Iranian Journal of Applied Linguistics (IJAL), Vol. 15, No. 1, March 2012, 73-95

Constructing and Validating Computerized Dynamic Assessment of L2 Reading Comprehension

Reza Pishghadam [a]
Associate professor in TEFL, Ferdowsi University of Mashhad, Iran

Elyas Barabadi
MA in TEFL, Ferdowsi University of Mashhad, Iran

Received 4 August 2011; revised 20 November 2011; accepted 2 February 2012

Abstract

The main purpose of this study was to construct and validate a Computerized version of Dynamic Assessment (C-DA) and examine its effectiveness in enhancing reading comprehension. Feasibility and concern for psychometric properties of testing are issues that have limited the use of DA approaches. In this study, C-DA is offered as a solution for overcoming such limitations. To this end, a software package named Computerized Dynamic Reading Test (CDRT) was developed. The software is capable of providing test takers with strategy-based hints. For each test taker, two scores are assigned by the software; a non-dynamic score which is based on test takers' first try of each item and a dynamic score which is based on the average hints they have employed. One hundred and four university students took the test. The findings of the study indicated that while observing the psychometric standards of testing namely, reliability and validity, C-DA was useful both in improving students' reading comprehension ability and in obtaining information about their potentiality for learning which goes beyond and over the initial performance level. While some test takers made the best use of the

[a] *E-mail address*: pishghadam@um.ac.ir,
Corresponding address: English Language Department, Ferdowsi University of Mashhad, Park square, PO box 9177948974

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

hints and could enhance their comprehension of the text, others could not use them to their advantage. The Information obtained from DA enables teachers to provide students with more individualized and consequently more effective instruction.

Keywords: Dynamic assessment; Non-dynamic assessment; Computerized dynamic assessment; Strategy and reading comprehension

Introduction

Following the innovations in the field of psychometrics, a new movement has emerged out of Vygotsky`s (1978) ideas which has opened up new horizons for testers and teachers. This new movement questioning the major tenets of traditional tests as totally static and deterministic, has introduced Dynamic Assessment (DA) as an emergenistic and postmodern notion in testing. Unlike traditional tests in which there is no dialog between teacher and learner during examination, in DA the assessor enters into a dialog with the learner to find out the current level of performance on the task and to share with him possible ways in which that performance might be enhanced on a subsequent occasion. In fact, in DA, assessment and learning are seen inextricably mingled and not as separate processes (Haywood & Lidz, 2007; Haywood & Tzuriel, 2002; Lantolf, 2009; Lidz & Gindis, 2003).

In recent years, the growing importance of DA in L2 (Albeeva, 2008; Aljaafreh & Lantolf, 1994; Poehner, 2008; Poehner & Lantolf, 2005) in general, and L2 reading comprehension (Antón, 2009; Kozulin & Garb, 2002, 2004) in particular has been recognized. With the integration of assessment and instruction, DA tries to achieve two main purposes: to improve students' development of certain abilities e.g., reading comprehension and to understand about their potentiality for learning by focusing on the cognitive processes of learning. Tzuriel (2003) states that DA better shows children's potentiality for learning than static one, and also it is more useful in enhancing the learning how to learn. He believes that learning processes can be understood more easily and effectively through DA. This new strand of assessment is also more in line with new perspectives toward reading instruction and assessment. Afflerbach (2007) for instance, believes that readers' development should be the main consequence of reading assessment. In fact, readers' development is the cognitive component of reading assessment which focuses on skills and strategies used by developing readers; such process-oriented reading

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

assessment allows teachers to assess in the midst of learning.

Although DA offers several advantages over traditional tests, it is not used widely in educational context due to some problems associated with its application. The administration of DA seems to be time consuming and ensuring its reliability and validity has been a major concern among DA practitioners (Haywood & Lidz, 2007; Haywood & Tzuriel, 2002; Poehner, 2008). According to Haney and Evans (1999) other problems are related to lack of adequate knowledge base and expertise in this field. In sum, feasibility and concern for psychometric properties of testing are issues that have limited the use of DA approaches. The present study offers Computerized Dynamic Assessment (C-DA) as a workable solution to the above mentioned problems. In what follows, first DA and Vygotsky's Zone of Proximal Development (ZPD) which is the theory behind it are explained in more detail, and then the uses of DA in L2 context are discussed briefly.

Background

Dynamic vs. Non-dynamic Assessment

DA has been used and defined by many researchers in slightly different ways but the common thread running through all these definitions is that assessment and instruction are viewed as two facets of the same activity. Here are two definitions given by some experts in the field of DA:

Lidz and Gindis (2003, p. 99) define DA as "an approach to understanding individual differences and their implications for instruction that embeds intervention within assessment procedure". Hasson and Joffe (2007, p. 10). Likewise, contend that "DA includes a range of methods and materials to assess individuals' potentiality for learning". Its aim is to reveal the maximum level of performance by providing mediation in the course of assessment session. Mediation refers to the kind of help given to the learners by the teacher within their ZPD, and consequently enhances their development. This help can be in the form of leading questions, hints or prompts. The important point here is that the teachers as mediators should be sensitive to learners' ZPD so that the mediation could improve their development not simply help them complete the task (Wertsch, 1984).

