study, Kozulin and Garb developed a formula to operationalize student learning potential:

$$LPS = \frac{(S_{post} - S_{pre})}{MaxS} + \frac{S_{post}}{MaxS}$$

where S_{pre} and S_{post} refer to nondynamic and dynamic scores respectively and $MaxS$ is a maximum obtainable score or the highest dynamic score on a given test. Using this formula, Kozulin and Garb's (2002) suggested that DA has the potential to be used as a way of unlocking the potential of individual test takers for future learning by taking into account their differing ability to learn with assistance.

PURPOSE OF THE STUDY

As mentioned earlier, the efficiency of computerized delivery of mediation in DA has been confirmed with regard to reading and listening comprehension by some researchers. The main purpose of the current study was to examine to what extent a computerized dynamic test of grammar can contribute to test takers' development of grammatical knowledge of L2. Besides, examining DA's ability to reveal test takers' potential for learning was another focus of the study. As such, the study aimed at answering the following research questions:

1. Is there any significant difference between the students' scores in computerized dynamic assessment and computerized nondynamic (traditional) assessment?
2. Is C-DA capable of revealing test takers' potential for learning?
3. How do the learning potentials of high and low knowledgeable learners differ through computerized mediation?

METHOD

Participants

The sample of the study consisted of 83 Iranian university students. The majority of the test takers were BA and MA students majoring in English (TEFL, Literature, and Linguistics). Of all the participants, there were only three PhD students in TEFL and six participants from non-English majors (e.g. Geochemistry and Political sciences). The reason why MA and PhD students were also included in the study was that based on the results of the pilot study, the second section of the test was found to be challenging even for MA students. The students who participated in the study were from various Iranian state-run universities, including Shiraz University, Tehran University, Ferdowsi University of Mashhad and Allameh Tabataba'i University. All the participants were between 18 and 34 years with a mean age of 28. They were selected on the basis of their availability and willingness to take the test. For all of the participants, Persian was their first language and English was their second language.

Instrumentation

The instrument used in this study is a software package which is capable of dynamically testing the grammatical knowledge of test takers by offering predetermined hints in case they make a mistake. The software is comprised of three parts: introduction, the main part or the dynamic tests and the scoring file. In the introduction part, the test takers are asked to fill out a form related to their personal characteristics such as age, gender, major, etc. The introduction also gives test takers a short description of DA. The main part consists of two dynamic grammar tests arranged in the order of difficulty. Each test has 20 items, and each item is followed by five hints in case the test taker cannot answer the item correctly. Finally, upon completion of the test, a scoring file with the following information is generated: two scores for each student (dynamic and nondynamic), the number of hints used for each item and the total time spent on the test.

Data Collection Procedure

In order to develop the software package capable of assessing students' grammatical knowledge in a dynamic-adaptive way, a three phase

procedure was followed: test preparation and piloting, software preparation, and administration of the test.

Test preparation

To prepare the items of the computerized grammar test, initially 50 grammar items were taken from the book *12 SAT Practice Tests* by Black and Anestis (2008, 2011). The reason we selected this book was that all the items in this book including grammar items are rated based on their difficulty level. Knowing the difficulty level was of high importance since the Dynamic Grammar Test developed in this study had a similar feature to adaptive tests in a sense that it consisted of two subtests arranged in the order of difficulty. Accordingly, the difficulty level of each item was the starting point for dividing the items into two subtests, namely, the easy and difficult test. To achieve this aim, items with difficulty level of one and two on the scale of five in this book were selected for the easy test while those with the difficulty level of four and five were selected for the difficult test. Items with the difficulty level of three were ignored because we wanted to make sure that the two versions of the test were really different especially in a DA test in which the provision of mediation diminishes the difference between easy and difficult items. Of the large number of grammar items in these two books, 50 items (25 easy and 25 difficult items) representing different grammatical points were selected for our purpose in this study. All these items were in MC format. However, they were rewritten into other formats to better serve the purpose of a DA test. The five types of questions used in this study were: 1) Identifying Error, 2) Filling in the blanks, 3) Specifying the additional word or phrase, 4) Writing the most appropriate form of the given word or phrase, and 5) Rephrasing the underlined part. Due to the changes made to the item format, it was likely that the difficulty level of the items might have changed from that mentioned for the original test; therefore, it was considered necessary to pilot the 50 item test with students of different proficiency levels in order to make sure that the difficult and easy tests were distinct enough. Hence, 26 Iranian learners of English took the test in its traditional paper and pencil format. Test piloting helped us be more specific concerning the difficulty level of items after changing their format.

