

Dynamic assessment in EFL classrooms: Assessing listening comprehension in three proficiency levels

Hashemi Shahraki, Sara ✉

English Department, University of Isfahan, Iran (sara_m_hashemi@yahoo.com)

Ketabi, Saeed

English Department, University of Isfahan, Iran (ketabi@fgn.ui.ac.ir)

Barati, Hossein

English Department, University of Isfahan, Iran (h.barati@gmail.com)



ISSN: 2243-7703
Online ISSN: 2243-7711

OPEN ACCESS

Received: 7 December 2014

Revised: 4 January 2015

Accepted: 10 January 2015

Available Online: 12 January 2015

DOI: 10.5861/ijrse.2015.1024

Abstract

The use of dynamic assessment (DA), grounded in Vygotskian Sociocultural theory, in classrooms is believed to have the potential to provide a situation for creating a group of learners' Zone of Proximal Development (ZPD) (Poehener, 2009). The present study explored the implementation of DA in English as Foreign Language (EFL) classrooms on groups of learners at different proficiency levels in the context of listening comprehension. 146 Iranian EFL learners at three proficiency levels were selected for this study (experimental groups=71 and control groups=75). A multi-assessment procedure in the format of dynamic and non-dynamic pretest-enrichment phase-dynamic and non-dynamic posttest was conducted. During the nine-week group dynamic assessment procedures, mediational strategies were only given to the experimental groups. The quantitative data analysis revealed that through mediated interactions within the group's ZPD, group dynamic assessment is able to determine the learners' developed abilities in listening comprehension while at the same time support the development of individual learners in this skill. Moreover, it was found that the level of proficiency of the learners did not have a significant effect on learners' gains from group dynamic assessment procedures. These findings can have implications for all classroom teachers that the use of DA in classroom setting cannot only be beneficial to them but also be considered as a strategic learning and assessment method that can meet both the learners' and teacher's needs.

Keywords: classroom dynamic assessment; sociocultural theory; zone of proximal development; listening comprehension ability; proficiency level

Dynamic assessment in EFL classrooms: Assessing listening comprehension in three proficiency levels

1. Introduction

In most of the Second Language (L2) English classes (whether in schools, universities, and private institutes) teachers have often witnessed students who perform relatively well in class but poorly on classroom or final tests. When only the independent performance of a learner on a test is used to evaluate and interpret his/her knowledge of a specific skill or feature, according to Vygotsky (1998) much will remain unclear about the ability of that learner. In order to reflect the dynamism and ongoing nature of the learning and development of learners, assessment processes are needed which not only tap into the fully formed capacities of the learners but also into the emerging ones (Lidz, 1991; Sternberg & Grigorenko, 2002). Dynamic Assessment (DA) rooted in Vygotskian Socio-Cultural Theory (SCT) of mind prescribes mediated teacher-examinee dialog during assessment procedure as examinees face problems (Sternberg & Grigorenko, 2002). This type of assessment assesses the zone of proximal development (ZPD), mostly described as the difference between an individual's independent functioning and the level of performance she/he may reach in co-operation with others (Vygotsky, 1978; Lantolf & Thorne, 2006). According to Vygotsky's innovative insight, the learning of a learner most effectively occurs within her/his ZPD under appropriate mediation. Therefore, by offering appropriate mediation, a teacher can diagnose the fully developed abilities of a learner revealed through her/his independent performance, as well as abilities that are in the process of forming (Sternberg & Grigorenko 2002; Haywood & Lidz, 2007).