Poehner and Lantolf (2005) point out that although DA has its origin in the

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

work of Vygotsky (1978), it was his colleague Luria (1961) who contrasted statistical with dynamic approaches to assessment. In the former, individuals' solo performance on a test is indicator of his capabilities; the latter requires their responsiveness to mediation as well during assessment. In Grigorenko and Sternberg's words (2002, vii), "if the goal of assessment is to modify learners' performance during assessment, it is dynamic otherwise static". Indeed, working in the ZPD which is the distinguishing feature of DA and non-dynamic assessment (NDA) is not simply to help learners get through the task but to help them gain a fundamental understanding of the task so that they can move from the given activity to other new ones; that is, transcendence (Lantolf, 2009; Poehner & Lantolf, 2005).

The mounting concern of DA is to identify the conditions that limit students' access to their intelligence and to remove them and then assess abilities again. In other words, it tries to assess what would be possible under more optimal condition that includes intervention. In fact, DA attempts to identify non-intellective variables such as unfamiliarity with materials and content of the test and to redress them (Haywood & Lidz, 2003). DA has the potential to rule out situational bias by providing children with the "rules of the game" (Jacobs, 2001; Kester, Peña & Gillam, 2001; Peña, Iglesias & Lidz, 2001).

"The use of DA reflects the need for a kind of assessment that informs instruction rather than result in diagnostic labels and classification for program eligibility" (Haywood & Lidz, p. 74). The authors further contend that DA gives us some specific information concerning students' learning potential that is not obtainable from traditional approaches to assessment. However, traditional standardized tests provide us with other valuable information about learners. Indeed, psycho-educational assessment process is dependent upon data gathered from various sources, of which DA is one. In fact, DA can be used along with traditional tests, to fill the gap left by these tests or to correct some of their errors especially when teachers or psychologists exclusively rely on these static tests.

Poehner and Lantolf (2005) explained that there are two general approaches to DA, interventionist and interactionist; the former is rooted in quantitative interpretation of ZPD while the latter is rooted in more qualitative interpretation of ZPD. Interventionist DA remains roughly similar to traditional standardized tests by providing standardized predetermined forms of assistance (graduated hints).

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

Moreover, quantification of the results and concern over psychometric properties of the test are highlighted; in contrast, in interactinist DA, assistance is the result of ongoing interaction between assessor and student regardless of the effort required for each individual test taker (Poehner, 2007). Sternberg and Grigorenko (2002) classify DA approaches as following either a *sandwich* or *cake* format. The sandwich format is consisted of three phases: pretest, mediation and posttest. The pretest which is administered to establish a baseline measure is followed by a mediation phase and finally there is a posttest to check out the effect of mediation on students' performance. Contrary to sandwich format, in cake format, mediation is given during assessment session in response to students' emerging needs.

Since Vygotsky's notion of ZPD is the theoretical underpinning of DA, (Kozulin & Gindis, 2007) we will explain it in detail before moving on to methodological aspects of this study.

Zone of Proximal Development; The Theoretical Foundation of DA

Poehner and Lantolf (2005) point out that the concept of ZPD implies that potential development differs from actual development, meaning that the latter cannot be a predictor of the former. Moreover, potential development is by no means a priori prediction; instead it emerges from mediated activity between mediator and the learner. According to Poehner (2007, p. 326), "ZPD is Vygotsky's approach to understanding and supporting cognitive development". Kozulin and Gindis (2007) observe that applied research is one of the three contexts in which the notion of ZPD has been used by Vygotsky (1978); "the applied aspect focuses on the difference between the child's individual and aided performance" p. 353). Similarly, Langford (2005) in his discussion of DA in relation to ZPD states that assessment of the child performance while assisted by others is a better predictor of what the child can do in future than assessment without such assistance. The latter according to Poehner (2007) only indicates those abilities which have been internalized while the former indicates those abilities that are still forming-the next or proximal level of development. Kozulin (2003) mentions three ways in which ZPD is related to DA: (1) the emerging and maturing psychological functions of the child are highlighted; (2) it recognizes the value of assisted performance as a legitimate parameter of assessment procedure; (3) ZPD helps to differentiate between actual performance and learning potential of the child. As Lidz and Gindis (2003) observe: “Vygotsky’s major objection to standardized tests was that they confused latent capacities with developed ones" (p. 102). Lantolf (2009) suggests

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

that "a successful education must be sensitive to learners' zone of proximal development and this requires the dialectical integration of instruction and assessment into a seamless and dynamic activity" (p. 355). To illuminate this dialectical principle of instruction and assessment, Lantolf tactfully suggests an analogy between Marx's (1844/ 1972) view concerning the dialectical unity of production and consumption and Vygotsky's (1997) view of instruction and assessment. For Marx, both production and consumption were components of the same process; that is, without production, no consumption and vice versa and also for Vygotsky, effective instruction is not possible without assessment and any effective assessment entails instruction.

