Article

LANGUAGE
TEACHING
RESEARCH

Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA)

Language Teaching Research
17(3) 323–342

DOI: 10.1177/1362168813482935
ltr.sagepub.com
SAGE

Matthew E Poehner
The Pennsylvania State University, USA

James P Lantolf
The Pennsylvania State University, USA

Abstract

Dynamic Assessment (DA) derives from Vygotsky's (1987) insight that the use of mediation, attuned to learner needs, enables learners to perform beyond their current level of functioning, thereby providing insights into emerging capabilities. The instructional quality of mediation means that this process simultaneously creates possibilities for development, even over the course of a single session, a phenomenon referred to as 'microgenesis' (Wertsch, 1985). An important feature of DA is transfer, or transcendence: the process of tracking learner development as it unfolds not only with regard to control of tasks included in an assessment but also more complex and difficult tasks that build on those included in the original instrument (Campione, Brown, Ferrera, & Bryant, 1984; Feuerstein, Rand, & Hoffman, 1979; Poehner, 2007). The present study reports on the use of DA principles in tests of second language (L2) listening and reading comprehension delivered through an online format. The tests, which incorporate mediation, include transfer items to determine the extent to which learner development is supported during the test. This process generates three scores: actual (to capture unmediated performance); mediated (reflecting responsiveness to assistance provided on each test item); and a learning potential score (based on gain between actual and mediated performance) that indicates how much investment in future instructional activity is likely required for development to move forward.

Keywords

Computer-based testing, dynamic assessment, L2 comprehension assessment, learning potential, sociocultural theory

Corresponding author:
Matthew E Poehner, Department of Curriculum and Instruction, The Pennsylvania State University, 159 Chambers Building, University Park, PA 16802, USA.
Email: mep158@psu.edu

1 Introduction

Much has been made in popular media over the notion of 'teaching to the test' (Ravitch, 2010; Sacks, 1999), a phenomenon usually viewed negatively because it limits students' educational experiences to those believed to best prepare them for the narrow context of high-stakes, often standardized, testing. Such discussions focus on the quality of teaching and learning prior to test administration, assuming that this process is suspended during test administration, and that learner abilities are relatively stable throughout the course of the testing procedure. Indeed, in the psychometric literature, the possibility of learner abilities changing during a test is referred to as 'instrument decay' and is seen as highly problematic for drawing inferences on the basis of test performance (Glutting & McDermott, 1990).

A markedly different set of assumptions is proposed in Dynamic Assessment (DA). DA is predicated on the thesis that human abilities emerge through participation in activities in which individuals are guided and supported by others and through the use of available cultural resources. That is, the development of knowledge and cognitive abilities is contingent upon sociocultural forms of mediation (Sternberg & Grigorenko, 2002). This premise is the cornerstone of Vygotsky's (1978, 1987) psychological theory. The most well known of Vygotsky's (1987) formulations, the Zone of Proximal Development (ZPD), is both a concrete expression of his social ontology of mind and the theoretical basis of DA. As Poehner (2008) explains, DA procedures include in their diagnoses both learner independent and externally mediated functioning. The former is interpreted as an indication of abilities that have fully formed, as is the case in most forms of assessment. What differentiates DA from traditional assessment is that collaborative functioning with others is given equal, if not greater, attention. Following Vygotsky's (1987) discussion of the ZPD, responsiveness to mediation is an indication that a learner's abilities are in the process of forming and therefore provide a fuller picture of their developmental trajectory. Indeed, second language (L2) DA researchers (e.g. Ableeva, 2010; Poehner, 2008) argue that leading learner development through ZPD activity can occur over the course of a multi-session assessment program or even during a single interaction, a phenomenon Wertsch (1985) has termed 'microgenesis'.

While microgenesis has primarily been explored in the context of moment-to-moment interaction between learners and mediators (see Aljaafreh & Lantolf, 1994; Lantolf & Poehner, 2011; Poehner, 2008), the present article reports on a project that incorporates principles of mediation into a computerized approach to DA. Specifically, the study explores the concept of microgenesis in the context of a Computerized DA (C-DA) of L2 Chinese, French, and Russian listening and reading comprehension. The tests were designed to distinguish learners' independent from mediated performance, to predict their learning potential, and to uncover evidence of learning through integration of transfer items into the tests.

Before describing the tests and the results of their initial administration, we will first review the relevant literature on C-DA, transfer of learning, and learning potential. We also point out that in developing the tests with a focus on efficiency and breath of administration, it was necessary to compromise the more fine-grained and individually negotiated modes of mediation that have characterized most L2 DA research to date. Potential

advantages and disadvantages of standardized procedures are elaborated in the next section. We are not reporting the results of a controlled experiment but rather an initial attempt to develop an online DA that can be administered to large numbers of students and that generates results in a manner that are useful for teachers and learners.[1]

II Mediation, transfer, and the diagnostics of development

While some DA researchers, most notably Feuerstein and colleagues (Feuerstein, Rand, & Rynders, 1988), have favored an individualistic and dialogic approach to mediation, others, such as Budoff (1987) and his team, have opted for a standardized approach (see Poehner, 2008; Sternberg & Grigorenko, 2002). The terms 'interventionist DA' and 'interactionist DA' (Lantolf & Poehner, 2004) offer a useful distinction between approaches that favor scripted prompts arranged hierarchically and administered sequentially (interventionist) from those in which mediation is negotiated in mediator-learner dialogue (interactionist). Interventionist approaches yield results that are more readily comparable across learners, showcasing for instance the number of prompts required during a given task (Budoff, 1987). Interactionist DA offers greater opportunities to support microgenesis because mediation can be more precisely attuned to a learner's (or a group's) needs. The majority of L2 DA studies has adopted an interactionist orientation in a one-on-one format (Ableeva, 2010; Poehner, 2008), although interventionist DA has also been examined in classroom settings (Davin, this issue; Lantolf & Poehner, 2011).

