

Designing and validating a test battery of computerized dynamic assessment of grammar

Ghasem Modarresi¹

Assistant Professor, Department of English, Science and Research branch, Islamic Azad University, Tehran, Iran

Seyyed Mohammad Alavi

Associate Professor, University of Tehran

Received on April, 2014

Accepted on December 15, 2014

Abstract

Dynamic Assessment (DA) is the integration of assessment and instruction into a unified activity which derives from Vygotsky's (1978) theory of the ZPD. An important strand of research that will solidify a central place for DA in the L₂ domain is computerized DA (C-DA) that can be used to assess a large number of students simultaneously while observing the psychometric properties of testing. The present study aimed at designing and validating a test battery of C-DA of grammar for EFL learners named Computerized Dynamic Grammar Test (abbreviated as CDGT). The software reports three scores for a test taker: a non-mediated score, a mediated score and a learning potential score. A pool of 122 homogeneous BA and MA students from different universities participated in this study. The results obtained from the test takers' scores showed that C-DA is effective in helping students increase their performance and promote their learning development. Data analysis also indicated that C-DA is more effective for low achievers than for high achievers. A major

¹ Corresponding author: qasem.modarresi@gmail.com

implication of the study is that C-DA can be incorporated in informal and formal testing situations.

Keywords: computerized dynamic assessment, dynamic assessment, English grammar, learning potential score, non-dynamic assessment

1. Introduction

Dynamic Assessment (DA), as a theoretical framework for research undertaken in applied linguistics, is grounded in Vygotsky's (1978) writings on the Zone of Proximal Development (ZPD) which is the difference between what an individual can do independently and what they can do with assistance or mediation. Central to the ZPD, is the role of mediation, and DA, as an ontogenesis, emergentist and post-modernist trend in testing, integrates "two key elements of mediation and instruction into a unified activity to promote learner development" (Lantolf & Poehner, 2004, p. 50).

According to McNamara (1997), what is needed is a paradigm shift whereby instruction and assessment could be reintegrated as a single pedagogical activity. Likewise, Shohamy (2001) argues that teaching and assessment are oppositional activities since teaching and assessment are generally viewed as separate. In recent years, the great importance of DA in L2 (Albeeva, 2008; Aljaafreh & Lantolf, 1994; Anton, 2009; Poehner, 2008; Poehner & Lantolf, 2005) has been acknowledged. However, the major problem of DA lies in the issues of its reliability, validity, and application. In DA, the abilities are not regarded as stable traits that makes it hard for researchers and practitioners to ensure its reliability and validity (Poehner, 2005), although recent interpretivist approaches to test validity suggest that reliability may not be necessary after all and a test is valid if it promotes learner's development (Lynch, 2003).

Another problem of DA is the time consuming administration procedures (Haywood & Lidz, 2007; Poehner, 2005). De Beer (2006) notes the time problem regarding DA application in real educational settings, and Poehner (2008) adds that when teachers

may have classes of up to 100 learners, the feasibility of DA can be considered quite a challenge.

Van Lier (2004) points out that an area of research that is currently much in need of research is the role of technology in SLA, and that ZPD can shape the theoretical framework of research projects undertaken in Computer-assisted Language Learning. Significant studies in the general DA literature have reported the use of C-DA in second language acquisition (Birjandi & Ebadi, 2012; Jacobs, 2001; Pishghadam & Barabadi, 2012; Poehner & Lantolf, 2013; Tzuriel & Shamir, 2002). Each of these designs was programmed following a unique approach to provide standardized mediation during the procedure. What mainly prompted this study was the novelty of this particular area of research. Using C-DA, the present study attempted to cope with the above mentioned problems of reliability, validity, and feasibility. C-DA mostly follows an interventionist model with mediation offered from a menu of predetermined clues, hints, and leading questions selected in a lock-step fashion by the computer (Poehner, 2008).

It is also worth mentioning that C-DA does not follow a one-size-fit-all model for all students since it offers a range of hints from the most implicit to the most explicit for any single item. As computer facilities are becoming increasingly accessible, their role in mediating should become the focus of research (Hyland & Hyland, 2006). They believe that computers can empower students. The researchers of the present study set out to assess the grammatical knowledge of the learners based on the principles of DA, and in so doing, they sought an alternative way to test grammar, and they came up with communicative test of grammar which is compatible with the new perspectives and recent findings regarding grammar teaching and testing in second language classrooms.

It has already been shown that DA is effective in promoting learners' development (Guthke & Beckmann, 2000; Lantolf & Poehner, 2004). However, the major significance of the current study is that C-DA is closely related to teaching methodology. It may, indeed, promote prompt-based language teaching as opposed to spoon-feeding education. A successful teacher does not teach

from A to Z; rather they provide learners with prompts. In the model, the researchers designed, the invisible teacher is in the test, who supervises, monitors, and scores the test. In addition, more than just giving the hints is the quality of the hints provided by the teacher. Writing hints or prompts is a creative, and, therefore, challenging task. A good hint would lead the learner to the desired outcome. This property, as well, highlights and is in line with the process-oriented nature of grammar. As grammar is related to Logical Intelligence (Pishghadam & Moafian, 2008), writing prompts intelligently is teaching itself.

2. Background

2.1 Dynamic Assessment vs. Non-dynamic Assessment

The fundamental difference between DA and non-dynamic assessment (NDA) derives from Vygotsky's theorizing in the ZPD that is based on a fundamentally different understanding of the future. In NDA, actual development is sought rather than potential development. That is to say, NDA is based on the past-to-present model of assessment, while DA is based on the present-to-future model of assessment (Valsiner, 2001). Extensive research within DA has been carried out in the Netherlands, Germany, the United States, Canada, Belgium, Europe in general, the United Kingdom and South Africa (Murphy, 2011). The new trends within psychological assessment suggest DA methods as complementary to mainstream assessment (Stiggins, 2005).