The presence of a rich research literature on DA in psychological and educational research since its introduction in the 1950s reflects it to be an important tool in gathering quality information about learning (e.g. Feuerstein, Rand, Hoffman, & Miller, 1980; Budoff, 1987). However, it is less than 20 years that language educators have begun to examine the pedagogical applications of DA on L2 reading (e.g. Abdolrezapour, Tavakoli, & Ketabi, 2014; Ajideh & Nouradad, 2012; Kozulin & Garb, 2002) on L2 writing (e.g. Aljafreh & Lantolf, 1994; Alavi & Taghizadeh, 2014; Panahi, Birjandi, & Azabdaftari, 2013) and on L2 Speaking (e.g. Anton, 2009; Poehner, 2005). The L2 skill of listening comprehension has received the smallest share of L2 DA research (Ableeva, 2010; Hidri, 2014). Ableeva (2010) reports on a DA project carried out on university level intermediate learners of French to diagnose and promote listening ability. Ableeva's study followed a multi-assessment procedure in the format of non-dynamic and dynamic pretest - enrichment program - non-dynamic and dynamic posttest - non-dynamic and dynamic transfer. In her study, Ableeva observed that a lack of lexical knowledge of the L2, problems stemming from phonology, limited knowledge of the L2 culture, and issues with discourse level grammar caused poor performance in L2 listening comprehension. She also witnessed that the students were able to transfer their knowledge to some though not all of the new more complex passages, i.e. there were limits to what individuals are capable of doing under mediation. In another study on L2 listening comprehension within SCT, Hidri (2014) compared static and dynamic testing of L2 listening comprehension at university level. She found in her study that through DA better insights into learners' cognitive and meta-cognitive processes could be gained than through the traditional static assessment.

One reason that L2 listening comprehension is the least researched skill could be its complexity. According to Vandergrift (2007) this skill demands the listener to use a wide variety of knowledge sources to interpret incoming data rapidly. Another reason based on Buck's (2001) beliefs is that oral texts exist in real time and need to be processed quickly; when the text is over, only a mental representation remains in the mind of learners. He claims that since these processes are covert, listening is a difficult skill to research. In order to probe deeper in the processes and factors that influence successful L2 listening comprehension and to have a diagnostic assessment of this language skill, assessing this skill in learners' through DA attains great significance.

Browsing the literature of L2 DA research (e.g. Ableeva, 2010; Anton, 2009), it is observed that most

researchers have had a strong preference for one-to-one (teacher–learner) DA procedures outside the classroom. This could be due to the fact that this type of interaction in classroom setting will be time-consuming and tiresome. However, SCT practitioners agree that it is possible for the mediator to simultaneously provide group of learners with mediating prompts and comments in order to co-construct several ZPDs and also push the entire group's ZPD forward (Poehner & Lantolf, 2005; Poehner, 2009). Poehner (2009) introduces a new framework known as group dynamic assessment (G-DA) to investigate the use of DA with groups of classroom L2 learners rather than individuals. Moreover, Vygotsky's definition of ZPD as the "optimum time for teaching both the group and individual ZPD" (Vygotsky, 1998, p. 204) seems to be a confirmation of the assumption that DA in a group context can be used to promote a number of ZPDs simultaneously. Poehner (2009) clarifies the notion of G-DA and maintains that in this type of DA the same rubrics for offering mediation as in individualized interactions are applied. However, in G-DA the focus is not on one individual but on an entire class. Nevertheless, a major challenge facing the implementation of DA in the classroom context is that how it should be applied in a place where the teacher interacts with not a single ZPD but a group of ZPDs (Poehner, 2009). This could be known as one reason that the classroom context has received a small share of DA research. Some studies (e.g., Lantolf and Poehner, 2011; Davin, 2013) have recently tried to implement DA procedures with groups of classroom L2 learners, but this matter is still in its infancy.

Gibbons' (2003) study is known to be an early attempt on co-constructing a group's ZPD by mediation; however, her study was not overtly framed as a DA. She found that the ways students and teacher used to co-construct meaning in a shared experience had a great impact on students' progress. This in turn could stretch their ZPD to more complex domains. Another study concerning the implementation of DA in classroom context is reported in Lantolf and Poehner (2011). They analyzed the efforts of an elementary school L2 Spanish teacher aiming to improve oral proficiency of her learners through G-DA. They observed the process of development was gradual for some and abrupt for others. Moreover, they witnessed that it was sometimes possible to push the development of the group of students forward with the co-construction of a ZPD with an individual. Davin's (2013) study concerns the integration of DA and instructional conversation (IC) within classroom setting. In her study, an L2 Spanish teacher made use of this integration to teach and assess a grammatical structure to elementary-level learners. The findings of her study revealed that G-DA and the IC are compatible and have the potential to promote development and improve assessment in the language classroom.