Having given the test to these university students, the researchers of the study analyzed the items. The results of item analysis were

interesting because some of the items that were initially considered easy came to be difficult and the other way around. This seemed logical considering the change made to the format of the items and the fact that their original difficulty was decided judgmentally by the writers. As such, based on the difficulty level determined through the pilot study, the items had to be re-categorized. Items with difficulty level of .62 and above and .32 and below were selected for the difficult and easy tests respectively. Moreover, in order to make sure that these two tests were adequately different from each other in terms of difficulty, items with difficulty level between .32 and .62 were omitted (10 items). Accordingly, the final test used for the Computerized Dynamic Test of grammar was left with 40 items, 20 items for each subtest. The most important phase of test preparation from a DA perspective; that is, preparation of appropriate hints, followed item preparation. It was the most important since the main objective of DA which is the learner's development is totally dependent upon the quality of mediation (hints). For each question, five hints arranged from the most implicit to the most explicit were prepared. To prepare appropriate hints, the researchers of the study first benefited from the careful analysis of the test takers' responses and their feedback to each question in the piloting phase. At the same time, several well-known test books including Pamela's (2004) *12 SAT Practice Tests* series which contain a separate section named *Detailed Answer Key*, *Barron's How to Prepare for the TOEFL*, and Phillips' (2003) *Preparation Course for the TOEFL Test* were consulted. When the computerized dynamic test was fully prepared, it was piloted again with 20 EFL university students to study the effectiveness of the hints. Upon receiving feedback from them, the hints were reanalyzed and some adjustments were made to make them more understandable, and hence more attuned to test takers' ZPD. Ultimately, the final version of the test including the items as well as the hints was reviewed by two of the professors at Shiraz University, and some minor changes were made.

The Software Preparation

The software program used in this study was made using Visual Studio. This software consists of two different sections: in the first section the test takers are asked to fill out a form related to their personal characteristics including, name, major, degree, gender, age, and email address. The second section includes the tests. At first, the test takers are

presented with the easy test consisting of 20 questions. As mentioned before, the test takers are provided with predetermined hints arranged from the most implicit to the most explicit. If a given test taker could not answer a question correctly with the first four hints, the software would provide the correct answer in the fifth hint. The number of hints used in the first five questions of the easy test helps estimate the proficiency level of the test takers, and is the basis to decide whether the test taker should go on with the first test or be directed to the second test which is more difficult. On average, if a given test taker makes use of ten hints or below, the test is considered easy for that test taker, and he/she would be directed to the second test which is more difficult. In other words, for test takers whose average use of hints is two or below, the test is within their ZAD. Therefore, they need the second test which is more sensitive to their ZPD. This partial adaptation of the test takers' ability to the difficulty level of the test could partially obviate one of the main shortcomings of C-DA; namely, the nonsensitivity of mediation to test takers' ZPD.

The software has been designed in such a way that any PC can run it easily; it can be installed properly on any PC provided that NET Framework software is already installed. As soon as the test takers finish the test, a scoring file in Word format appears on the desktop which contains the following information:

1. The test taker's personal information.
2. Test taker's nondynamic score: This score is calculated according to the students' first attempt at each item. This score is calculated regardless of the number of hints the test taker used. However, in order to make it comparable with the dynamic score of the test, it is calculated on a scale of 0 to 100 points; five points for each item. For example, one test taker (Mina, a pseudonym) who answered five questions correctly on the difficult test using no hints earned a nondynamic score of 25.
3. Test takers' dynamic score: The number of hints used by test takers is the defining point for calculating their dynamic score. Since there are 100 hints for each test; five hints for each question, it is possible to calculate their dynamic score by subtracting the number of used hints from the total number of hints. Back to the test taker in the previous

example, her dynamic score on the difficult test was 59 since she had used 41 hints.