Based on the Vygotsky's ZPD, Lantolf and Poehner (2004) made a clear distinction between the two general approaches to DA: *Interactionist* DA and *Interventionist* DA. *Interactionist* DA finds its origins in Vygotsky's qualitative, interpretation of the ZPD which encourages us not to measure but to focus on students' development, and this can only be accomplished through interaction and cooperation. Thus, mediation in the interactionist model emerges from the interaction between the teacher as the mediator and the learner, accordingly responding to the learner's ZPD. *Interventionist* DA is rooted in Vygotsky's quantitative interpretation of the ZPD as a difference score. It is currently

utilized in either of two formats: 1) A pretest-treatment-post-test experimental approach and 2) item-by-item assistance selected from a prefabricated menu of hints during the administration of a test (Poehner, 2008). A distinct advantage of interventionist DA is that it can be conducted with high numbers of individuals simultaneously via computer since it does not necessitate a face-to-face interaction.

2.2 ZPD and English Grammar

There is considerable evidence for the usefulness of the theoretical construct of DA in grammar instruction (Aljaafreh & Lantolf, 1994; Antón, 2003; Lantolf & Aljaafreh, 1995; Nassaji & Swain, 2000; Poehner, 2005). Aljaafreh and Lantolf (1994) examined the use of high frequency features of English (tense morphology, articles, model verbs, and prepositions) in the written performance of three ESL learners and reported that a shift from explicit mediation to a more implicit mediation contributed to students' development. Nassaji and Swain (2000), in a case study of two learners, provided feedback within the learners' ZPD to complement Aljaafreh and Lantolf's (1994) findings. Their study showed that help provided within the ZPD was more effective than help offered randomly.

According to Lantolf and Aljaafreh (1995), as learners displayed greater independence from the tutor's guidance and improved accuracy in their use of the relevant forms, development through the ZPD was observed over time. However, they argued that learner development was not a smooth linear process; instead, it followed the type of irregular trajectory covered by Vygotsky's description of development as a revolutionary process. Poehner (2005), in a phase of his dissertation, asked the participants to orally produce a past-tense narrative in French based on video clips from the film *Nine Months*. Based on the video clip, the learners must use the past tense including the *passé composé* and the *imparfait*. The findings demonstrate that DA is an effective means of understanding learners' abilities and helping them to overcome linguistic problems.

2.3 Computerized DA Software Programs

The initial attempt to employ computerized mediation was KIDTALK developed by Jacobs (1998) in which pre-school and school-age children were directed through a series of computer-based activities designed to evaluate their language aptitude. The program provided children with samples from an invented language based on Swahili that the researchers referred to as Kidtalk. Guthke and Beckmann (2000) developed computerized versions of the *Leipzig Lerntest* (LLT) that work similarly to KIDTALK. They designed a computerized LLT that is also adjustable to individuals' needs whereby training tasks are presented when examinees make errors. The program could be administered individually or in groups. Like non-computerized DA, the central issue in the Lerntest and KIDTALK procedures is the extent to which the assessment purposes and the available resources allow individualized mediation. Indeed, in some contexts, the Lerntest program is appropriate. In other settings, the human-computer collaborative format described by Tzuriel and Shamir (2002) will certainly be attractive because it further increases the possibility of working within individuals' ZPDs as explained below. They developed a computerized version of a DA procedure through which children were assessed based on their seriation thinking, an ability linked to performance in mathematical thinking through which the children arrange their thought in a series. One group of children received computer-based mediation, supplemented with human mediation when necessary, and another group was provided with computer-mediated assistance, endorsed by human mediation, and the other group given only human-mediated assistance. The first and second groups significantly outperformed the last one. However, the study did not include a group of children that only received the computerized mediation.

Birjandi and Ebadi (2012) examined the micro-genetic development of the oral abilities of foreign language learners by means of Google Wave (GW) and Skype for a period of three months. They concluded that the students' responsiveness is significantly associated with their level of ZPD regarding the time

they spent on each item. Pishghadam and Barabadi (2012) reported on the construction and validation of a C-DA software program known as Computerized Dynamic Reading Test to be used as an instrument in promoting learners' reading comprehension skills. The test presented two scores for each individual: a 'non-dynamic score' and a 'dynamic score' (to use their own terms) which was based on test takers' first try of each item and the average hints employed by them, respectively. The results highlighted the usefulness of C-DA in improving students' reading comprehension ability and in presenting information concerning their potentiality for learning.

Poehner and Lantolf (2013) reported on the use of DA principles in tests of L2 listening and reading comprehension offered through an online format. The results indicated both unassisted and assisted performance on the tests as well as the Learning Potential Score (LPS) which makes distinction between mediated and unmediated performance to predict how learners are likely respond to future instruction. Their study shed light on the significance of C-DA administered via the internet in learners' development. Modarresi and Alavi (2013) provided a comprehensive and critical overview of C-DA software programs designed to promote learners' development. They explored the data collection procedures and data analyses of the computer-based dynamic assessment of learners in first and second language contexts.