In conclusion, it is the notion of ZPD that gives us a present-to-future understanding of individuals' future. Rooted in this notion, DA tries to target the emergent future of individuals by dialectical integration of instruction and assessment (Valsiner, 2001, cited in Poehner & Lantolf, 2005).

DA in L2 Context

Poehner (2007) and Poehner and Lantolf (2005) demonstrated the relevance of DA for L2 classrooms. The specific examples that they present of the use of DA for speaking ability confirm the value of DA for a better understanding of learners' development. Following an interactionist approach to DA, Albeeva (2008) examined the effects of DA on L2 listening comprehension. The result of her study showed that DA in comparison with its counterpart, NDA, is better able to help teachers identify the problematic areas that impede listening comprehension, and as a result students' development which is the end goal of assessment is more easily achieved. Through mediation, Albeeva was able to identify the source of comprehension problems that in one case related to a single lexical item and in another to cultural knowledge .

Birjandi and Ebadi (2010) examined the micro-genetic development of L2 through DA using the internet. Their case study consisted of two participants with the same ability according to traditional static tests. Through DA, they could differentiate between the students' abilities in terms of their responsiveness to online mediation offered by the mediator through the Net. They claimed that the students' responsiveness in terms of time they spent on each item is significantly related to their level of ZPD. The higher the ZPD, the less time they spend on tasks while interacting with mediator. Kozulin and Garb (2002, 2004) considered the

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

usefulness of DA for students studying English as a foreign language. The results of their study confirmed that a DA approach evaluates not just the existent knowledge of students that can be seriously affected by cultural differences but focuses on students' ability to benefit from interaction. Their studies also confirmed that students' text comprehension can be improved significantly via mediating to them relevant cognitive strategies.

Most of DA studies in L2 context reviewed above are interactionist and follow a sandwich format. Almost all of these studies explored the use of DA for a limited number of students, and also they did not take into account the psychometric properties of testing. However, this latter point does not mean that these DA studies lack reliability and validity rather their understanding and interpretation of reliability and validity is quite different. Moreover, the mediation provided to students in such studies which follow a sandwich format is administered in a non-dynamic manner. In other words, mediation phase is administered separately from assessment session, and in fact, assessment and instruction are not fully integrated (Poehner, 2008).

This study explores the use of C-DA as a solution to the above mentioned problems. Being administered in a highly standardized way, C-DA offers three advantages: 1) reliability and validity are taken into account; 2) many students can be assessed dynamically, and 3) mediation is given at the time of assessment not in a separate session.

Purpose of the Study

As mentioned above, the literature on DA shows that many if not all of the studies conducted in L2 context are either interactionist or follow a sandwich format. In case of the former which is very demanding and time-consuming, only few students can be tested, and in case of the latter, mediation is not given in a dynamic way. Due to the problems associated with the application of DA in education and second language learning studies, this study intends to design a Computerized version of DA whose feasibility and psychometric properties are guaranteed. With that in mind, we aim to provide answers to the following two questions:

1. Does C-DA observe the psychometric properties of standardized tests?
2. Is C-DA able to make a distinction between an individual's potential and actual levels of reading performance?

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

Method

Participants

Our sample comprised 104 university students, primarily drawn from TEFL students at MA levels. The participants were randomly selected from different Iranian universities. The mean age of the sampling was 28 years. All of the participants were between 18 and 44 years of age. For all of the participants, Persian was their first language and English was their second language. In order to make our sampling fairly homogenous in terms of their level of proficiency, we just included those students whose non-dynamic scores on the Computerized Dynamic Test used in this study fell one standard deviation below or above the mean and ignored the rest. By doing so, we were left by 77 students (27 males and 50 females) who were at the same level of language proficiency.

Instrument

The only instrument used in this study was Computerized Dynamic Reading Test (CDRT). This designed software is both capable of giving hints to students when students make mistakes and also providing the teacher with two scores (dynamic and non-dynamic).

Test Construction Procedure

We followed a three-step procedure in order to ensure reliability and validity of dynamic test as much as possible. The participants who were randomly selected from three Iranian universities took the test in their universities. The participants in each group were asked to attend a two hour meeting to take the test so that all the participants could work under the same conditions. In the following, we will explain in detail the steps taken to design the test.