Another study done in G-DA is Hashemi Shahraki, Ketabi and Barati (2015). In their study, they focused on the effects of group DA on learners' pragmatic knowledge of conversational implicatures in the context of listening. Their study was conducted on 50 English as Foreign Language (EFL) learners at the intermediate level. Through the analysis of the mediated interactions between the learners and the mediator, they detected nine mediational strategies effective for the development of listening abilities which ranged in their degree of explicitness. Moreover, Hashemi Shahraki and associates found that the group DA procedures implemented in their study led to the improvement of learners' listening ability especially their pragmatic understanding of conversational implicatures; however, all the learners did not improve in the same amount. Based on the literature survey done on L2 DA, it appears that the L2 listening comprehension of language learners has not yet been adequately analyzed through dynamic assessment especially through G-DA in the context of classroom.

In order to have a deeper understanding of the problems EFL learners face in listening comprehension and to gain some insight on the nature of this skill, this study implemented DA to groups of learners at different proficiency levels in EFL classrooms. In other words, the present study aimed to determine whether G-DA can promote the development of L2 listening comprehension ability in elementary, lower intermediate, and upper intermediate language learners. Moreover, it attempted to find out which proficiency level took the greatest advantage of the mediation received. The rationale for using proficiency level as one variable in this study was to examine Vygotsky's (1978) claim about the actual proficiency level of individuals in the context of L2 learning. Vygotsky (1978) pointed out that the size of a child's ZPD can be more revealing about his/her development in learning a feature than that child's actual proficiency level; however, his claim was about child learning and not adult learning, and most importantly, it was not about learning an L2.

An important point to consider is that the current study reports on a part of a larger, developmental study examining the applicability of G-DA to L2 listening comprehension with the aim of organizing classroom interactions in terms of the learners' ZPD. Although the full study employed both qualitative and quantitative data analyses, the present paper focuses on the latter. This study followed a quasi-experimental mixed method research design and aligned itself with the Socio-cultural Theory of Mind using a concurrent interactionist G-DA approach. In order to deeply examine the problem at issue, the following research question was addressed:

- Does G-DA enhance the development of listening comprehension ability of EFL elementary, lower intermediate, and upper intermediate learners? If so, which proficiency level benefits the most?

Based on the previous research and the aforementioned question, the following directional hypothesis is formulated:

Hypothesis: (a) G-DA would enhance the development of listening comprehension ability of EFL elementary, lower intermediate, and upper intermediate learners, and (b) all proficiency levels would benefit the same.

2. Materials and Method

2.1 Participants

The statistical population under study included EFL learners who enrolled at a language institute in Iran. In order to select participants, the non-random purposive sampling method (Mackey & Gass, 2005) was applied. To this end, six intact classes (two elementary-level, two lower intermediate-level, and two upper intermediate-level) were selected. The 175 learners of all the above-mentioned classes were female Persian native speakers in the age range of 14-18. The 60 learners of the two elementary level classes had taken English courses for approximately one year, the 58 learners of the two intermediate level classes had taken English courses for approximately two years, and the 57 learners of the two high-intermediate level classes had taken English courses for approximately three years in this institute. None had exposure to English apart from attending this English course. Initially, the Oxford's Quick Placement Test (QPT) was given to the learners in order to select those whose levels are, based on the placement guide of this test, at the elementary, lower intermediate, and upper intermediate levels. Based on the results, 56 learners whose scores fell within the range of elementary i.e. 16-23, 52 learners whose scores fell within the range of lower intermediate i.e. 24-30, and 53 learners whose scores fell within the range of upper intermediate i.e. 40-47 were selected to participate in the study. Then, from each level one class was considered as the experimental group and the other as the control group. A number of the participants were excluded from the study since they did not attend all sessions during the nine-week G-DA procedure. This left 22, 25, and 24 participants respectively in the elementary, lower intermediate, and upper intermediate experimental groups, and 25, 24, and 26 participants respectively in the elementary, lower intermediate, and upper intermediate control groups.