The present study focuses on the use of C-DA in grammar instruction since recent research in SLA recognizes the need for attention to grammar and has led to a re-examination of the importance of grammar (Nassaji & Fotos, 2011). As for the importance of grammar in second language learning, Batstone (1994) explains that grammar is a component of discourse, a requisite feature of reading and speaking, and is not easy to separate in any clear-cut way from vocabulary. Indeed, effective communication in a language would be seriously impaired without an ability to put grammar to use in a variety of situations. The researchers of the present study, therefore, aim to provide answers to the following four questions:

1. Does a test battery of C-DA observe the psychometric properties of standardized tests?
2. Does C-DA have any significant effect on EFL learners' grammatical knowledge?
3. Is C-DA able to make a distinction between a learner's potential and actual levels of performance?
4. Do high and low achievers significantly differ in their use of mediation in the form of hints?

3. Method

3.1 Participants

The sample consisted of 177 university students, including 126 BA and 51MA students at intermediate level majoring in TEFL, English Literature, and Translation Studies. The participants were selected based on accidental sampling from nine universities of Iran. The mean age of the sampling was 24 years; they were between 20 and 36 years old. In order to make our sampling fairly homogenous in terms of their level of proficiency, the researchers just included those students whose scores obtained from NDA on CDGT used in this study fell one standard deviation below and above the mean. In so doing, 55 participants were removed from the study since their non-mediated scores on this CDGT did not fall between the mean ($M=75$) and one standard deviation ($SD=40$) (the total number one could attain was 200). Since at some universities the English laboratory was not computerized, the researchers also created a Weblog named ghmodarresi.blogfa.com and put the software link on the site. Having completed the test, the participants sent the file created on their computers containing demographic information and their scores to the E-mail.

3.2 Instrument

The only instrument used in this study was Computerized Dynamic Grammar Test (CDGT). The test is comprised of 40 question items, each item including 5 hints. This program is capable of giving hints to students when students make mistakes and also providing the

teacher with two scores (mediated and non-mediated). The software through which the program was designed was Microsoft Visual Studio-Net (2010) and the programming language was #C. To make the background of the tests and the hints clearer, Photoshop CS12 was used.

3.3 Test Construction Procedure

The current study adopted a straightforward procedure including three steps including 1) test preparation, 2) software preparation, and 3) test piloting to ensure the reliability and validity of the dynamic test as much as possible.

3.3.1 Test Preparation

The most important step of the study was to construct and validate a Test Battery of C-DA of Grammar. To find an appropriate grammar test for this dynamic test, the researchers examined three versions of TOEFL materials as well as six grammar textbooks. One way to ensure reliability is to make use of tasks whose reliabilities are already well-established in the static or non-dynamic mode (Haywood & Tzuriel, 2002). In other words, the test battery will enjoy a higher reliability if the tasks used in DA have high reliability in the static mode; it is for this reason that the test sample used in this study was selected from standardized TOEFL test *“How to prepare for the TOEFL test: Test of English as a foreign language”* by Sharpe (2001). Although changes were made in the original formats of the tests, the non-dynamic test still enjoyed content validity. Since the scope of structural patterns cannot be captured in a single test battery, the researchers tried to cover the most frequently-used patterns. In so doing, they categorized the structural patterns into 10 major categories including 40 subcategories. As a matter of fact, the list of grammatical structures for the test preparation is too vast to be covered in the test, but the researchers believed that a valid test of grammar must entail the basic grammatical patterns. Thus, the list mentioned below was selected meticulously, covering the most frequent structural patterns of English grammar as it is also used in TOEFL tests. Altogether,

fortunately, the number of sub-categories chosen by the researchers was 40 that equals the number of grammar questions used in paper-based TOEFL examinations. The researchers also tried to construct each of the test items so as to cover the most frequent structural patterns of the sub-category under question as much as possible. For example, while writing the test for past tense, they attempted to include different forms of past tense (i.e. simple past, past continuous, and past perfect) in a test item. Indeed, the discourse-based tests of grammar help make it possible. In this way, the researchers took the criterion of content validity into due consideration.

Next, the researchers prepared 5 hints for each item. The hints matched the structural patterns covered in the test battery. Students were allowed to answer each item within 4 minutes. Regarding the hints, since the grammatical patterns are labeled by abstract names such as connectors, present continuous, or adverbs of indefinite frequency, the researchers used examples together with the labels to help students understand them. The C-DA program allowed the students to take the test in 2 hours and 40 minutes. If they failed to answer an item within 4 minutes, they would lose that item automatically. Having selected appropriate grammatical structures, the researchers, then, set out to prepare the items for the test. They needed to change the format of the test items. The original test was in multiple-choice format which was not appropriate to be used here since the researchers offered hints to the students to find the answer. If a multiple-choice format were used, as soon as a student was given a hint, he or she would know that the answer was wrong and they would be left with three alternatives, and by receiving the second hint, they would have only two alternatives, and so on. Thus, they could guess the correct answer from the remaining choices. The researchers decided to design the communicative test of grammar as an appropriate test format.

Therefore, the application of discourse-based grammar has two major merits. First, since new perspectives on grammar consider the discourse in emergentist and sociolinguistic terms, the current study is an early attempt to assess grammar at the discourse level in C-DA. Second, the learners could identify the errors while giving hints

to learners would not help them guess the right answer. The following test is an example of a discourse-based grammar test designed by the researchers to assess knowledge of relative clauses. After the items were prepared, five hints were prepared for each item. In DA, the quality of hints is very important as different learners may have different ZPDs for the same incorrect forms, meaning learners require different levels of assistance. The first hint was the most implicit and the last hint was the most explicit. The assistance was given on a progressive scale varying from implicit to specific answers. Indeed, each time a learner answered a question incorrectly, computerized mediation was provided with increasing explicitness.