An important difference between interventionist and interactionist DA concerns the feasibility of employing the procedure with large groups of learners. L2 DA research has illustrated the benefits of interactionism in probing learner understanding of relevant concepts and pursuing various dimensions of a task to arrive at nuanced diagnoses of abilities while simultaneously promoting greater understanding of relevant concepts involved (see Poehner & van Compernolle, 2011). Equally clear, however, is that such approaches are labor intensive and may not be feasible in all contexts. Although standardization limits mediator flexibility, it enhances efficiency in terms of the number of students that can be simultaneously assessed at the cost of restricting the prompts to a one-size-fits-all format.

To our knowledge, only three studies in the general DA literature have reported the use of Computerized DA (Guthke & Beckmann, 2000; Jacobs, 2001; Tzuriel & Shamir, 2002). Each has designed a unique approach to delivering standardized mediation during the procedure. While a full discussion of these models of C-DA is beyond the scope of the present article (see Poehner, 2008), we briefly describe them below.

1 Computerized Dynamic Assessment (C-DA)

The earliest attempt to employ computerized mediation is the 'Lerntest' (learning test) developed by Guthke and Beckmann (2000). The test relied on a series of prompts designed to evaluate various cognitive abilities, including language aptitude (for

discussion, see Poehner, 2008). An incorrect response triggered a tutorial program that reviewed relevant concepts and principles and provided examples. The test then resumed and the learner re-attempted the task in question before moving on. The final diagnosis of learner abilities included those items for which a tutorial was needed and whether the learner then responded correctly to the problematic items. Those items for which a tutorial was not needed are interpreted to reflect a learner's zone of actual development (independent performance) and those for which a tutorial resulted in a correct response were assumed to comprise a learners' ZPD. Aside from their introduction of this model of C-DA, the authors have not reported any subsequent studies of its use.

Jacobs (2001) describes a computer software program known as KIDTALK that assesses pre-school children's language aptitude. The program begins with a video featuring puppets that introduce features of an invented language ('kidtalk'). This is followed by a series of questions designed to assess understanding of the language. The test pauses whenever a learner answers incorrectly and the relevant portion of the introductory video is replayed; the test then resumes, and if the child again answers incorrectly the video is played a final time before allowing a third attempt at the question and moving on. Like the Lerntest, then, KIDTALK allows only one form of mediation (a tutorial module in the former and a replaying of the introductory video in the case of the latter) rather than a graduated series of prompts that become available to learners if they are needed.

The matter of tailoring mediation to learners is taken up in the C-DA approach devised by Tzuriel and Shamir (2002). Working in the domain of cognitive psychology, these authors sought to evaluate young children's seriational thinking abilities, believed to be central to success in learning mathematics. Following an interventionist model, if learners fail to respond correctly to a series of tasks, hints are provided that are arranged from implicit ('try again') to explicit (pointing out important features of a task or offering reminders of relevant principles). Teachers are also present during the administration of the test and are free to provide supplemental support similar to interactionist DA (e.g. check learner understanding of the task, offer different hints and explanations), which is also reported in the final diagnostic. According to Tzuriel and Shamir (2002), the procedure provides more in-depth diagnoses of learner abilities and creates more learning opportunities when teachers are present than when mediation is provided exclusively by the computer.

2 *Transfer*

Transfer holds that introducing assessment tasks that follow the same principles as earlier ones but are more difficult or complex can offer insights into whether learners have internalized mediation previously offered. The most robust theoretical argument for the value of investigating learners' performance on transfer tasks was elaborated by Feuerstein and his colleagues at the International Center for the Enhancement of Learning Potential in Israel (Feuerstein et al., 1988). Researchers there have endeavored to diagnose and support the development of cognitive functions among learners with special needs as well as ethnic and cultural minorities in the context of their interactionist approach to DA known as Mediated Learning Experience (MLE). Standardization is

eschewed in MLE sessions as assessors are expected to use every means available to probe learner functioning and jointly work toward task solutions, all the while guiding learners toward greater understanding of, and control over, relevant concepts and operations (see Sternberg & Grigorenko, 2002, for details). According to Feuerstein, the outcome of MLE is to both move learners to a point where they are able to independently complete tasks that were previously only possible with mediator support and also to help learners orient to more challenging and complicated tasks. Transfer in this model involves tracking learners as they progress from tasks of one difficulty level to the next, noting qualitatively whether and how much mediation they required to make these transitions.