2.2 Instruments

The following materials were used:

Quick Placement Test (QPT) - In order to determine the exact level of the participants in this study, the Quick Placement Test (QPT) developed by Oxford University Press and Cambridge ESOL (2001) was used. Based on the research notes of ESOL Examinations, this test serves as a flexible test of English language proficiency. The test has a computer-based and a paper-and pen version. For the purpose of this study, the paper-and-pen version was used which has 60 multiple-choice format questions. This version consists of two parts: Part one (1-40) deals with simple grammar and vocabulary items which all candidates should answer. Part two (40-60) is concerned with a bit more difficult multiple choice items and cloze test, which is for higher ability

learners (Geranpayeh, 2003). All of the test items have gone through Cambridge quality control procedures (Beeston, 2000). In this study, the learners of the elementary and lower intermediate levels only took the first part of the test. However, the upper intermediate learners took part two as well since they answered more than 35 items out of the 40 items of part one correctly. The specified time allotted for the completion of this test is 30 minutes. The reliability (internal consistency) value of this test is 0.9 and has been validated by more than 6000 students in 20 countries.

The pretest and posttest - The pretest and posttest in this study was the Interactions/Mosaic Listening Placement Test (McGraw-Hill ESL/ ELT College). This test is developed by the McGraw-Hill ESL/ ELT College as a placement test for the series of the books of Interactions/Mosaic Listening/Speaking (2008, 6th Ed., McGraw-Hill Press) to guide teachers and course managers to choose the appropriate-level book for their learners. This test assesses the ability of L2 learners to listen for a wide variety of real-life purposes. In other words, the test might measure the ability of the test takers to listen for the gist of a whole extract, for a particular detail or the speaker's opinion. The reason for choosing this test was to guide the mediator (the researcher) to choose the appropriate-level listening test tasks from the series of Interactions/Mosaic Listening/Speaking books for the participants of the study for the enrichment phase. Another justification for choosing this test was that it was quite a new test in this institute; thus, surely it was not available to the students before the test administration. Consequently, the results gained from this test are quite reliable. There are a total of fifty questions in multiple-choice format. This test takes about 40 minutes to administer. To estimate the quality criteria of this test such as its reliability and validity, this test was piloted. Therefore, 20 elementary, 20 lower intermediate, and 20 upper intermediate learners, whose general characteristics were exactly identical to the participants in the study, were selected to participate in the piloting procedure. The reliability of this test was computed according to Cronbach's alpha formula which was 0.93. The listening section of FCE test was used as a standard criterion to estimate the criterion-related validity of the test which was 0.89.

Tests of the enrichment phase - 21 Listening Tests (LTs), i.e. 7 LTs for each proficiency level, were used for the seven-session enrichment phase in this study. The criterion for the selection of LTs was the scores of the participants on the Interactions/Mosaic Listening Placement test. The scores of the elementary level learners were in the range of 2-13; thus, based on the test guidelines the Interactions Access book from the series of Interactions/Mosaic Listening/Speaking books (2008, 6th Ed., McGraw-Hill Press) was appropriate for them. According to the test guidelines, Interactions 2 and Mosaic 1 books were determined to be appropriate for the lower intermediate and upper intermediate levels respectively since the scores of the lower intermediate learners were in the range of 24-31, and those of the upper intermediate learners were in the range of 33-42. The information gained through pre-testing the participants guided the selection of the appropriate content for the LTs. To this end, focusing on the grammatical constructions and lexical knowledge which were problematic for the learners of each proficiency level, 21 LTs were selected from the appropriate-level books. The comprehension questions of the LTs are of multiple-choice type. The items of the tests are based upon factual information in the text and cannot be answered correctly without having listened to and understood the relevant part of the text. These items test the participants' ability to listen for the gist of a whole extract, for a particular detail or the speaker's opinion, and making inferences about what they hear.