Test item and its hints:

Instructions: You have 4 minutes to answer each question. If you can answer an item correctly in your first attempt, a score of 5 is awarded for that item. This is your non-mediated score. If you answer the item in your second attempt, a score of 4 is awarded and so on until the correct answer is revealed in the fifth hint and a score of 0 is earned for the item. This is your mediated score.

Several people were injured this morning when a lorry which was carrying pipes overturned in the center of town and hit two cars. Ambulances called to the scene took a long time to get through the rush hour traffic. People who saw the accident say the lorry hit the cars after it swerved to avoid a pile of stones leaving in the road.

Hint 1 → That's not the right answer, try again.

Hint 2 → Look at the relative clauses in the test. They are defining relative clauses used to include essential information.

Hint 3 → There are four relative clauses here. Sometimes we can leave the *relative pronoun + auxiliary verb* out of the clause. For example, *The man **who is watering the garden** is*

my uncle can be reduced to *The man watering the garden is my uncle*.

Hint 4 → Pay attention to the last sentence. We can form clauses with a present participle (e.g., *watering*) in active sentences and a past participle (e.g., *watered*) in passive sentences.

Hint 5 → The right answer is **left** NOT leaving.

3.3.2 Development of the Software

When the tests of grammar were constructed and the hints were provided for the learners, the next step was to design the software package. Before designing the software, the researchers asked three experts in teaching grammar, to read the test and check for different aspects of the items including face validity, content validity and the format of the individual items. The researchers revised the test according to their feedback. The software was designed in a way that could be run by all operating systems. On the opening page of the software, test takers needed to provide personal information (e.g., their names, age and majors). The next page of the software gives test takers a brief and simple description of the software and DA as well in English. After reading the description, test takers started the test. By starting the test, the items appeared on the screen. The computerized mediation ends automatically when test takers find the correct answer to the question. The maximum level of mediation each student receives is five. That is, if a student gives a wrong answer to an item, the software provides them with hints until they get to the right answer in the fifth hint. In other words, the test generates two weighted numerical scores. If an initial response to an item is correct, a score of 5 is awarded to that item. If a second attempt at the same item produces a correct response, a score of 4 is awarded and so on until the correct answer is revealed and a score of 0 is earned for the item. The total number of 5s are summed and reported as the ‘actual score’ for the test. This represents the

learner's actual (i.e. unmediated) performance. The total points earned for mediated responses are summed and reported as the 'mediated score'. Then, the Learning Potential Score (LSP) is calculated using Kozulin and Garb's (2002) formula that takes account of the difference between actual and mediated scores. When the test is over, a scoring file is created on the desktop. The following information about each test taker is stored in this file.

1. Test takers' non-mediated scores: This score is calculated according to the students' scores obtained from their non-dynamic performance or their first try. In fact, this score is the same as that obtained in traditional tests. To make it comparable with the score based on DA of the test, the researchers calculate this score on a scale of 0 to 200 points; five points for each item.
2. Test takers' mediated scores: This score is calculated according to the students' scores obtained from their dynamic performance or their use of the hints. The number of hints used by each test taker is subtracted from the total number of hints which is 200. The number that is obtained by this subtraction is the score based on DA. For instance, imagine that a student uses two hints for the first 20 items of the test; that is, two hints for each of these items. This student's score is 160 which is calculated by subtracting the number of hints used by him (here 40 hints) from 200. The non-mediated score of the same student would be 100 because this student has given wrong answers to the first 20 items of the test, and only after receiving hints they have managed to provide the right answers.
3. The LPS: This score is calculated using Kozulin and Garb's (2002) formula which takes account of the difference between actual and mediated scores.
4. The number of hints used in each item: The software subtracts the number of hints used by each test taker from the total number of hints. It means that for each hint that is used, one point is deducted from the total score, which is 200.

3.3.3 Test Piloting

The researchers conducted a pilot study and collected the relevant information regarding the usefulness of the test. Through piloting, the test was revised so that some items were changed and some removed. The researchers administered the test to a pilot group of 80 students who had roughly the same language proficiency level as the participants of the study to standardize the test. The sample participated here was not selected for the study. Particularly, during this pre-testing, the researchers noted the feedback either cognitively or emotionally from the test takers who expressed their opinions orally after they took the test e.g., their feelings and reactions to C-DA. After conducting the pilot study, the researchers made some modifications regarding the test content, difficulty level of the items, length of the items, the usefulness and the quality of hints, and the feasibility, and the quality of the software package. Here are some examples:

1. In the first draft of the software, three minutes was allotted for each item, but regarding the test takers' feedback during the pilot study, the time was increased to four minutes. Nearly all of the participants believed that they could not study the hints carefully, and they would choose the next hints as soon as possible because they thought that they are pressed for time. Therefore, the researchers again selected 10 participants and recorded the time for them to answer a given item, and they concluded that four minutes is a more reasonable duration for each item.
2. In the first draft, item 40 beginning with the sentence "it is like a chair with a curse" was vague for the students so that a sentence was added to the test to make it clear.
3. Item 10 and item 18 were lexically difficult for the students, so these items were simplified.
4. Regarding the quality of the hints, in item 8, the second hint focused on the use of *that*-clauses after verbs, adjectives and nouns, but the labels were vague, therefore, the researchers added examples to make the concepts clear. In item 1, the forth hint had typo which was corrected.
5. As for the length of the items, in the first draft the length of the items had not been paid due attention, and some items were

longer than the others. For instance, item 5 and item 21 were too long and the students said that the time was not enough for those items. The researchers revised the long items while preserving the coherent content of the items. All of the items were checked to have a length of three to four lines while displayed on the screen.