The approach to DA that offers the most extensive study of tracking learner response to new challenges is the Graduated Prompt Approach (GPA) formulated by Anne Brown and colleagues (Brown & Ferrara, 1985; Campione et al., 1984). GPA makes use of scripted prompts, arranged from implicit to explicit, and offered to learners in a standardized format whenever they encounter difficulties responding to test items. Weighted scores reflect the level of prompting required for each item, and this is interpreted as an indication of how close learners are to independent performance. In GPA, just as learners differ with regard to their weighted scores, they may be further differentiated according to their performance on transfer tasks. In a study examining reading comprehension among special needs learners and their 'normal' peers, the initial DA uncovered marked differences in the level of mediation learners in the two groups required to complete the tasks (Brown & Ferrera, 1985). Following an intervention program aimed at teaching relevant strategies and principles, the DA items were re-administered. It was found that differences between the groups all but disappeared, with learners requiring little or no mediation. However, some learners in both groups struggled as changes were introduced to the initial DA tasks. In some cases, the performance of special needs learners deteriorated as the tasks became more distant from the initial ones. This suggests, according to the researchers, that while the intervention had moved the initial assessment tasks from the learners' ZPD to their zones of actual development, differences persisted with regard to their current ZPD, which only became apparent through examining transfer performance.

3 *Transfer in L2 DA*

To date two L2 DA studies have included transfer in their procedures. In Poehner's (2007) study on oral proficiency of L2 French, initial DA sessions determined that learners' control over the preterit and imperfect distinction was emerging, or 'ripening' to use Vygotsky's (1987) term. Although learners were unable to use aspect appropriately during independent narration, through interaction with a mediator they revealed a partial understanding of the concept. A subsequent enrichment program based on learner responsiveness to mediation focused specifically on this feature of the language. Following enrichment, learners repeated the initial DA narration tasks as well as two transfer (TR) narrations. The prompts used for TR narrations were more complex than the original prompt, based on the movie *Nine months*. The first TR task involved a complex, highly graphic and emotional scene from *The pianist* depicting the violence of the Warsaw Ghetto during World War II. The second task asked learners to recount a written

excerpt from Voltaire's *Candide* that followed eighteenth-century literary conventions and therefore included lexical items as well as verb forms that were unfamiliar to the learners.

As in Brown's research with children, Poehner (2007) found that learner performance on the initial DA tasks was more uniform than during TR. For example, although one learner consistently made appropriate aspectual choices during her TR narration, she expressed the need for verification of her choices from the mediator. Another participant exhibited marked deterioration in her ability to temporally frame the TR narratives. When asked to verbalize the reasons behind her aspectual choices, she oscillated between explanations that invoked the systematic linguistic concepts presented during enrichment and statements that reflected the piecemeal rules-of-thumb derived from previous French courses.

Ableeva (2010) extended L2 DA research into the domain of listening comprehension. While the aural texts employed in the initial assessments and during enrichment followed an interview format, in which French native speakers responded to a range of topics (e.g. cuisine, politics, cinema), the TR sessions incorporated excerpts from a television documentary and a radio commercial. According to Ableeva, the latter two texts included more cultural detail, and required knowledge of specialized vocabulary and complex grammatical structures than the interview texts. While learner performance on the original tasks improved following enrichment, performance on the TR tasks varied, with a TV documentary on smoking in French restaurants (near transfer) resulting in better comprehension than did a 20-second radio commercial for a Belgian restaurant chain (far transfer).

Transfer is an important component of DA because increasing the difficulty level of tasks allows one to better diagnose the extent to which learner ability has developed and to predict how likely learners are to react to future instruction. For this reason transfer items were incorporated into the C-DA tests, although following a different format than in Poehner's and Ableeva's respective projects.

4 *Learning potential score*

Sternberg and Grigorenko (2002) point out that a common approach for quantifying performance during DA is to report both independent and mediated scores. This is clearly in line with Vygotsky's (1987) argument for the need to take account of the difference between actual and potential development that comprises the ZPD. The C-DA tests therefore generate mediated and non-mediated (i.e. independent) scores. However, efforts have also been undertaken to represent the difference between these scores through inclusion of a learning potential score (LPS), as some DA proponents have argued that the difference between mediated and non-mediated performance is indicative of a construct termed 'learning potential'. In both Feuerstein's MLE as well as Budoff's (1987) Learning Potential Assessment (LPA) approach to DA, the goal is to capture learning ability on the basis of 'responsiveness' to mediation rather than on more 'static notions' of assessment (Lidz, 1991, p. 26). We are, however, not in full agreement with how learning potential is frequently discussed in the DA literature, where it is often described as a latent intellectual capacity that predicts future responsiveness to

instruction (Feuerstein et al., 1988). Specifically, we are not comfortable with the claim that DA brings out 'hidden' and 'fixed' potential. Consequently, we do not subscribe to the view that ability is somehow hidden inside the person only to be released through mediation. A more sympathetic position with respect to Vygotsky's theory would be to conceive of learning potential as 'openness to mediation' that enables individuals to appropriate (take over) the meditational process and to use it to regulate their own intellectual activity. As Leont'ev (1981) reminds us, abilities are not innate but 'are formed in the course of a person's life, and particularly in the course of performing the corresponding activity' (p. 79).

Budoff categorized individuals who completed initial assessments as high scorers and low scorers (according to their initial test performance) and as gainers or nongainers, depending on their performance on a retest administered following enrichment. While Feuerstein was also concerned with differential responsiveness to mediation, his approach was more focused on 'child analysis' rather than on 'test analysis', Budoff's primary concern (Lidz, 1991, p. 27). In earlier work (Lantolf & Poehner, 2004) we expressed a preference for Feuerstein's orientation, noting that it is more sensitive to individual abilities. At the same time, however, we acknowledged that it sacrifices efficiency given that it is difficult to simultaneously administered individualized assessments to large numbers of learners. The current project, as mentioned, foregrounds efficiency and is more in line with Budoff's orientation.