Test Preparation

Preparing a valid test for this study was one of the most difficult stages of doing the current research. To find appropriate passages for this dynamic test, many versions of TOEFL, IELTS and GRE books were studied by the researchers. The criteria for selecting the passages were the following: to be in line with strategy-based mediation e.g., hints; not to be biased against or in favor of particular students; and finally to have a readability level suitable for Iranian university students majoring in English including both MA and BA students. In view of what was said, the two reading passages used in this Computerized Test were selected from the book

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

"preparation course for the TOFFL Test" by Philips (2007). Each passage is followed by ten items. For each item, 5 hints have been prepared. The hints were constructed mainly based on the reading skills covered in the above-mentioned book. Students had 20 minutes for reading each passage and 4 minutes for answering each item. On the whole, they had 2 hours to take the test. If they failed to answer an item within 4 minutes, they would lose that item. Finally, upon completion of the test, a scoring file is generated on desktop which contains the following information: two scores for each student, the number of hints used in each item and also the total time spent on the test.

Having selected appropriate passages, we then set out to prepare items for each passage. The original items of the TOEFL iBT could not be exactly copied here since they had multiple choice format which is not compatible with dynamic testing in which we are intentionally giving students hints to find the answer. If a multiple-choice format was used, as soon as a student was given a hint, he or she would know that the answer was wrong and in fact, he was left with 3 alternatives, and in the case of receiving the second hint he had just two alternatives, and so on. In other words, provision of each hint meant the deletion of one alternative. To avoid this, open-ended questions had to be designed so that giving hints to test takers did not make them guess the right answer. The following are the kinds of items used in this C-DA.

1. Identification items, e.g., identification of a word, phrase or a sentence from the text,
2. Writing the appropriate answer in the blank space e.g., to choose from the text or write of their own,

For example: The internal <u>dynamics</u> of the decision unit are necessary for reaching a good decision. Write the word from the passage (only paragraphs 4, 5 and 6) that has the same meaning as "dynamics "In the following box.

3. Choosing the appropriate alternatives from among a number of alternatives given in the item. Here is an example:

   Which of the following have been stated about *lateral line system* in the passage?

   - Type the Roman numeral or numerals in the following box. [the correct answer may be one numeral at least and maximally 3 number].

   I. it contains lines of pores

   II. In fish, it is similar to sense organs in other animals.

   III. It can detect movement in the water

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

IV. It creates a visual effect

4. Insertion items. Test takers are asked to put a given sentence somewhere in the passage that best fits and completes the meaning. Such items require the test takers to pay especial attention to the linguistic context before and after the insertion square for any ideas that the given sentence fits the appropriate place in the text (Philip, 2007). In other words, test takers have to be cognizant of the various cohesive devices that are used to show connections and signal relationship (Nuttal, 1996).

The original passages (non-dynamic form), were followed by 13 and 18 items respectively and each was given 30 minutes to be answered. In the dynamic form of the test, the number of items was reduced to 10 for each passage. These ten items roughly mapped on to the skills covered in TOEFL iBT book.

After passages and items were prepared, the next step was to prepare 5 hints for each item. The hints were arranged from the most implicit to the most explicit. Naturally, in the first hint which is the most implicit one, the purpose is just to signal to the test takers that their answer is wrong and in this way, they are given the chance to go back to the item and try it again, and in the last hint test takers are provided with the right answer. These two hints usually take the following format in the test:
Hint 1 → Your answer is wrong, try again.
Hint 5 → The right answer is …….

Although the first and the last hint of each item remained fixed through the test, the rest of the hints were mainly strategy-based, and their format and composition varied from item to item depending on the skill involved in that item. In fact, these three hints were mainly taken from the "How to answer the question" section of TOEFL, IELTS and GRE books. In this part of these books which directly follows each skill of reading, there are some helpful guidelines for students about how to answer reading comprehension items. In other words, the above mentioned books along with some other books on reading were a rich source of inspiration for those three hints. An example of an item in the dynamic test is presented in Appendix A.

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

The Software Preparation

The software package has been designed in such a way that any PC can run it easily. It can be installed properly on any computer provided that it has NET Framework software installed on it. On the opening page of the software, test takers need to type the required information about their names, age and majors in the blank spaces. The next page of the software provides test takers with a short and simple description of the software and dynamic assessment as well in English. After reading the description, test takers can start the test. By starting the test, the first passage and its first item appear on the screen. They have 20 minutes to read the text. Having studied the passage, test takers can start answering the items. While test takers are answering the items, they have an unimpeded view of the passage. They have four minutes for each item, and if they cannot answer an item within four minutes, the software will move automatically to the next item, and test takers will miss that item. If a student gives a wrong answer to an item, the software will provide him with hints until he gets to the right answer in the fifth hint. When the test is over, a scoring file is created on the desktop. As it is shown in Appendix B, the following information about each test taker is stored in this file:

1. Test taker non-dynamic score. This score is calculated according to the students' first try of each item. In fact, this score is exactly the same as that obtained in traditional tests. To make it comparable with dynamic score of the test, we calculated this score on a scale of 0 to 100 points; five points for each item.
2. Test takers' dynamic score. This score is calculated according to students 'use of hints. However, it should be noted that the same hints are given to all test takers in the same order and indeed what makes a difference between test takers in terms of ultimate score is whether they have used more the explicit or the implicit ones on average. The number of hints used by each test taker is subtracted from the total number of hints which is 100. The number that is obtained by this subtraction is the dynamic score. For instance, imagine that a student uses two hints for the first ten items of the test; that is, two hints for each of these items. This student's dynamic score is 80 which is calculated by subtracting the number of hints used by him (here 20 hints) from 100. The non-dynamic score of the same student would be 50 because this student has given wrong answer to the first 10 items of the test, and only after receiving hints he was able to get to the right answers.
3. The number of hints used in each item. The software also takes into account those items missed by test takers by marking the letter "M" in front of them.