6. As for the feasibility of the test, the researchers modified the description of C-DA would appeared on the second page of the software to make the students familiar with the C-DA and the test itself. In fact, in the first draft students were not aware of the scores they would gain from the test. Therefore, the three scores, including non-mediated, mediated, and LPS, were explained for them in the final version.
7. And finally, the design of the software was changed several times regarding the font, size, and color of the texts and the hints. Moreover, there were major improvements regarding the way the last hint was displayed for each item, the timing of the test, and how the time limit was shown for each item. In the modified version, when the last hint is displayed, the test taker has time to reflect on it. It means that the test does not go automatically to the next item even if the time (four minutes) is finished, unless the test taker clicks on OK. Moreover, in the draft version, by each click, a hint was displayed even if the arrow was on the blank space and not on the text, and this fault was noted by participants and corrected in the final version.

3.4 Data Analysis

The internal consistency of both dynamic and non-dynamic tests was assessed with K-R 21 method of estimating reliability. To determine the statistical significance of the difference between the means of mediated scores and non-mediated scores, paired samples t-test was run to see if DA results in significant improvement of test takers' performance. The Pearson product- moment correlation coefficient was run to estimate the concurrent validity of the test, and also to examine the significance of difference between mediated scores and non-mediated scores.

To estimate the Learning Potential Score, Kozulin and Garb's (2002) formula was used. The formula is used to differentiate between high learning potential and low learning potential students. This score indicates the potentiality of DA procedures for measuring students' potential. The formula is as follows:

$$LPS = \frac{(mediatedS - actualS)}{MaxS} + \frac{mediatedS}{MaxS}$$

Where:

mediatedS = mediated score

actualS = non-mediated score

MaxS = the highest obtainable score

4. Results

4.1 Reliability and Validity of CDGT

To estimate the reliability and validity of the test, KR-21 method of estimating reliability and Pearson product-moment correlation coefficient was used. The estimated reliability of non-dynamic and dynamic tests and concurrent validity of the dynamic test is displayed in Table 1.

Table 1: Reliability and validity of the test

	<i>Non-mediated</i>	<i>Mediated</i>
Reliability	.92	.71
Validity		.83

The test enjoyed acceptable reliability and validity. As for the construct validity of the test, recent interpretivist approaches to test validity suggest that a test is valid if it promotes learner's development (Lynch, 2003). Lantolf and Poehner (2008) explain that DA practitioners must also address another construct, namely, development, and validity is defined as the extent to which a test promotes learner development and development is understood in the

interaction between the mediator and the learners. Therefore, the test takers' significant increase of mean scores from non-mediated scores ($M=70.89$, $SD=23.20$) to mediated scores ($M=117.02$, $SD=27.37$) showed that the test has construct validity.

4.2 Test Takers' Performance on CDGT and Its Non-mediated Test

The second question of the study examined the significant difference between the students' non-mediated scores and their mediated scores. There was a statistically significant increase in CDGT scores from the non-mediated test ($M=70.89$, $SD=23.20$) to the mediated test [$M=117.02$, $SD=27.37$, $t_{(152)}=33.40$, $p<.05$] as displayed in Table 2.

Table 2: Paired samples t-test

	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>Sig (2-tailed)</i>	<i>P</i>
Non-mediated	70.89	23.20	152	33.40	0.000	0.05
Mediated	117.02	27.37				

While on the non-mediated test no students could achieve a high score (e.g., 134 or higher), on the mediated test, 32% of the students managed to get a score of 134 or above. There were also students whose scores on the mediated test were more than 115 while on the non-mediated test the highest score was 115.

The researchers also calculated the effect size or the relative magnitude of the difference between the means of non-mediated scores and mediated scores to see how big the difference between the means was. The eta squared statistic calculated manually by means of eta squared formula was 0.90 which is a large effect, following the guidelines proposed by Cohen (1988) as follows: .01=small effect, .06=moderate effect, .14=large effect. Therefore, the eta squared value showed a substantial difference in the students' scores before and after mediation.

4.3 Learners' Potential and Actual Levels of Performance

Research question 3 concerned the effectiveness of mediation in learning development and the extent to which learners benefit from explicit assistance. In other words, the third question concerns the potential ability of hint-based mediation to promote success for all and to see how students differ in their abilities to learn from assistance. The results from Pearson product-moment correlation coefficient, data analyses of students with the same non-mediated scores, and the overall percentage of each hint used provided a positive answer to the question. Indeed, all of the learners benefitted from mediation but they varied in their use of the hints. The obvious fact is that there are learners who need more explicit assistance or hints to respond to the test items. To provide a better picture of the individual differences in their level of potential development, Table 3 shows how two different students with the same non-mediated scores benefitted from mediation.

Table 3: Students’ non-mediated, mediated and learning potential scores

Students	Non-Mediated S.	Mediated S.	LP S	Non-mediated S.	Mediated S.	LPS	Non-mediated S.	Mediated S.	LPS potential score
A	35	45	.30	65	77	.44	95	127	.79
B	35	103	.95	65	114	1.11	95	133	.87
A	45	71	.48	75	102	.64	105	115	.62
B	45	106	.83	75	147	1.09	105	159	1.06
A	55	88	.60	85	123	.80	115	155	.97
B	55	133	1.05	85	160	1.17	115	163	1.05

The table presents the highest and lowest mediated scores for the non-mediated counterparts. Excluding the scores not one SD below and above the mean, the lowest non-mediated score was 35 and the highest one was 115 and included in the table are the highest and lowest mediated scores for the non-mediated scores of 35, 45, 55, 65, 75, 85, 95, 105 and 115, as representatives.