Building on the work of Budoff and Feuerstein, Kozulin and Garb (2002) proposed a formula for calculating LPS to capture the distinction between independent and mediated performance in a single score. In a study assessing ESL reading comprehension, Kozulin and Garb (2002) proposed a formula for quantifying learning potential that can then be used to characterize learners as high, mid-level, or low, gainers. The authors observed a significant negative correlation between the scores generated during actual performance and those that resulted from mediated learning. They interpreted this to indicate that actual scores do not predict responsiveness to mediation. They also observed a substantial effect size, whereby scores improved by more than one standard deviation as a result of mediation. What is more, several students with identical actual scores performed differently following mediation when they were once again asked to perform independently. For instance, two low-performing students both scored 29% on the ESL test under independent conditions; following mediation, one scored 38% and the other scored 59%. Similarly, two high-performing students both scored 62% initially, but after mediation one only improved to 65% while the other scored 82% on the comprehension test (Kozulin & Garb, 2002, p. 121). The formula given in Figure 1 operationalizes LPS.

Kozulin and Garb (2002) argued that the formula 'provides a theoretical basis for distinguishing between high learning potential and low learning potential students' (p. 121).[2] Using the above scores they observed that of the low performers one generated an LPS of .47 while the other produced a score of .89. Similarly for the high performers, one produced an LPS of .68 but the other manifested a considerably higher potential, generating a score of 1.02. An especially significant point is that one of the low performers had a higher LPS than one of the high performers (e.g. .89 vs. .68) and therefore we would expect this learner to react more favorably to instruction than the high performer

$$\text{LPS} = \frac{(\text{S post-mediation} - \text{S actual})}{\text{Max S}} + \frac{\text{S post-mediation}}{\text{Max S}}$$

Abbreviated as follows:

$$\text{LPS} = \frac{2\,(\text{S post-mediation}) - \text{S actual}}{\text{Max S}}$$

Figure 1. Formula for learning potential score (LPS).
Note. LPS: learning potential score; S: score.
Source. Kozulin & Garb, 2002, p. 121.

with a lower LPS. According to Kozulin and Garb (2002) this type of information is not 'obtainable by means of standard examination' (p. 122). Moreover, it attests to the full significance of Vygotsky's (1978) stance on development and the ZPD.

Kozulin and Garb (2002, p. 121) organized the learners in their EFL study into three subgroups based on their LPS scores: High scorers (LPS ≥ 1.0); mid-range scorers (.88 ≥ LPS ≤ .79) and Low scorers (LPS ≤ .71). An LPS of 1 or higher indicates a positive change of one standard deviation, while scores under 1 show a change of less than one standard deviation. We adopted the scoring procedure for C-DA tests as well as described in the next section.

III Description of the assessment instruments[3]

In our initial attempt to develop C-DA tests, we decided to focus on language comprehension rather than production and to integrate a multiple-choice response format with standardized meditational prompts.[4] The tests of listening comprehension consist of a series of short passages that learners hear twice prior to presentation of the questions followed by multiple-choice response options. The reading tests diverge somewhat with the Chinese test comprised of a series of short passages, and the French test consisting of much longer passages. The questions and response options are the same as in the listening test. All questions and options are in English to avoid a situation wherein learners comprehend the passage but encounter difficulties with the language in which the questions and responses are framed. To maximize the number of prompts, the tests incorporate a correct choice and four distracters. To the extent possible, the prompts are sensitive to the particular construct that an item is intended to assess. For reading comprehension the constructs are lexis, grammar, discourse, and culture. The listening tests include these constructs plus phonology.

To establish appropriate levels of mediation, each test question was piloted in a one-to-one administration during which a mediator interacted with individual learners. This process allowed for the identification of prompts and mediating moves

(e.g. narrowing the search space within a text) that proved helpful and that were subsequently written into the C-DA program. Thus, the mediating prompts for each item on the C-DA tests focus learner attention on the key portion of a text where the correct response is found. With each successive prompt the search space is reduced and, when relevant, a prompt also suggests that a specific piece of information or a relationship between two propositions in a text should be kept in mind as an item is re-attempted. For instance, the French reading test includes a brief first-person narrative that reflects on weekends the author spent as a child with his grandparents in the country. The accompanying test item asks students which part of those memories is most meaningful for the author (for screen shots of the text, item, and mediation sequence, see the appendix). Successful comprehension is contingent upon a rather complex three-part construction wherein the author states that while he was fond of all aspects of those weekends, *ce que j'aimais le plus* ('what I liked the most'), *ce n'était pas les animaux* ('it wasn't the animals'), *mais plutôt les soirées calmes avec mon grand-père* ('but rather the quiet evenings with my grandfather'). In addition to drawing attention to relevant parts of the text the prompts also include statements such as 'Does the author say what he liked most here [in the highlighted portion]?' and 'The author says it wasn't the animals he liked the most, but …'