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

This mark shows that test takers could not answer the items within four minutes.

4. The total time spent on the test.

Test Piloting

Our major goal of piloting the test was to collect information on the test usefulness for the purpose of making revisions in the test. It was in this phase of test making that a number of ELT experts were asked to substantiate the content validity of the test. To standardize the test, the researchers administered the test to a pilot group of 10 students who had roughly the same language proficiency level as the participants of the study, but they were not selected for the study. Specifically, this pre-testing was done to get qualitative feedback on the test e.g., their feelings and reactions toward C-DA, and also to achieve the following objectives.

1. Making some modifications in the content of test including both the items and hints,

2. Making some modifications in the software package. Having received constructive feedback on the general layout of the software initially from some experts in ELT and later on from test takers, a couple of major changes were made in the software as well. For instance, initially test takers did not have complete view of the text on the screen. Therefore, they had to scroll up and down to read the text, and this made them uncomfortable when they were taking the text. The problem could be overcome by putting the entire text on the screen so that test takers could see the whole text at a singe glance. However, the problem of not seeing the whole passage still remains for those computers with low configurations and the monitors with low resolution.

Data Analysis

The internal consistency of both dynamic and non-dynamic tests is examined using KR-21 method of estimating reliability. The t-test was used to determine the statistical significance of the difference between means of dynamic and non-dynamic scores in order to find out if C-DA has been successful in enhancing students' development of reading comprehension.

The Pearson product moment correlation coefficient was run between both dynamic and non-dynamic scores and also between non-dynamic and gain scores. The former correlation coefficient indicates the effect size and also the concurrent validity of DA and the latter tells us about students' potential for learning

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

The Learning Potential Score (LPS) formula developed by Kosulin and Garb (2002) was used to estimate students' potential for learning. In fact, this score along with correlation between non-dynamic and gain scores are good indicators of considerable potential of DA procedures for measuring students' potential.

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

where
Spost = dynamic scores
Spre = non-dynamic scores
MaxS = the highest dynamic score gained in this test

Results

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
The reliability and validity of the test

	Non-dynamic	Dynamic
Reliability	0.64	.70
Validity		.66

According to Farhadi, Jafarpur and Birjandi (1994, p. 154), "validity and reliability coefficients below .50 are considered low, .50 to .75 are considered moderate, and .75 to .90 or above are considered high"; so, the tests used in this study have moderate reliability (r = .64, .70) and validity (.66).

According to some experts in the field of DA (Poehner, 2007; Lidz & Macrine, 2002), if DA results in significant improvement of test takers' performance, it has construct validity. Comparison of non-dynamic with dynamic gains of the 77 students participated in this study indicates a change of mean scores from 47 to 76. In order to determine the statistical significance of the difference between means on these two sets of scores, we used t-test. The results of the t-test reveals (Table 3)

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

that there is a significant difference between NDA and DA ($t = 3.9$, $p < .05$). It means that the test takers have outperformed in dynamic test.

Table 2
The t-value and standard deviation for dynamic and non-dynamic tests

Two types of test	mean (SD)	df	t	P
Non-dynamic test Dynamic test	47 (7/90) 76 (7/38)	152	3.90	.05

Another way to show that whether test takers actually benefited from mediation is via calculation of the effect size. The effect size can be measured in a number of ways of which Pearson r correlation is one. In this case, the value of effect size varies between -1 to +1. According to Cohen (1992) the effect size is low if the value of r varies around 0.1. The effect size is called medium if r varies around 0.3, and the effect size is called large if r varies more than 0.5. The correlation between dynamic and non-dynamic scores in this study is (.66) which is considered a large effect size (Table 1). Figure 1 is a display of the fact that while on the non-dynamic test no student could get a high score (e.g., 75 or higher), on the dynamic test the majority of test takers (65%) could obtain a score of 75 or higher.