An interesting feature of the table is that the two identical non-mediated scores (e.g., 35) do not necessarily map onto the same mediated scores. For example, two learners produced the same non-mediated score of 35; however, the first learner produced the mediated score of 45 while the second produced a much higher

mediated score of 103, an indication that the learner responded more favorably to mediation. Besides, while the correlation between non-mediated scores and mediated scores was high for all of the test takers (0.83), the correlation between non-mediated and mediated scores for the lowest and highest mediated scores and their non-mediated counterparts including 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115 was moderate (0.73) (Table 4). Therefore, it can be concluded that while all the test takers benefit from the mediation, they perform differently on the mediated test.

Table 4: Correlations between non-mediation scores and mediation scores

		Non-mediated score	Mediated score
Non-mediated scores	Pearson Correlation	1	.732**
	Sig. (2-tailed)		.000
	N	122	122
Mediated scores	Pearson Correlation	.732**	1
	Sig. (2-tailed)	.000	
	N	122	122

To see the extent to which students made use of explicit mediation in the form of hints, the overall percentage of each hint used by them was calculated. The number of hints used by the test takers yielded interesting results. Figure 1 displays the percentage of each hint used by the test takers. As already mentioned, the hints are offered from the most implicit (hint 1) to the most explicit. Students used the hints from the most to the least as follows: 1) hint 2 (25.8%), 2) hint 4 (19.6%), 3) hint 3 (18.7%), 4) hint 5 (18.5%), and 5) hint 1 (17.3%). The second hint provides the test takers with the area that each question item is dealing with without any explicit guidance as to where they look for the answer in the test. Therefore, when the test takers are given the opportunity to focus on the question, having in mind the topic, they could perform well on the test.

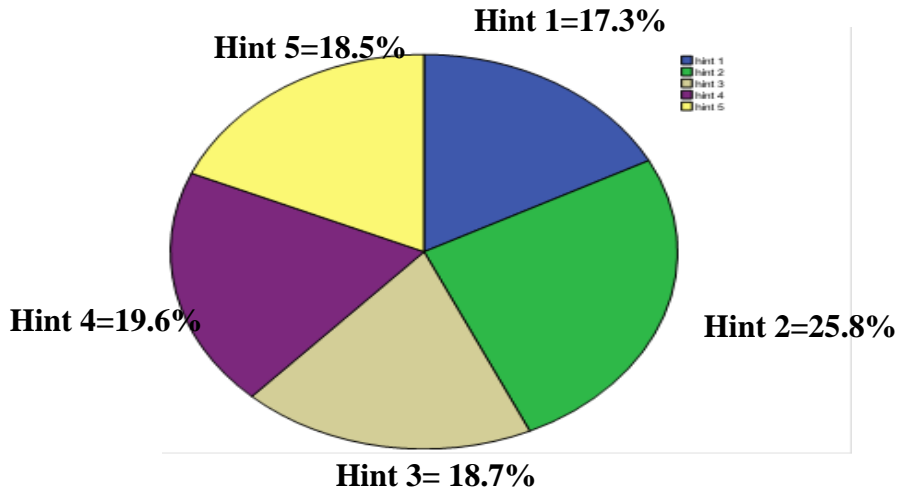


Figure 1: Overall percentage of each hint used

It should be noted that the students benefitted from other hints except hint 5, which mentions the right answer, roughly the same that shows students have differing abilities, and they vary in the use of hints, some of which need more explicit knowledge to reactivate their grammatical knowledge. As for the students' lack of ability to find the answer using all hints, it should be noted that, according to Ellis (2008), to understand ZPD, it is helpful to distinguish three levels of development. Vygotsky (1978) distinguished the actual developmental level and a level of potential development. The third level not commonly mentioned by sociocultural theories is the level that lies beyond the learner, that is, the learner is unable to perform the task even if assistance is provided (see Modarresi & Jalilzadeh, 2011).

4.4 High and Low Achievers and Their Uses of Mediation

To find out whether high and low achievers differ in their use of mediation in the form of hints, the learning potential scores (LPSs) of all test takers were measured. This score, indeed, measures the size of ZPD, and as proposed by Kozulin and Garb (2002), it can be used to make a distinction between high learning potential and low learning potential students. For example, by looking at the last row

of Table 3, we can see the LPSs of the test takers with the same non-mediated scores. Students who progressed well from non-mediated score to mediated score generated high LPSs while those who had slow progress generated low LPSs. For example, in the case of the two students whose non-mediated scores were 55, one of them who progressed considerably (from 55 to 133) generated a high LPS of 1.05 while the other who had slow progress (from 55 to 88) generated a low LPS of .60.

The LPSs of the students in this study ranged from .30 to 1.17. The learners, in this study, were categorized into three subgroups based on their LPS scores: High scorers ($LPS \geq .94$); mid-range scorers ($.93 \geq LPS \geq .77$) and low scorers ($LPS \leq .76$). The percentages of the learners in these three subgroups were 32%, 31%, and 36%, respectively. Just as there were some students with the same non-mediated scores but different potential scores, there were students with different non-mediated scores but with approximately or exactly the same learning potential scores. For example, as displayed in Table 4.4, if we compare the learner who had a low non-mediated score of 55 with the learner who had a high non-mediated score of 115, we see that despite their different non-mediated scores, they generated the same LPS of 1.05.