4. The number of hints used for each item.

Given that the software program is able to provide such information in a user-friendly manner, the process of data collection was not difficult for the researchers. Having access to the software, every test taker could run the program easily and take the test on his/her own. The following section deals specifically with the process of data collection.

At the outset of the study, it was scheduled preferably to have most of the participants, if not all, attend a two-hour meeting to take the test so that all the participants could work under the same conditions. However, since the university classes were closed for the end of the term break by the time the software was completed, most of the participants took the test individually. Only 11 participants could attend a two-hour meeting in language laboratory of Shiraz University and take the test together; the rest of the participants were given a choice of having the software e-mailed to them, or given to them in person. Having taken the test, the participants sent their scoring files to the researchers' emails.

Data Analysis

The data collected were analyzed using *t* test to determine the statistical significance of the difference between the dynamic and nondynamic mean scores. Also to understand about the strength of this difference, eta squared statistic was applied (Dornyei, 2007). Finally, the learning potential score (LPS) formula developed by Kosulin and Garb (2002) was used to estimate the learners' potential for learning.

RESULTS

Out Of 83 participants in this study, 38 took the easy test. In other words, these 38 participants' scores on the first five questions of the easy test were below 16 meaning that the first test was close to their ZPD, and hence appropriate for them. The remainder of the participants (45 participants) received a score of 16 or above meaning that the first test was within their zone of actual development (ZAD). Accordingly, they were directed to the more difficult test which was within their ZPD. In

what follows the results of the study are presented in three sections in line with the three research questions of the study.

Comparing the Participants' Scores in Computerized Dynamic Assessment and Computerized Nondynamic (Traditional) Assessment

Table 1 indicates the descriptive statistics for the test takers' performance on the easy test. Comparison of nondynamic gains with dynamic gains of the 38 test takers who took the easy test indicated a change of mean scores from 35.7 (S.D. = 5.64) to 63.9 (S.D. = 5.13). Likewise, as indicated in Table 2, the comparison of nondynamic and dynamic scores of the 45 students who took the difficult dynamic test indicated a change of mean scores from 35.11 (S.D. = 18.29) to 63.38 (S.D. = 15.02).

Table 1: Descriptive statistics and paired sample *t* test for the easy test

	<i>M</i>	<i>N</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
NDA	35.79	38	5.64	-28	3	.000
DA	63.97	38	5.13			

As Tables 1 and 2 indicate, it is evident that providing test takers with graduated hints via computerized dynamic test made great contribution to their grammatical knowledge and hence their significant increase in their dynamic scores. In order to determine the statistical significance of the difference between these two sets of scores in each test, paired sample *t* test was performed. The results (Table 1 & 2) show that there was a significant difference between the DA and NDA scores in both the easy and the difficult test ($P < .000$ for both tests).

Table 2: Descriptive statistics and paired sample *t* test for the difficult test

	<i>M</i>	<i>N</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
NDA	35.11	45	18.29	-25	44	.000
DA	63.38	45	15.02			

Although the results presented above indicated that the difference between DA and NDA scores was unlikely to occur by chance, we needed to make sure about the strength and magnitude of this difference. To achieve this aim, the effect size statistic was used. As suggested by Dornyei (2007), eta squared formula for calculating this statistic is appropriate. The effect size values were .95 and .93 for the easy and the difficult test, respectively. Based on Cohen (1988), the effect sizes for both tests were quite large indicating that there was a substantial difference between the dynamic and nondynamic scores.