Throughout the tests, an initial incorrect response is met with the following statement, presented aurally in the listening test and in writing in the reading test: 'Sorry. That's not the right answer.' The first prompt is then presented, either by highlighting in colored font or by auditory repetition of a chunk of text that the learner should pay attention to. The learner then selects from among the remaining four response options. Following a second incorrect response the learner is again informed that the response is incorrect: 'Sorry. That's not right either.' At this point the scope of the search space is further narrowed. This process continues for a total of four attempts, at which point the correct answer is revealed accompanied by an explanation. At any point where the correct response is selected the option of viewing an explanation supporting the correct response is provided. This procedure is intended to help learners who answer correctly but who may be uncertain of their response or who guess the correct answer.[5]

The tests were administered to students enrolled in intermediate (fourth semester) university courses. We do not believe that the relative proficiency of the French and Chinese students is equivalent because of the relative difficulty of each language for speakers of English as a first language. The French reading test includes related passages of variable length while the Chinese test comprises a series of shorter independent passages. The motivation for the difference is that the development team for Chinese did not believe intermediate level students in this language were capable of dealing with an extended text, while the French team considered this option reasonable.

1 Scoring procedure

The tests generate two weighted numerical scores, and an LPS. If an initial response to an item is correct, a score of 4 is awarded for that item. If a second attempt at the same item produces a correct response, a score of 3 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 4s are

1. Which sentence is not true?
   a. Simon studies every night.
   b. Simon wants to teach in Marseille when he completes his degree.
   c. Computer programming is one of Simon's passions.
   d. Simon goes to the movies on Saturdays.
   e. Simon currently lives in Toulouse.

Figure 2. Computerized dynamic assessment French listening item (lexical).

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'. The LPS is then calculated using Kozulin and Garb's formula, which, as we have discussed, takes account of the difference between actual and mediated scores.[6]

2 *Transfer and non-transfer items in C-DA*

Transfer items were integrated into each test without any overt identification of their function.[7] The French tests included five transfer items, as did the Chinese listening test; however, the Chinese reading test included eight transfer items. The transfer items paralleled the other test items in terms of response options, prompts and focus on a particular construct relevant to comprehension. The sole difference between transfer and non-transfer items concerned their level of difficulty. As explained, each test item targeted a particular dimension of reading or listening comprehension, and this was also true of the transfer items. Of course, the transfer items did not intend to simply determine whether learners could recall a particular lexical item or grammatical form that had appeared earlier in the test. Rather, we were interested in determining if they were better able to interpret written or aural texts or if they continued to experience the same level of struggle with regard to the lexical, grammatical, discursive, cultural, and phonological dimensions of the texts.

To appreciate more clearly how transfer items function we will consider two items from the French listening test which target lexical knowledge. For item 1 (see Figure 2) the correct response is contingent upon recognizing the basic vocabulary used to describe the lifestyle of Simon, a university student in Toulouse. The text briefly sketches Simon's goal of becoming a teacher, his interest in computer programming, and his habit of accompanying his girlfriend Sylvie to the cinema on Saturdays. The challenge posed by this item is that learners are asked to identify the incorrect statement from among the five options presented. To have full confidence in responding, therefore, learners must comprehend nearly the entire passage and cannot rely on catching the appropriate line that contains the answer. That said, the vocabulary employed in this particular text would likely be encountered by most learners of French very early on in their study of the language. Moreover, the context – the life and schedule of a university student – is probably familiar to most learners.

11. According to the passage:
    a. The family is vegetarian.
    b. The family almost always has a dessert.
    c. The family eats mostly meat.
    d. The family rarely has dessert.
    e. The family always eats cheese with their meal.

Figure 3. Computerized dynamic assessment French listening transfer item (lexical).

The correct answer is (a) 'Simon studies every night.' In the text, the speaker explains that Simon does indeed study Monday through Friday but that on Saturday night he and Sylvie go to the cinema. In this way, option (a) will appear to be true to learners who understood the phrase *Simon étudie tous les soirs* ('Simon studies every evening') but who failed to comprehend the full utterance, *Du lundi au vendredi, Simon étudie tous les soirs mais le samedi, il va au cinéma avec son amie Sylvie* ('Monday through Friday, Simon studies every evening but on Saturdays he goes to the cinema with his girlfriend Sylvie'). Thus, while the passage provides a good deal of detail for learners to sift through, its presentation is relatively straightforward. This stands in contrast to the complexity of the transfer items on the test, such as item 11 (Figure 3).

The aural passage accompanying item 11 is a description of the speaker's family dinner routine. The speaker explains the several courses that the family enjoys and lists examples of the kinds of dishes commonly served at each course. Although item 11 appears less demanding than item 1 – it only asks learners to identify the correct statement rather than eliminating four correct statements to determine an erroneous one – the passage it pertains to is more sophisticated.

The correct answer is (b) as is stated explicitly in the passage. However, this is one detail mentioned in a list-like series of statements that might easily point learners to other options. For instance, the speaker states that the family eats *tous les types de viandes* ('all types of meat') and goes on to mention beef, pork, duck, and rabbit but also states that main courses are composed primarily of vegetables rather than meat and adds examples of the vegetables the family frequently enjoys (beans, peas, carrots, and potatoes). She also notes that her father eats cheese after the main course. Each of the distracters, then, might appeal to learners. Even the reference to desserts, which is the key to determining that (b) is correct, includes the phrase *tout le temps* ('all the time') that learners must recognize in order to understand precisely what is being said about desserts.