To answer the last question, the participants were divided into two subgroups of low achievers (from 35 to 70) and high achievers (from 75 to 115), based on their non-mediated scores. There were 61 participants in each group. While the high achiever subgroup could increase its mean score on the mediated test by 44 points, the low achiever subgroup could increase it to 48. Therefore, the latter subgroup achieved a rather bigger increase than the former. An important point here is that there are some low performers who could generate a higher LPS than high performers (e.g., 65 vs. 95). This learner with a higher LPS of 1.11 reacted more favorably to instruction than the high performer with a lower LPS of .87. It can be concluded that DA is especially useful to learners who gain more scores on static tests.

5. Discussion and Conclusion

In this research, the major focus was on designing and implementing a test battery of C-DA, named CDGT, as an assessment instrument which generates different types of scores with regard to English grammar. The distinct advantage of the test is its applicability and feasibility. A large number of students participated in this model of dynamic assessment simultaneously. The results obtained from the test were reported quantitatively and interpreted clearly. The study also aimed to take advantage of the technology to alleviate the time-constraints that many teachers face in their classrooms. Moreover, in line with Crook's (1991) remarks, computers can serve as a human partner, or as an invisible teacher within the ZPD, and technology makes the computerized tool relevant to the focused intervention activities. Indeed, the computerized format of DA could reply to the major criticisms against DA with regard to reliability, content validity, concurrent validity, construct validity and applicability. There are just a few studies that set out to observe the reliability and validity of their instrument (e.g., Guthke & Beckman, 2000; Jacobs, 2001; Poehner & Lantolf, 2013).

The statistical analyses performed in this study provided useful evidence in favor of the reliability and validity of the CDGT. Moreover, in this study, low performers benefitted more from mediation than high performers, and this shows that DA is probably more helpful for low achievers without taking into account the source of their performance.

The study examined the effectiveness of the provision of structured mediation for EFL learners using computers. The results obtained from the study are in line with the previous research corroborating the observation that mediation in the forms of hints and leading questions improves learners' language skills and learning potential (Kozulin & Garb, 2002; Poehner, 2007). In the present study, the performance of the students increased significantly from the non-mediated test to the mediated test on a test battery of grammar.

Moreover, the results of the study reported that some students with the same non-mediated scores performed differently on the mediated test. Indeed, some of them benefitted more from the

mediation and they generated higher LPSs than their counterparts. The results of the study are similar to those of Pishghadam and Barabadi (2012) and Poehner and Lantolf (2013) who carried out their C-DA research on reading comprehension skills and concluded that some students with the same non-mediated scores perform quite differently on the mediated test. In the present study, the students made use of both implicit and explicit hints which indicates that explicit assistance is conducive to their knowledge of grammar.

The test calculates the LPSs of the individuals by taking the difference between non-mediated and mediated performance into account in order to predict learners' responsiveness to future instruction. Of course, the issue of high or low LPS does not mean that only those students who have higher LPS are able to continue language study. Indeed, learning potential is not a capacity with fixed amounts, but it can be increased through mediation and this is in agreement with Vygotsky's suggestion that mediation should be offered in a way that promotes success for all. Therefore, students' LPSs can be used as their level of potential development for placement decisions.

During the last two decades, there has been a shift in language paradigm from a reductionist, structuralism perspective to an anti-reductionist, and communicative perspective and focus on form has been incorporated with the emphasis on meaning. When students learn grammar at the discourse level, they can understand the grammatical structures and use their knowledge in real context. In this study, CDGT included discourse-based test of grammar that could simultaneously assess and instruct the students and is in line with the view of grammar as a dynamic system not a static system which incorporates form, meaning and use (Larsen Freeman, 1997). We can conclude that the software model of DA, acting as an other-regulator or tutor, could reduce the emphasis on learning strategies involving memorization and repetition and highlight learning strategies which entail consciousness raising, focus on form, self-awareness, problem-solving, and self-discovery.

C-DA as a new discovery in the field of testing and assessment for foreign language learners proved to be effective in enhancing learners' development while observing the psychometric properties

of testing. Its feasibility with a large number of students facilitates its use both in language classrooms and in high-stake testing situations.

As for the language learners, a test designed based on C-DA allows for self-assessment and reassessment; they can assess their progress and measure their ZPDs and reevaluate themselves as many times as needed, and based on their individual needs, they are provided with an optimal amount of assistance through which the mediation is presented in a gradual progression from an implicit to an explicit design. C-DA not only is a valuable source for the learner to assess their language knowledge but also helps them develop effective learning strategies like directed attention, self-evaluation, and self-discovery. Moreover, C-DA is not just limited to the classroom; it can be used at home or outside the classroom. The software package of C-DA is user-friendly and does not require computer expertise since it can be posted on the internet, or stored on portable flash drives.

As for the major implications for language teachers, C-DA places the classroom teacher as a guide or facilitator. C-DA provides the teacher with rich feedback regarding the quality of their hints to see the number of hints used to respond to an item and help them decide to keep, modify, or remove a hint so as to accommodate the learners' needs. Teachers are recommended to practice writing creative hints to teach learners using hint-based education or "development-oriented pedagogy" (Poehner & Lantolf, 2013, p. 15) as opposed to spoon-feeding education or teaching from A to Z. Furthermore, teachers can use C-DA along with human mediation in the classroom, offering more fine-tuned hints to the learners when they would likely need extra help from the teacher on a one-on-one basis. The combination of human mediation and computer mediation would produce more encouraging results for the learners.

C-DA also carries profound implications for test developers. Since C-DA can be administered to large numbers of students and reports of the students are automatically generated, test developers are suggested to integrate it in formal testing situations and high stakes tests like University Entrance Examination, IELTS, and

TOEFL. They can use C-DA along traditional standardized tests for comprehensive exams in Middle schools and High schools to make a compromise between criterion-referenced assessment and development-referenced assessment. They can use C-DA for diagnostic and placement decisions in English institutions where teaching can be attuned to students' ZPDs. They are recommended to maintain a professional distance from the negative washback and impact of "teaching to the test" practice.