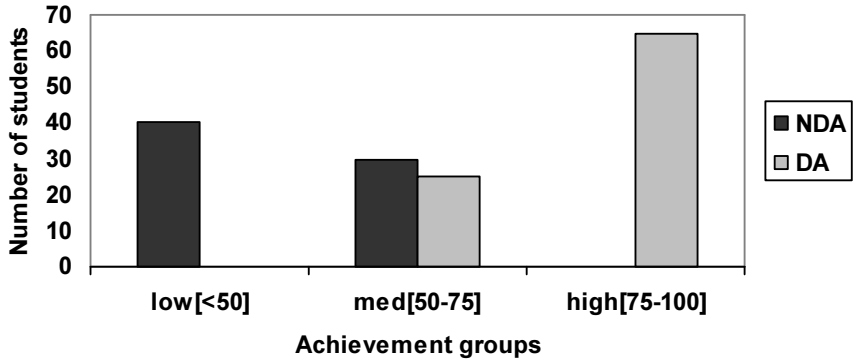

Figure 1: Students in different achievement groups in dynamic and non-dynamic tests

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

While dynamic and non-dynamic scores are highly correlated (.66), the correlation between gain scores and non-dynamic scores is drastically reduced to .33 (see Table 3). This means that the non-dynamic scores reflect test takers' actual performance level, and the scores on dynamic test reflect their learning potential; something which remains hidden in traditional assessment.

Table 3

Non-dynamic scores and their relationship to dynamic and gain scores

Correlation between non-dynamic and dynamic scores	Correlation between non-dynamic and gain scores
.66	.33

Table 4 clearly shows how two different students with the same non-dynamic score could benefit from mediation. Though for each of non-dynamic scores, the highest and the lowest dynamic counterparts have been presented in Table 4, almost all test takers could improve their performance level in varying degrees. This is true of all test takers; whether with high or low non-dynamic scores. So, we can affirmatively answer the second question posed in this study. Indeed, the value of DA in general and EFL dynamic assessment in particular lies in its ability in distinguishing between test takers' initial performance level and their learning potential.

Table 4

Students' dynamic, non-dynamic, gain and learning potential scores

Non–dynamic scores	35	35	40	40	45	45	50	50	55	55	60	60
Dynamic scores	71	65	84	55	86	69	86	68	86	75	89	74
Gain scores	55	46	73	25	74	43	72	36	68	44	72	35
LPS[a]	1/19	1/06	1/43	0/81	1/42	1/03	1/36	0/96	1/30	1/06	1/32	0/98

[a] Learning potential score

In order to determine test takers' learning potential, Kozulin and Garb's formula (1992) was used. As the above authors contend "this formula provides a theoretical basis for distinguishing between high learning potential and low learning potential students" (p. 73). In the last row of Table 4 are presented the learning potential

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

scores of the test takers with the same non-dynamic scores. Those students who made considerable progress from non-dynamic to dynamic test had high LPS, and those who made slow progress had low LPS. Consider the example of those two students whose non-dynamic scores were 40. Student M. who made a significant progress (from 40 to 84) has a very high LPS = 1.43, while student S. who could achieve only a small increase (from 40 to 55) has a very low LPS = .81.

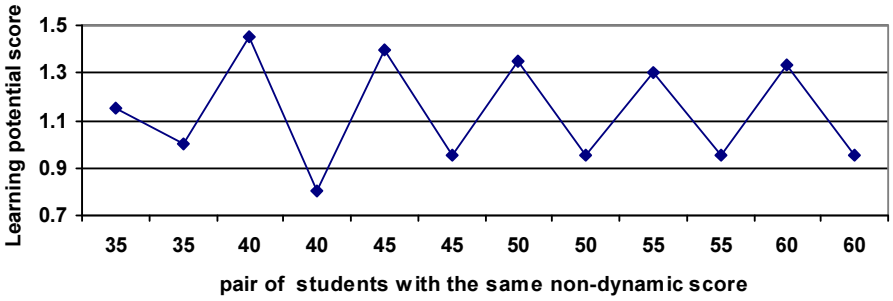

Figure 2: Distribution of learning potential scores among test takers with the same non-dynamic score presented in Table 4

As one can see in Figure 2, the LPS of test takers in this study ranges from 0.81 to 1/43. So, if we assume that 1/3 and above it as the high learning potential score, and 1 and below it as low learning potential, it can be observed that there are some students with the same non-dynamic score but different LPS and also there are some test takers with different non-dynamic scores but with almost the same LPS. This is indicator of the fact that learning potential gives us a better understanding of test takers' learning ability. That students benefited quite in varying degrees indicates that DA has higher statistical validity by reducing the test bias through teaching students the rules of the game; that is, strategic mediation telling them how to approach reading the passage more effectively.