In order to make the LTs reliable and valid for the purpose of the study and to find out the appropriate time that should be allotted for their completion, the 21 LTs were piloted. To this end, 20 elementary, 20 lower intermediate, and 20 upper intermediate learners, having the same general characteristics as the participants in the study, were selected. In the piloting procedure, the 20 elementary learners took the 7 LTs of the elementary level. The 7 LTs of the lower intermediate and the 7 LTs of the upper intermediate level were given to the 20 lower intermediate and 20 upper intermediate learners respectively. Using Cronbach's alpha formula the reliability values of the 21 LTs were calculated. The 7 LTs of the elementary level yielded the reliability of 0.85, 0.92, 0.77, 0.79, 0.89, 0.87, and 0.78 respectively. The estimated reliability values for the 7 LTs of the lower intermediate level are 0.90, 0.94, 0.84, 0.79, 0.89, 0.93, and 0.86 respectively. And the piloting results demonstrated that 7 LTs of the upper intermediate level had the reliability values of 0.73, 0.80, 0.84, 0.87, 0.85,

0.92, and 0.88 respectively.

To assure the content validity of the 21 LTs selected for the purpose of this study, these tests were given to a group of 5 testing experts at University of Isfahan to be reviewed and judged. They qualified the content of the tests of each level to be appropriate for measuring the listening ability of the students at that relevant level; however, they suggested some modifications to be made in the items and choices of the tests. Based on the findings of piloting the 21 LTs, depending on the length of the listening text in each LT, 10 - 15 minutes' time was deemed to be required for the completion of each listening test task.

2.3 Procedure

The multi-assessment GDA procedures in this study were conducted in three group settings (namely, elementary, lower intermediate, and upper intermediate learners) for a time span of 9 weeks. Following Ableeva (2010) and Poehner (2008) the G-DA procedures were administered in the non-dynamic and dynamic pretest - enrichment phase - non-dynamic and dynamic posttest format as shown below:

- Week 1 → Non-dynamic and dynamic pretest
- Weeks 2-8 → Enrichment phase (7 LTs for each proficiency level: Non-dynamic and dynamic assessments)
- Week 9 → Non-dynamic and dynamic posttest

In week 1, the pretest was given to the participants of the experimental group and control group in the elementary, lower intermediate and upper intermediate-level classes. After administering the pretest in the non-dynamic format to the experimental and control groups of all proficiency levels, the mediator (the teacher) implemented this test dynamically only to the experimental groups of the three proficiency levels. The dynamic administration of the pretest was for the purpose of gaining information about learners' problems in L2 listening comprehension to guide the mediator's efforts to select LTs for the enrichment phase and thereby mediate the learners.

The DA of the learners in the experimental groups of all proficiency levels of this study in the classroom context was conducted as explained below. After the participants answered the test individually, the test sheets were collected. Then the listening texts of the test were replayed portion by portion for the class to provide their recalls. Upon the students' failure to recall the content of the sentence, the mediator intervened and provided mediation. Mediation is defined here as assisting prompts intended to promote development that are offered from implicit to explicit. However, the mediation offered in this study was not pre-specified and emerged from the mediator's ongoing collaborations with learners, i.e. using the interactionist approach. This type of flexible and fine-tuned mediation is assumed to be more effective in bringing about learning than standardized mediation (Vygotsky, 1998). Moreover, the concurrent approach to group DA was used to offer mediation. In other words, to mediate the learners, when a learner faced a problem, the mediator provided mediation in response to her problem. However, the interaction shifted rapidly between primary and secondary interactants since one learner's question or comment set the stage for another's contribution. In this way, the mediator dialogued with the entire group. Prior to the dynamic implementation of any test in this study, the learners were told that their active presence and contributions in the class were essential to their learning. Moreover, learners were warned that being silent and inactive in class would be interpreted as an inability in their listening comprehension skill. This was done to assure their active participation in the G-DA interactions.

During weeks 2-8, the enrichment phase took place over a seven-week period which included seven sessions held once a week. The experimental groups of all proficiency levels first answered an LT individually in each session. Then in that same session, the learners of these groups took that LT dynamically through the group DA procedure explained for the dynamic administration of the pretest. The participants of the control groups of all

proficiency levels, during weeks 2-8, did not receive any mediation. However, during this period, each week they took one LT in the non-dynamic format. Thus, they answered the 7 LTs of the enrichment program independently in order to eliminate the practice effects of the test materials (Mackey & Gass, 2005).