Capability of the Computerized Dynamic Assessment in Revealing the Test Takers' Potential for Learning

Providing information concerning test takers' potential for further learning and development is another distinguishing feature of DA in comparison to traditional tests. The second research question specifically addressed the ability of DA to assess the size of students' ZPD. Using Kozulin and Garb's (2002) formula for calculating learning potential score (LPS), we tried to examine DA as a way of unlocking the potential of individual test takers for future learning by taking into account their differing ability to learn with assistance. Consider how LPS of the test taker mentioned in section 4.3. is calculated:

Mina's NDA score: 25

Her DA score: 60

The maximum DA score on the difficult test: 91

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

where S pre and S post were nondynamic and dynamic scores in our study, and Max was a maximum obtainable score or the highest dynamic score which was 91 in this case.

$$HerLPS = \frac{(60 - 25)}{91} + \frac{60}{91} = 1.04$$

As can be seen in Table 3, the test takers' LPSs on the easy test ranged from .86 to 1.46, and on the difficult test, from .63 to 1.37. In fact, LPS indicates that the improvement of test takers' performance on dynamic test was not equal. Thus, through this score, it was possible to

differentiate among test takers with the same NDA score. Those students who made considerable progress from nondynamic to dynamic test had high LPS, and those who made slow progress had low LPS. Once again, consider the test taker mentioned above with LPS of 1.04 on the difficult test. Another test taker with the same nondynamic score of 25 had an LPS of .88. So the two test takers were different in terms of their potential for learning though they had the same nondynamic score. Similarly, two test takers with the same nondynamic score of 40 progressed at different rate on dynamic test. One of them received an LPS of 1.06, and the other an LPS of 1.46.

Table 3: Descriptive Statistics of test takers' LPS on the easy and difficult test

Type of test	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
The easy test	38	.86	1.46	1.21	.13
The difficult test	45	.63	1.37	.99	.16

In order to see if LPS could differentiate among the learners with the same NDA score, we compared eight test takers with the same NDA score on the easy test. Figure 1 clearly shows how different these eight test takers are regarding their LPSs. If we consider those LPSs which lie between one and two standard deviations above the mean ($M = 1.34$ to 1.47) as high learning potential, and those LPSs which lie between one and two standard deviations below the mean as low learning potential ($M = 0.95$ to 1.08 to), it is evident that test takers' LPSs on this test were not the same. For example, consider the two test takers who scored 35 on nondynamic test. One could increase his DA score to 50 whereas the other could receive a DA score of 70. The differing gains of these two test takers are reflected in their LPSs which are .86 and 1.4 respectively. This shows that while from the point of view of a traditional test, grammatical knowledge of those students with the same NDA score is considered the same, the learners' LPS and in turn their dynamic scores could differentiate among them by considering their ZPD along with their ZAD.

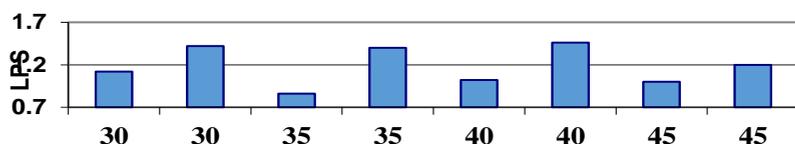


Figure 1: Distribution of learning potential scores among test takers with the same nondynamic score

Likewise, in order to show how C-DA was capable of discerning test takers' potential for learning on the difficult test, LPSs of ten students with the same nondynamic scores on the difficult test were compared (see Figure 2.). Again, if we consider those LPSs which lie between one and two standard deviations above the mean (1.15-1.31) as high learning potential, and those LPSs which lie between one and two standard deviations below the mean as low learning potential (0.83-0.67), a significant difference in their LPSs is observed.