Discussion

Determining reliability in DA procedures is difficult since as Haywood and Lidz (2007, p. 329) state "one sets out deliberately to change the performance of examinees". One way to ensure reliability according to Haywood and Tzuriel (2002) is to make use of tasks whose reliabilities are already well-established in the static or non-dynamic mode. In other words, the dynamic test will enjoy a higher

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

reliability if the tasks used in DA have high reliability in static mode; it was for this reason that the reading comprehension tests used in this study were selected from TOEFL iBT. Although some minor changes were made in the original formats of the tests, the non-dynamic test still enjoyed an acceptable reliability. According to Poehner (2007), reliability takes on considerable importance in DA especially among those who argue that DA should meet the standards of psychometric testing. C-DA enabled us to administer the test in a highly standardized way. This standardized approach, though, did not provide truly individualized intervention in response to emerging needs of test takers could, to some extent, ensure the reliability of dynamic test (r = .70). The moderate reliabilities of non-dynamic and dynamic tests in this study demonstrate the overall effectiveness of mediation provided in the form of hints. Concerning the importance of reliability in DA procedures, it should be noted that the gain in dynamic test can be attributed to intervention only when both the pre-test and post-test instruments e.g., non-dynamic and dynamic tests have acceptable reliabilities; otherwise, change in DA, as Haywood and Lidz (2007) put it "could be merely random or unsystematic variability in the test" (p. 329).

As for the issue of validity; first the content and concurrent validities are discussed and then the most important of all, construct validity is explained. The content validity in DA seems to play a role similar to that in NDA since DA is more concerned with the "how" of assessment than with the "what" of assessment (Haywood & Tzuriel, 2007). By the way, the content validity was substantiated by a number of ELT specialists and also by selection of tests from TOEFL iBT. Moreover, the concurrent validity of C-DA was found to be moderate. As mentioned earlier, the significant gain of students from non-dynamic to dynamic test explored the issue of 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. This study exhibits that C-DA has helped test takers to improve their scores.

The results of our study indicate that the integration of instruction and assessment via computer in which pre-determined hints (mediation) are given to test takers is both feasible and effective not only in improvement of students' performance in EFL reading comprehension but also in providing us with information concerning students' learning potential. These findings are consistent with Haywood and Lidz (2007, p. 2) who state that "if the results of DA are to be

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

directly relevant to educational setting, the content must be related to educational domain e.g., reading comprehension"; to put it in another way: DA procedures are useful both in the field of cognitive performance and also in curricular domains such as EFL reading comprehension. The findings of our study are also similar to those of Poehner (2007) and Kozulin and Garb (2002, 2004). Poehner demonstrated how DA interactions can promote development and provide insights into the learners' functioning. Similarly, Kozulin and Garb (2002, 2004) indicated that DA is effective in both improving students' reading ability and understanding about their potential for learning. Obtaining information about learners' potential allows us to have a true picture of their abilities. Anton (2009) drew similar conclusion by saying that teachers will misrepresent learners' abilities if they consider only the results of traditional tests.

When comparing test taker' performance on non-dynamic test with their performance on dynamic test, it can be observed that their improvement from non-dynamic test on average is about 30%. By providing students the "rules of the game", DA is capable of revealing the maximum level of performance by providing mediation (Pena, Iglesia & Lidz, 2001). The better performance of students on dynamic test can be interpreted in another way. By automatically providing mediation when needed and also by automatically generating each individual's profile, C-DA allows for individuals' self-assessment. In addition to being a tool for assessment of reading comprehension, it makes test takers more involved in their process of learning. Such a test can help students overcome their non-intellective factors such as lack of motivation, fear of failure and anxiety by making the second language assessment more learner-friendly. When students become part of the whole process of language learning and also aware of their progress, they tend to take more charge of their own learning and consequently make the most of their cognitive capacity. Indeed, DA according to Haywood and Lidz (2007) is part of idiographic enterprise in which each individual is not compared to others rather comparison is within person not with reference to performance of others. Thus, when comparing their own performance on dynamic and non-dynamic test, they become aware of their own progress.

Conclusion

In our view, using dynamic assessment including C-DA is an important step to shifting the paradigm of "teaching to the test movement" which is undemocratic (Shohamy, 2001) and capitalistic (Kanpol, 1999) in nature, to a "testing to the

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

teaching movement" whose honest goal would be to help students learn something. One point which should be emphasized is that the inclusion of DA does not imply the exclusion of traditional tests from the system of language education. In fact, both of these types of assessment are complementary rather than contradictory. Moreover, the outcomes of the study reveal the fact that computers can help us to put into practice the major principles of DA, allowing teachers to have mass testing in class.

In the end, it is our hope that our instrument can pave the way for more research into this area, making teachers and testers employ more DA in language education. In this study, the effect of C-DA was examined on reading comprehension. Other studies can be done in relation to other skills e.g., listening, writing and grammar. Also, researchers are recommended to replicate this study to demonstrate its validity and also to confirm the effectiveness of C-DA in reading comprehension.

Note on Conributors:

Reza Pishghadam *is an associate professor of TEFL at Ferdowsi University of Mashhad. He has published more than one hundred articles and books in different domains of English language education. In 2007, he was selected to become a member of the National Association of Elites of Iran. In 2010, he was classified as the top researcher in humanities by the Ministry of Sciences, Research, and Technology of Iran. Emails: rpishghadam@yahoo.com/ pishghadam@um.ac.ir*

Elyas Barabadi *is a PhD student in TEFL in Shiraz University. He did his MA degree in TEFL in Ferdowsi University of Mashhad. His research interests are: Psychology and Sociology of language education. His email address is: Elyasbarabadi@yahoo.com.*

References

Afflerbach, P. (2007). *Understanding and using reading assessment.* Newark, DE: International Reading Association.