Item 11 is more difficult than 1 due both to the lexical density of the passage as well as its content (i.e. French dining habits rather than the more familiar context of university student life). Again, the point in comparing these two items is not simply to determine whether learners appropriate and transfer a given lexical item (e.g. *tout le temps*) from an earlier to a later test item but to ascertain whether learners show signs of improvement in recognizing and interpreting words and phrases in the target language more generally (an important component of reading ability) over the course of the test.

Table 1. Maximum score, mean, actual and mediated scores, gain scores, reliability coefficients.

	Chinese listening	Chinese reading	French listening
Number of learners	62	80	21
Maximum score	60	76	40
Mean actual score	29.92 (15.57)	38.51 (18.12)	27.05 (8.19)
Mean mediated score	41.71 (10.63)	57.42 (11.44)	38.24 (4.65)
Mean gain score	14.63 (6.54)	19.35 (8.25)	7.19 (4.31)
Reliability coefficient	.955	.955	.890

Table 2. Two-tail *t*-tests and effect size comparing actual × mediated scores.

	Chinese listening	Chinese reading	French listening
t-value	3.143	4.945	2.319
Significance	$p < .01$	$p < .01$	$p < .025$
Effect size	1.39	1.65	1.54

IV Analysis

1 General statistical analysis of the tests

Table 1 presents the descriptive statistics and reliability coefficients for the three C-DA tests under consideration. The mean actual and mediated scores are based on performance without the transfer items. The transfer scores, which we analyse separately in conjunction with the LPSs, serve as an indication of learning during the test-taking process and therefore should not be integrated with the regular test items; see Table 2.

Reliability coefficients were calculated using the automatic calculator for K-R 21 available at http://www.cedu.niu.edu/~walker/calculators/kr.asp (accessed March 2013). The gain scores indicate the change between the actual and mediated components of the tests. As expected, performance improved under mediation. Notice also that with mediation, performance became more homogeneous as indicated by the decrease in the standard deviation from the actual to the mediated scores.

Two-tailed *t*-tests comparing the difference between actual and mediated performance, as reflected in gain scores, across all three tests are statistically significant. The effect size, which is substantial for each test, was calculated using the gain between actual and mediated scores divided by the standard deviation of the mediated score.

2 Evidence of learning

To uncover evidence of learning during the tests, we analysed the transfer items. Space does not permit us to analyse the performance of each learner. We therefore limit our analysis to several examples that illustrate the value of distinguishing actual from mediated scores and of the importance of incorporating transfer into dynamic assessments.

Table 3. Individual, actual, and mediated scores, learning potential scores (LPSs) and transfer scores: Chinese listening test.

Learner	Actual*	Mediated	LPS	Transfer**
1	8	40	1.20	16
2	40	54	1.13	12
3	48	57	1.10	24
4	40	53	1.10	16
5	40	52	1.07	28
6	16	40	1.07	16
7	28	44	1.00	12
8	32	46	1.00	12
9	28	43	0.97	12
10	36	46	0.93	8
11	24	40	0.93	8
12	24	34	0.73	12
13	16	20	0.70	8
14	16	28	0.67	8

*Notes. * Maximum score = 76; ** Maximum score = 32.*

Table 3 provides actual and mediated scores on the main portion of the test as well as scores for the transfer items along with LPSs for several learners on the Chinese Listening Test. Keep in mind that mediated scores do not directly reflect learning, but they are an indication of how well learners responded to mediation: the mechanism through which learning occurs during the test.

The first feature of Table 3 that we draw attention to is that identical actual scores do not necessarily map onto the same mediated scores. For example, learners 6, 13, and 14 (learner numbers are assigned for convenience only and do not reflect any aspect of test performance) produced the same actual score of 16; however, learners 13 and 14 produced similar mediated scores – 29 and 28, respectively – while learner 6 produced a much higher mediated score, an indication that the learner responded more favorably to mediation. This learner also generated a high LPS (1.07) and was able to respond correctly to half (4 of 8) of the transfer items without mediation,[8] but learners 13 and 14 generated low LPSs and subsequently answered only two of the transfer items correctly. Next, consider the rather extreme example of learner 1, who produced a low actual score of 8 (out of a maximum of 60) but who scored 40 under mediation. Consequently, this student received an LPS in the high range and went on to score reasonably well on the transfer items: the same as learner 8 but higher than learners 13 and 14, who both earned higher actual scores. Finally, several learners (9–12) received higher actual scores than did learners 1 and 6 but did not improve as much under mediation and therefore produced mid-range LPSs and in fact scored lower on the transfer items than did 1 and 6. Learner 12 generated an LPS in the low range but still did better on the transfer items than did learners 9–11.

We caution against assuming a direct mapping across all scoring categories. Indeed, significant low to moderate correlation coefficients between LPSs and transfer scores

Table 4. Mean and standard deviation of transfer scores grouped by learning potential score (LPS) range.