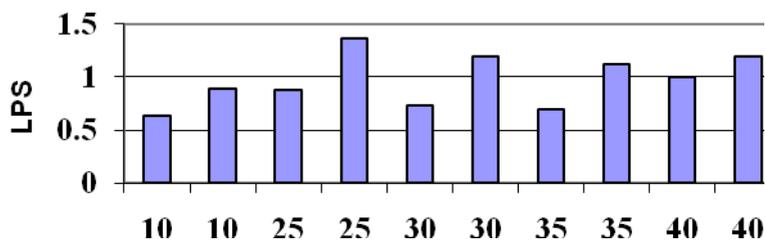


Figure 2: Distribution of learning potential scores among test takers with the same nondynamic score presented in Table 3

Learning Potentials of High and Low Knowledgeable Learners through Computerized Mediation

One of the main assumptions within the DA procedures is that mediation, will, in general be more effective for low achievers; no matter their low achievement is due to cultural, socio-economic or academic reasons (Peña, Iglesias & Lidz, 2001; Tzuriel & Kaufman, 1999). The third

research question specifically dealt with this issue by asking whether low and high proficiency level students benefited differently from mediation in the form of graduated hints. As indicated in Table 3, the mean LPS of those who took the easy test was 1.21 while the mean LPS of those who took the difficult test was .99. An independent-samples *t* test was conducted to compare the mean LPSs for those who took the easy and those who took the difficult test. As can be observed in Table 4, there was a statistically significant difference in mean LPSs of the two groups ($P < .000$). The magnitude of the difference was large (eta squared = .9).

Table 4: Independent Samples *t* Test for LPSs on the easy and difficult tests

	<i>t</i>	<i>df</i>	<i>p</i>	Std. Error Difference
LPS Equal variances assumed	6.44	80	.000	.03

Finally, it is worth noting that the computerized dynamic test developed in this study was partially adaptive since it could direct students to the second subtest which was more difficult in case that the first subtest was considered easy for them. In other words, if their average use of hints was less than 10 in the first five questions, the test was considered as easy by the software, and they were directed to the second test which was more difficult.

DISCUSSION

This study sought to explore the feasibility of computerized delivery of mediation in three ways: (a) whether there is any significant effect of DA procedure on test takers' grammatical ability, (b) whether DA is able to distinguish between test takers' potential and actual level of performance, and (c) whether high or low proficiency level students could make the maximum use of mediation provided in the form of hints.

Regarding the first research questions, the findings of the current study indicated that the computerized grammar test was able to improve the test takers' grammatical knowledge significantly. The results of this facet of the study are consistent with those obtained by other researchers in other areas of L2 such as reading comprehension (Pishghadam & Barabadi, 2012), reading and listening comprehension (Poehner & Lantolf, 2013) and pragmatics (Tajeddin & Tayebipour, 2012). All these

studies including the current one could create a supportive atmosphere aiming and prioritizing test takers' further learning and development by taking into account both test takers ZAD (zone of actual development) and ZPD. While traditional (non-dynamic) tests can only account for the intramental, self-regulated, and fully-internalized abilities of the test takers, DA takes into account not only these abilities but also those which are other-regulated (intermental). However, the significant gain of test takers from non-dynamic to dynamic test can be attributed to non-intellective factors. As Pishghadam and Barabadi's study (2012) indicated, non-intellective factors such as lack of motivation, fear of failure, and inattentiveness can be the cause of incorrect response by test takers. In the like manner, many test takers in this study could get to the right answer when they received the first two hints which were the most implicit. In other words, although the first two hints were rather independent of the grammatical point in question, they helped the test takers overcome these non-intellective factors that might have caused them lose the whole score in a non-dynamic test. Test takers' significant gain on dynamic tests of grammar can be considered as evidence for their construct validity. According to some DA practitioners (Haywood & Lidz, 2007; Lidz & Macrine, 2001; Poehner, 2008), construct validity is understood as the extent to which DA enhances individuals' development.

Concerning DA's capacity to provide information about the test takers' potential for learning (2ND research question), a discussion of LPS as proposed by Kozulin and Garb (2002) seems necessary. According to Kozulin and Garb (2002), a high LPS means that the learner's ZPD level is close to their ZAD level. That is, the targeted ability is on the verge of internalization or self-regulation. On the other hand, a low LPS shows that the test taker is in need of much more mediation and external help to internalize the learning point in question. In line with this conceptualization, it was indicated that the test takers with low LPS in this study made use of much more mediation in the form of hints than those test takers with high LPS. This pattern of results is in line with Kozulin and Garb's (2002). In their study, LPS could differentiate between the test takers with the same nondynamic score. Similarly, other DA researchers such as Poehner and Lantolf (2005) and Anton (2009), though not referring to the notion of LPS, reported in their studies that DA could differentiate between the test takers with the same score in

nondynamic tests. That said, if the primary purpose of language assessment as Bachman and Palmer (2010) cogently argue, is to provide information that will help make more informed decisions that in turn will lead to beneficial consequences for the stakeholder especially test takers, a strong point can be made for DA in general and our version of DA in particular.