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. 112-134). London: Equinox.

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

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*, 465-483.

Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, *42*(3), 576-598.

Birjandi, P., & Ebadi, S. (2010 January). Dynamic assessment and transcendence of oral abilities in EFL context. *Paper presented in IELTI5, University of Tehran, Tehran, Iran.*

Cohen, J. (1992). Quantitative methods in psychology. *Psychological Bulletins*, *112*(1), 155-159.

Farhadi, H., Jafarpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. Tehran: SAMT.

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 University Press.

Haywood, H., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education, 77*(2), 40–63.

Jacobs, E. (2001). Screening test: The effects of adding dynamic assessment components to a computerized preschool language. *Communication Disorders Quarterly, 22* (4), 217-226.

Kanpol, B. (1999). *Critical pedagogy: An introduction*. London: Bergin & Garvey.

Kester, E. S., Peña, E. D., & Gillam, R. B. (2001). Outcomes of dynamic assessment with culturally and linguistically divers students: A comparison of three teaching methods within a test-teach-retest framework. *Journal of Cognitive Education and Psychology*, *2*(1), 42-59.

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International, 23*(1), 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.

Kozulin, A., & Gindis, B. (2007). Sociocultural theory and education of children with special needs, from defectology to remedial pedagogy. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 332-336). Cambridge: Cambridge University Press.

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching Journal*, *42*(3), 355-368.

Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin et al. (Eds.), *Vygotsky's educational theory in cultural context* (pp. 100-105). Newyork: Cambridge university press.

Lidz, C. S., & Macrine, S. L. (2002). An alternative approach to the identification of gifted culturally and linguistically diverse learners, The contribution of dynamic assessment. *Mensa Research Journal, 33*, 15-35.

Lurie, K., Pecsenye, M., & Robinson, A. (2009). *Cracking the GRE*. New York, NY: The Princeton Review, Inc.

Mc Carter, S., & Ash, J. (2001). *IELTS reading test*. Intelli Gene.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.

Peña, 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. (2001). *Longman complete course for the TOEFL test.* New York, NY: Addison-Wesley Longman, Inc.

Phillips, D. (2003). *Preparation course for the TOEFL test, the paper test*. New York, NY: Pearson Education. Inc.

Phillips, D. (2007). *Preparation course for the TOEFL test, iBT*. New York, NY: Pearson Education.Inc.

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, *9*, 233-265.

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.* Springer.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing, 18* (4), 373–391.

Sternberg, R. J., & Grigorenko, E. L., (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.

Tzuriel, D. (2000). Dynamic assessment of young children: educational and intervention perspectives. *Educational Psychology Review, 12* (4), 385-420.

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

Tzuriel, D. (2003). Foundations of dynamic assessment of young children. In A. S. H. Seng, L. K. H. Pou & T. O. Seng (Eds.), *Mediated learning experience with children. applications across contexts*. Singapore: McGrawHill.

Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning, 18*, 21-32.

Tzuriel, D., & Shamir, A. (2007). The effects of peer mediation with young children (PMYC) on children's cognitive modifiability. *British Journal of Educational Psychology, 77,* 143-165.

Vygotsky, L. (1978). *Mind in society.* US: President and Fellows of Harvard College.

Wertsch, J. V. (1984). The zone of proximal development: some conceptual issues. In B. Rogoff & J.V. Wertsch (Eds.), *Children's Learning in the "Zone of Proximal Development.* San Francisco, CA: Jossy-Bass.

Appendices

Appendix A

An Example of an Item in Dynamic Test

6. Indicate where the following sentence could be added to the passage. Click on the square [■] to add the sentence to one of the paragraphs: four, five or six.

"However, these same routines and procedures can also have an inhibiting effect on the ability of the organization to arrive at optimal decisions ".

a) Your answer is wrong, try again
b) You can add the sentence to paragraph four since the effect of routines and procedures on decision making is discussed only in this paragraph.
c) Look at the word "however" at the beginning of the given sentence. We use "however" when we want to add a fact or piece of information that is very different from what we have just said .for example "this is a cheap and simple process, however, there are dangers". Now, look where in the paragraph a fact has been said about "routines and procedures" that is different from this sentence.
d) pay attention to the sentence beginning with "pre-planned routines and ……."this is where a fact has been stated about routines; it says routines are necessary for decision making in large organization
e) the right answer is square 6D

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]

Appendix B
An Example of Scoring File Created on the Desktop
Name: Kermanshahi
Age: 23
Gender: Female
Major: Teaching English MA
Score gained with the use of hints: 90
The total number of hints used: 10
Score gained with no hint: 75

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of hints used	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	2	4	0

[Downloaded from c4i2016.khu.ac.ir on 2024-10-12]