To develop a more comprehensive picture of DA, especially C-DA, the researchers hope that this instrument can be used as an instrument for doing further research. In this study, the effect of C-DA is examined on the knowledge of grammar. Research is needed to be done in relation to other skills like listening, and writing. The design of the present study was within-group. Further research is recommended to follow a between-group design, which would compare a control group taking the original non-dynamic test and an experimental group taking the dynamic test. The test format used in this study is discourse-based in which test takers find the mistake but other test formats like cloze test or more open-ended types can be used in further studies on C-DA. The effect of gender and major of the participants were not controlled in this study. The homogeneity of the sample has not been met fully. It was checked based on the Z scores of all participants, keeping only those test takers whose non-dynamic scores fell within one standard deviation above and below the mean while it can be made homogeneous in another study by using a standardized placement test. Finally, further research is needed to show the internalization of mediation through inclusion of transfer items in another test battery of C-DA of grammar to prove if learners are able to transfer their mediated performance to non-mediated future performance.

References

- Albeeva, 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* (pp. 57-86). London: London, Equinox.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465-483.
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576-598.
- Batstone, R. (1994). *Grammar*. Oxford: Oxford University Press.
- Birjandi, P., & Ebadi, S. (2012). Microgenesis in dynamic assessment of L2 learners' sociocognitive development via web 2.0. *Procedia - Social and Behavioral Sciences* 32, 34-39.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed). New York: Lawrence Erlbaum Associates.
- Crook, C. (1991). Computers in the zone of proximal development: Implications for evaluation. *Computers and Education* 17(1), 81-91.
- De Beer, M. (2006). Dynamic testing: Practical solutions to some concerns. *SA Journal of Industrial Psychology*, 32 (4), 8-14.
- Ellis, R. (2008). Principles of Instructed second language acquisition. Retrieved on May 12th, 2015 from <http://www.cal.org/resources/digest/instructed2ndlang.html>
- Guthke, J., & Beckmann, J. F. (2000). The learning test concept: Application in practice. In C. S. Lidz & J. Elliott (Eds), *Dynamic assessment: Prevailing models and applications* (pp.17-69). Oxford, UK: Elsevier.
- 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.
- Hyland, K., & Hyland, F. (2006). Context and issues in feedback on L2 writing. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 1-19). Cambridge: CUP.
- Jacobs, E. L. (1998). Kidtalk: A computerized language screening test. *Journal of Computing in Childhood Education*, 9 (2), 113-131.
- Jacobs, E. L. (2001). The effects of adding dynamic assessment components to a computerized preschool language screening test. *Communication Disorders Quarterly*, 22 (4), 217-226.
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23(1), 112-127.

- Lantolf, J. P., & Aljaafreh, A. (1995). Second language learning in the zone of proximal development: A revolutionary experience. *International Journal of Educational Research*, 23(7), 619-632.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49-74.
- Larsen-Freeman, D. (1997). *Grammar dimensions: Form, meaning, and use*. Boston: Heinle & Heinle.
- Lidz, C. S., & J. G. Elliott. (2000). *Dynamic assessment: Prevailing models and applications*. Amsterdam: Elsevier.
- Lynch, B. K. (2003). *Language assessment and program evaluation*. Edinburgh: Edinburgh University Press.
- McNamara, T. (1997). Interaction in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446-466.
- Modarresi, Gh. & Alavi, S. M. (2013). A critical overview of computerized dynamic assessment software programs. *Elixir International Journal: Lang. & Testing* 65, 18-24.
- Modarresi, Gh., & Jalilzadeh, K. (2011). The role of mediation and ZPD in first language acquisition. *European Journal of Scientific Research*, 57 (1), 146-151.
- Murphy, P. (2011). *Dynamic assessment, intelligence and measurement*. UK: Wiley-Blackwell.
- Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classroom: Investigating form-focused instruction in communicative contexts*. New York: Routledge.
- Nassaji, H., & Swain, M. (2000). A Vygotskyan perspective towards corrective feedback in L2: The effect of random vs. negotiated help on the acquisition of English articles. *Language Awareness*, 9(1), 34-51.
- Pishghadam, R. & Moafian, F. (2008). The role of Iranian EFL teachers' multiple intelligences in their success in language teaching at high schools. *Pazhuhesh-e-Zabanha-ye-Kha-reji*, 42, 5-22.
- Pishghadam, R., & Barabadi, E. (2012). Constructing and validating computerized dynamic assessment of L2 reading comprehension. *IJAL*, 15 (1), 73-95.
- Poehner, M. E. (2005). Dynamic assessment of oral proficiency among advanced L2 learners of French, Ph.D. dissertation. The Pennsylvania State University, University Park, PA.

Designing and validating a test battery of computerized dynamic assessment

- 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*. New York: 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). Retrieved June 17th, 2013 from: <http://ltr.sagepub.com/content/early/2013/04/22/1362168813482935>.
- Sharpe, P. J. (2001). *How to prepare for the TOEFL test: Test of English as a foreign language*. New York: Barron's Educational Series, Inc.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Longman.
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
- Tzurriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning*, 18(1), 21-32.
- Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, 44(2/3), 84-97.
- Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Boston, MA: Kluwer.
- Vygotsky, L. (1978). *Mind in society*. US: President and Fellows of Harvard College.
- Vygotsky, L. S. (1986). *Thought and language*, Newly revised and edited by A. Kozulin. Cambridge, MA: M