Differentiation among test takers concerning their abilities and needs is not limited to LPS. By generating the scoring file for each test taker in which it is clear how many hints they have used in each question before they could get to the right answer, C-DA test of grammar enables L2 teachers to tailor their instruction to suit the specific needs of their learners. This result is in line with the claim made by Poehner, Zhang, and Lu (2014) who believe that C-DA can provide fine-grained diagnosis of test takers' L2 development. To illustrate, one of the test takers in this study, for example, used two hints on average in questions dealing with the verb tense. This shows that this aspect of language was on the verge of internalization. Hence, small amount of intervention or external help would suffice to move him from intermental plane to intramental plane. This same learner used four hints on average in questions dealing with parallel structures indicating that there was much room for the teacher to manoeuvre before this linguistic feature in question became internalized. So, DA as conceived in this study enables teachers to provide individualized instruction. Besides, knowing how many hints they have used for every question, "...learners may use diagnostic information from language assessment to make formative decisions about their own learning" (Bachman & Palmer, 2010, p. 87). Underlying processes used to answer a question by test takers can be considered part of this diagnosis. By tracking the learners' errors in terms of how many hints they used for each question, the software program can provide valuable clues about the processes of answering a question by test takers. Moreover, the total amount of time spent consulting the mediation (hints) was another advantage of this software. By knowing how much time a particular test taker used to get to the correct answer, we could understand about the comprehensibility of the hints for each test taker. However, these issues need more studies to delve into such advantages of C-DA.

The results of the current study indicated that there was a significant difference between the mean LPSs of those who took the easy and the

difficult test. In other words, the mediation brought greater benefit to the test takers who took the easy test. That low proficiency learners (those who took the easy test) made bigger gains in C-DA is in line with DA studies which indicate the relative superiority and usefulness of mediation for low-achievers than high achievers. Indeed, one of the main assumptions of DA is that individuals who have not received adequate mediated learning experience (e.g. low proficiency learners) in the past would benefit more from the mediation provided during DA sessions than those who had rich learning experiences (Haywood & Lidz, 2007; Tzuriel & Kaufman, 1999).

The results of the C-DA can be interpreted in the light of current views about validity which consider the process of test validation as building and substantiating an argument (Bachman & Palmer, 2010; Chapelle, 2012; Kane, 2011). Bachman and Palmer (2010), for instance, introduced an assessment use argument (AUA) model which is organized around a series of inferences that starts from test takers' performance to decisions which are made, and finally the consequences of those decisions. Here, we focus only on the inferential bridge between the test record and actual or intended interpretations about test takers' ability. In order to make any decisions about test takers, we need the results of an assessment which well represent the construct (e.g. grammatical knowledge) under question. In Bachman and Palmer's (2010) own words "when someone gives a language assessment he intends to interpret the performance on this assessment as an indicator of some aspect of the individual's language ability" (P. 89). Back to C-DA designed in the current study, and in concert with other DA studies (Ableeva, 2008; Anton, 2009; Birjandi & Ebadi, 2009; Kozulin & Garb, 2004; Lantolf & Poehner, 2008; Pishghadam & Barabadi, 2012), we believe that C-DA can provide us with a more comprehensive and precise profile of individuals' language ability by taking into account both their actual standing in a group based on their NDA scores, and their would-be (potential) standing based on their DA scores. However, it goes without saying that we did not design C-DA in this study through a thorough argumentation as proposed by Bachman and Palmer (2010) which includes four claims and their associated warrants. Our main objective was only to indicate that DA in general and C-DA in particular can lead to more valid inferences especially with regard to the inferential link from assessment records (test scores) to interpretations about test takers'

ability. To be more specific, test takers' LPS as described earlier can be a more valid indication of their ability than nondynamic scores which are solely based on their past achievement.