LPS range	Chinese listening	Chinese reading	French listening
High	$\bar{x}$ = 18.27 (8.56) n = 30	$\bar{x}$ = 12.22 (5.72) n = 55	$\bar{x}$ = 12.75 (5.31) n = 16
Mid	$\bar{x}$ = 8.38 (4.54) n = 20	$\bar{x}$ = 8.30 (5.49) n = 20	$\bar{x}$ = 9.61 (4.56) n = 5
Low	$\bar{x}$ = 9.67 (3.17) n = 12	–	–

were obtained for the Chinese reading (r = .301, p < .01) and listening (r = .538, p < .01) tests but not for the French reading test (r = .337, ns). Even though the correlation was not significant for the French listening test, we nevertheless observed similar scoring profiles among the learners as on the Chinese tests. Overall the learners who took the French test produced higher actual scores than did the learners who completed either of the Chinese tests, and even though they improved under mediation the ceiling for improvement was lower.

3 Comparing LPSs and transfer items

We also compared the differences between transfer scores on each of the tests grouped by LPSs (high, mid, low). Table 4 presents the relevant means and standard deviations. In several cases a comparison was not possible because there scores failed to fall within a particular range. A Mann–Whitney U for the Chinese Listening test revealed significant differences between the high and mid ranges (z = 3.94; p < .05) as well as between the high and low ranges (z = 3.05; p < .05) but no significant difference between the mid and low ranges (z = –.84). For the Chinese Reading test a significant difference was found between the high and mid ranges only (z = 2.53; p < .05). There were no scores in the low range on this test. No significant differences were found between the high and mid range transfer scores on the French Listening test (z = 1.11; p > .05), and no scores fell into the low range for the test. For the time being, it appears from both the correlations and the statistical comparison of means that the LPS has promise as a predictor of learning, at least during the test as measured by actual scores on the transfer items. It remains to be determined the extent to which LPSs predict learning outcomes in classroom settings.

V Discussion and conclusion

In this article we have described the format and results of an initial attempt at developing Computerized Dynamic Assessment instruments administered via the Internet. The instruments focus on reading and listening comprehension. The major contribution of the approach is to be able to administer C-DA to large numbers of students simultaneously and to report the results in a quantitative format that is easily interpretable. In contrast

with other measures of L2 abilities, these tests provide learners and educators with information pertaining not only to their actual abilities but also to their ZPD. That is, the scores reflect both unmediated and mediated performance on the tests. A further innovation offered by the tests is the inclusion of the LPS, which takes the difference between mediated and unmediated performance as the basis for predicting how learners are likely respond to future instruction. Finally, the tests embrace the DA principle of integrating the diagnosis and promotion of abilities by including a system of graduated prompting for each item as well as transfer items at the conclusion of each test to create a potential for microgenesis.

As more learners take each of the tests, including the French reading and the Russian tests (not discussed in this article), we expect to be able to develop a clearer picture of the relationship between LPS and transfer scores. It is also important to investigate the relationship between LPS and classroom performance of learners. With regard to testing instruments, integration of different categories of transfer items that take account of different near, far, and very far transfer needs to be explored. With regard to the classroom, it will be necessary to investigate how learners with different LPSs react to instruction. In keeping with Feuerstein's original conceptualization of learning potential, we would expect that learners with a low LPS would require more extensive and intensive instruction than those with high scores.

Our argument is not that LPS should be used to grant or deny access to language learning opportunities (e.g. permitting only high LPS learners to continue language study). As explained earlier, we do not accept that learning potential is an immutable capacity that learners have in fixed amounts; rather, we concur with Vygotsky that the source of human abilities, including language, resides in our mediated participation in activities with others. Moreover, we note that Vygotsky himself was preoccupied with issues of access and fairness in education and understood the provision of mediation attuned to individuals' needs as a means to promote success for all (for discussion, see Poehner, 2011). Thus, we see LPS as potentially quite relevant to placement decisions whereby learners receive instruction that is complementary not to their level of actual development but to their level of proximal development.

We also want to emphasize that we are not proposing the current test design as the only way to implement C-DA. As noted above, other models of DA have been proposed for computerized administration, and these may also serve as a point of departure for L2 DA researchers. We are in the beginning stages of what promises to be a lengthy and challenging process. For example, there is no reason to limit C-DA to the multiple-choice testing format. In fact, we are currently developing a language comprehension test that uses a cloze format with more open-ended types of mediation. We also see assessments that include language production tasks as an especially important, though challenging, context for C-DA. The CAST test developed by the Title VI National Language Resource Center at San Diego State University is an on-line speaking proficiency assessment that, in our view, presents an excellent opportunity to explore the feasibility of integrating mediation into its format.

Finally, we would like to propose that the integration of DA principles in a formal testing situation may best be viewed as a component of a broader program for development-oriented pedagogy. That is, C-DA and classroom-based DA offer a two-pronged approach

to diagnosing learner abilities, placing them in an instructional setting where teaching will be attuned to their ZPD, and then continuing to both challenge and support the learners in day-to-day classroom activities. Coordinating these formal and informal applications of DA offers a powerful approach to fully realizing the implications of Vygotsky's conceptualization of the ZPD in L2 education.