These two features of DA; that is, enhancing learners' development and providing information concerning their learning potential, can enable test developers and teachers to use assessment tools in what Shohamy (2005) calls "interactive, democratic, and constructive ways" (p. 101). The computerized dynamic test designed in this study, which was partially adaptive as well, like other forms of DA generated by other researchers can be characterized as interactive and constructive since the software provided test takers with mediation in the form of graduated hints helping them work out the grammatical problem. In other words, mediation can help learners construct their own knowledge of grammar. Also DA can be democratic for L2 learners and especially for L2 teachers. It will be democratic for learners since it tends to adopt a "present-to-future" (Valsiner, 2001) view toward their abilities. In other words, its main concern is with learners' potential for learning and helping them move forward no matter where they are standing at the time of assessment. As for L2 teachers, it should be noted that DA procedures do not "treat (L2) teachers as agents for carrying out orders"; instead, they empower teachers by letting them be authoritative and professional decision makers. In fact, DA can be considered as "alternative assessment procedure[s] that involve[s] teachers and are driven by teachers based on pedagogical considerations" (P. 101). Viewed from this perspective, DA can have a voice in teacher education and teacher professional development as well.

Ultimately, flexibility of computerized delivery of mediation is worth mentioning. According to Osolon-Buchanan and Drasgow (1999), "computer programming affords test developers the flexibility of dynamic selection of items to be presented and allows variations in the presentation of stimulus materials" (p. 2). In the present study, this feature; that is, flexibility, was actualized in a number of ways: 1) giving systematic mediation to test takers in case they made a mistake, 2) going beyond MC format by including other formats as discussed earlier, and more importantly, 3) the adjustment of the overall difficulty of the test to test takers' proficiency level. In fact, the software was capable of tailoring the overall difficulty of the test to the examinees' ability. Another advantage of the C-DA was related to the ease of administration

and scoring. Automatically providing mediation when needed and automatically generating the test taker' scoring file, the software program enabled the researchers to make DA more convenient, reliable, standardized, and affordable than noncomputerized DA. As such, it is possible to assess the ability of a large number of test takers dynamically in a standardized and systematic way.

CONCLUSION AND IMPLICATIONS

In our view, C-DA as designed in this study in which the overall difficulty of the dynamic test was adapted to the learners' proficiency level can be an innovation not only in the field of second language testing but also in the field of DA. C-DA is innovative in the field of DA since it enables teachers to assess a large number of students in a dynamic way at the same time. In fact, when computers can take over the role of expert mediators, DA no longer relies heavily on the presence of teachers and students in the classroom. Learners can interact with their computers as the expert mediator. Besides, by tracking learners' errors, C-DA enables both teachers and learners themselves to identify their strengths and weaknesses. Later, teachers can turn the focus of their instruction to their learners' problematic areas. C-DA allows for students' self assessment and reassessment; it encourages them to become part of the whole process of learning and assessment. So, with the availability of C-DA, students are no longer dependent upon teachers to be assessed and become aware of their progress; they can assess and reassess themselves as many times as needed.

On the other hand, C-DA is innovative in the field of L2 assessment by integrating instruction and assessment in order to boost learners' development, making assessment at the service of instruction not vice versa. Since DA procedures take into account both latent and developed capacities when assessing learners, it seems reasonable to suggest that DA be used along with traditional standardized tests. Therefore, it is important for teachers to recognize this important fact that the judicious use of these two types of assessment provides them with a more representative picture of learners' abilities; a picture that takes into account not only the current developed capabilities but also the emerging and maturing ones. Using the information obtained through DA, teachers need to understand how to avoid overestimating and underestimating their learners' abilities.

As mentioned before, the dynamic test used in this study was partially adaptive. In other words, the decision about the test-takers' proficiency was based on the first few (5) items. In fact, passing or failing these 5 items was arbitrarily determined as kind of cut-off point by the researchers of this study. Other researchers can think of developing a true dynamic CAT by adjusting item difficulty based on the test takers' response. However, this could be very challenging as the test should be both adaptive and provide mediations.

Finally, it should be mentioned that the findings of the study should be treated with caution due to the sampling employed. It is possible that the sample used in this study is not representative of the general population of the Iranian EFL learners since we used only those test takers who were at our disposal and expressed willingness to participate in the study.

Bio-data

Alireza Ahmadi is an associate professor of TEFL in the Department of Foreign Languages and Linguistics at Shiraz University, Iran. His main interests are language assessment and second language acquisition.

Elyas Barabadi is currently a Ph.D. student of TEFL in the Department of Foreign Languages and Linguistics at Shiraz University, Iran. His main interests are language assessment and second language acquisition.

References

- Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension. In J. P. Lantolf and M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 57-86). London: Equinox.
- Anton, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576-598.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Birjandi, P., & Ebadi, S. (2009). Issues in dynamic assessment. *English Language Teaching*, 2(4), 188-198.
- Black, C., & Anestis, M. (2008). *12 SAT practice tests and PSAT*. New York: McGraw-Hill.

- Black, C., & Anestis, M. (2011). *SAT*. New York: McGraw-Hill.
- Chapelle, C. H. (2012). Validity argument for language assessment: The framework is simple. *Language Testing*, 29(1), 19-27.
- Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Haney, M., & Evans, J. (1999). National survey of school psychologists regarding use of dynamic assessment and other nontraditional assessment techniques. *Psychology in the Schools*, 36(4), 295-304.
- Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy*, 23(1), 9-25.
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge: Cambridge University Press.
- Haywood, H., & Tzurriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education*, 77(2), 40-63.
- Kane, M. (2011). Validating score interpretations and uses. *Language Testing*, 29(3), 3-17.
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23, 112-127.
- Kozulin, A., & Garb, E. (2004). Dynamic assessment of literacy: English as a third Language. *European Journal of Psychology of Education*, 19(1), 65-77.
- Lantolf, J. P. (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.
- Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching Journal*, 42(3), 355-368.
- Lantolf, J. P., & Poehner, M. E. (2008). Dynamic Assessment. In E. Shohamy (Ed.), *The Encyclopedia of language and education* (vol. 7): *Language testing and assessment* (pp. 273-285). Cambridge: Cambridge University Press.
- Lidz, C. S., & Macrine, S. (2001). Identification of minority and immigrant students for gifted education: The contribution of dynamic assessment. *School Psychology International*, 22(1), 74-96.
- Pamela, J. S. (2004). *Barron's, how to prepare for the TOEFL*. New York: Barron's Educational Series, Inc.
- Pena, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, 10, 138-154.

- Phillips, D. (2003). *Preparation course for the TOEFL test, the paper test*. New York: Pearson Education.
- 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. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, 91(3), 323-340.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 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. (2014). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337-357.
- Shohamy, E. (2005). The power of tests over teachers: The power of teachers over tests. In D. J. Tedick (Ed.), *Second language teacher education* (pp. 101-112). Trenton, NJ: Lawrence Erlbaum associates.
- Tajeddin, Z., & Tayebipour, F. (2012). The effect of dynamic assessment on EFL Learners' acquisition of request and apology. *Journal of Teaching Language Skills*, 4(2), 87-118.
- Toe, A. (2012). Promoting EFL students' inferential reading skills through computerized dynamic assessment. *Language Learning & Technology*, 16(3), 10-20.
- Tzuriel, D., & Kaufman, R. (1999). Mediated learning and cognitive modifiability: dynamic assessment of young Ethiopian immigrant children to Israel. *Journal of Cross-Cultural Psychology*, 30, 359-380.
- Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning*, 18, 21-32.
- Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, 44, 84-97.
- Vygotsky, L. (1978). *Mind in society*. Cambridge: Cambridge University Press.