Notes

1. To date we have not administered the Russian tests to a sufficient number of learners to permit an appropriate analysis of results. In addition, during administration of the French tests, the server on which the tests were housed crashed and we lost a portion of the results, including learner performance on the reading test.
2. Kozulin and Garb (2002) identify a potential problem with regard to the reliability of the LPS stemming from the possibility that assessors with different meditational styles 'might reveal a somewhat different pattern of learning abilities in the same group of students' (p. 122). While this is indeed an important matter – especially for DA administered in an interactionist format – we do not believe it is an issue of direct concern for the approach we have adopted, given that mediation has been standardized. On the other hand, it is indeed possible that reliability could still be a concern since there are different ways of standardizing mediation, some of which could be more or less effective than others. We therefore agree with Kozulin and Garb when they point to a need for studies on LPS reliability (p. 123).
3. This project was funded by a grant from the United States Department of Education (Grant Award P017A080071). However, the contents of this article do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the Federal Government. Although we served as official co-PIs for the project, we want to acknowledge the indispensible contributions of co-researchers Rimma Ableeva, Katya Arshavskaya, Xiaofei Lu, Mei-Hsing Tsai, Adam van Compernolle, and Jie Zhang. We also acknowledge the contribution of Arlo Bensinger, who developed the computer platform for the administration and scoring of the tests.
4. Research on L2 comprehension assessments suggests that learners may engage in similar processes regardless of whether a multiple-choice or open-ended format is followed, provided that parallel texts and questions are employed. In a meta-analysis of L2 comprehension studies In'nami and Koizumi (2009), found positive correlations as high as .95 between these two assessment formats.
5. Allowing learners multiple attempts at each test item decreases the degrees of freedom they confront and therefore increases the possibility of guessing the correct answer. The reduction in the degrees of freedom after failed attempts means subsequent mediation might be doing less 'work' with regard to leading learners through a comprehension process. Nevertheless, based on our initial experience with the tests, we are convinced that the diagnosis of learner abilities is expanded in C-DA by determining not simply whether a learner answered an item correctly but by the number of mediating prompts required to do so. Furthermore, analysing learner performance on the transfer items included in the test can help to determine learners' use of guessing strategy. For example, if a learner guessed at each item either initially or in subsequent attempts, we would not expect that individual to learn much during the test. Of course, we cannot completely rule out guessing behavior, which is an issue for any multiple-choice test.
6. Inclusion of mediation in the testing process means that the abilities underlying learner performance during the test may in fact be a moving target. That is, learner abilities may change during the course of the assessment. In C-DA, microgenesis early in the test could affect an individual's performance on subsequent items, altering both their actual and mediated scores.

We do not view this as a problem in C-DA but as a favorable outcome. Moreover, avoiding this possibility would likely entail two administrations of the test, one in which no mediation is offered and a second that includes mediation. To be sure, this approach would be more time intensive, but researchers concerned with establishing a baseline of learner unmediated performance prior to DA might wish to explore this format.

7. Learner performance on each of the transfer items is reported in the qualitative profile.

References

Ableeva, R. (2010). Dynamic assessment of listening comprehension in second language learning. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA, USA.

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.

Brown, A., & Ferrara, R.A. (1985). Diagnosing zones of proximal development. In J.V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 273–305). Cambridge: Cambridge University Press.

Budoff, M. (1987). The validity of learning potential assessment. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 53–81). New York: Guilford Press.

Campione, J.C., Brown, A.L., Ferrera, R.A., & Bryant, N.R. (1984). The zone of proximal development: Implications for individual differences and learning. In B. Rogoff, & J.V. Wertsch (Eds.), *Children's learning in the 'zone of proximal development'* (pp. 77–91). San Francisco, CA: Jossy-Bass.

Feuerstein, R., Rand, Y., & Hoffman, M.B. (1979). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore, MD: University Park Press.

Feuerstein, R., Rand, Y., & Rynders, J.E. (1988). *Don't accept me as I am: Helping retarded performers excel*. New York: Plenum.

Glutting, J.J., & McDermott, P.A. (1990). Principles and problems in learning potential. In C.R. Reynolds & R.W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 296–347). New York: Guilford.

Guthke, J., & J.F. Beckmann. (2000). The learning test concept and its applications in practice. In C. S. Lidz & J.G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp.17–69). Amsterdam: Elsevier.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, *26*, 219–244.

Jacobs, E.L. (2001). The effects of adding dynamic assessment components to a computerized preschool language screening test. *Communication Disorders Quarterly*, *22*, 217–226.

Kozulin, A., & Garb, E. (2002) Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International*, *23*, 112–127.

Lantolf, J.P., & Poehner, M.E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, *1*, 49–74.

Lantolf, J.P., & Poehner, M.E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for L2 development. *Language Teaching Research*, *15*, 11–33.

Leont'ev, A.N. (1981). *Problems of the development of the mind*. Moscow: Progress.

Lidz, C. (1991). *Practioner's guide to dynamic assessment*. New York: Guilford Press.

Poehner, M.E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, *91*, 323–340.

Poehner, M.E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing.

Poehner, M.E. (2011). Dynamic assessment: Fairness through the prism of mediation. *Assessment in Education: Principles, Policy and Practice*, *18*, 99–112.

Poehner, M.E., & van Compernolle, R.A. (2011). Frames of interaction in dynamic assessment: Developmental diagnoses of second language learning. *Assessment in Education: Principles, Policy and Practice*, *18*, 183–198.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Publishing.

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

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

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L.S. (1987). Thinking and speech. In R.W. Rieber, & A.S. Carton (Eds.), *The collected works of L.S. Vygotsky: Volume 1: Problems of general psychology* (pp. 39–285). New York: Plenum.

Wertsch, J.V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

Appendix. Sample item from the French reading test illustrating mediation.