In week 10, the posttest was administered to the experimental and control groups of all proficiency levels non-dynamically. Nevertheless, in all proficiency levels, the experimental groups also took this test in the dynamic format. The test which was used as the pretest served as the posttest as well. The rationale of using the same test was to trace the effects of the enrichment phase on the students' L2 listening comprehension.

During the G-DA program in this study, each dynamic administration of the tests was audio and video-recorded. The teacher-learner interactions taken place in all tests were transcribed and coded in order to have access to all the mediating moves.

3. Results

Quantitative data analysis was used to answer the two parts of the research question which guided this study.

3.1 The First part of the research question

The aim of the first part of the research question of this study was to investigate whether G-DA enhanced the development of L2 listening ability in the three proficiency levels of elementary, lower intermediate and upper intermediate learners. To measure the effects of DA in the classroom context on the development of EFL learners' listening comprehension, all participants took the pretest and after the enrichment phase took the posttest. Table 1 displays the mean scores of the participants of this study on the pretest and posttest along with their standard deviations and the number of participants in each group.

Table 1

Descriptive statistics of the pretest and posttest of the three proficiency levels

Test Type	Experimental group			Control group		
	Elementary <i>N</i> =22	Lower Int. <i>N</i> =25	Upper Int. <i>N</i> =24	Elementary <i>N</i> =25	Lower Int. <i>N</i> =24	Upper Int. <i>N</i> =26
	<i>Mean</i>	<i>Mean</i>	<i>Mean</i>	<i>Mean</i>	<i>Mean</i>	<i>Mean</i>
	<i>SD</i>	<i>SD</i>	<i>SD</i>	<i>SD</i>	<i>SD</i>	<i>SD</i>
Pretest	7.63	27.72	36.00	7.56	27.83	36.08
	2.04	2.42	1.91	2.55	2.33	1.92
Posttest	12.86	34.04	41.58	7.60	27.92	36.46
	3.35	3.10	3.47	2.66	2.20	1.70

Note. pretest and posttest maximum possible score=50

In order to determine that both the experimental and control groups of all proficiency levels, i.e. elementary, lower intermediate, and upper intermediate were at the same level of listening ability at the outset of this study an independent samples t-test was run between the pretest scores of them.

Table 2

Comparison tests between the experimental and control groups of all proficiency levels on the pretest

Level	Comparison	<i>t</i>	<i>df</i>	<i>p</i>
Elementary	pretest experimental × pretest control	-0.107	45	0.916
Lower intermediate	pretest experimental × pretest control	0.153	47	0.879
Upper intermediate	pretest experimental × pretest control	0.142	48	0.888

Table 2 depicts that at the outset of the G-DA procedures there were no significant difference in the listening comprehension ability of the experimental and control groups of the three proficiency levels: elementary level [$t(45)=-0.107, p>0.05$], lower intermediate [$t(47)=0.153, p>0.05$], and upper intermediate [$t(48)=0.142, p>0.05$].

In order to detect the effect of G-DA procedures on the L2 listening comprehension of the learners, first the variables of the study had to be clarified. This study included three categorical independent variables: 1) study group with two levels of control and experimental, 2) proficiency level with three levels of elementary, lower intermediate and upper intermediate, and 3) time with two levels (time1: pretest and time2: posttest). The level of the learners (elementary vs. lower intermediate vs. upper intermediate) and study group (experimental vs. control) were between-subject factors and time was a within-subject factor in this study. The dependent factor was listening comprehension (the scores of the learners at each time interval). For this a 2×3×2 mixed between-within subjects ANOVA was applied for making comparisons between the six study groups at the beginning of the study and after the enrichment phase. The results of the multivariate tests are shown in Table 3.

Table 3

Multivariate tests on the pretest and posttest of the three proficiency levels

Effect (Wilks' Lambda)	Value	F	Hypothesis df	Error df	p	Partial eta squared
Time	0.349	261.390	1	140.00	0.000	0.651
Time*Study group	0.376	232.413	1	140.00	0.000	0.651
Time*Level	0.988	0.856	2	140.00	0.427	-
Time*Study group*Level	0.989	0.779	2	140.00	0.461	-

Note. Study group=Experimental or Control groups

As it is demonstrated in Table 2, no interaction effect was observed between time and level [Wilks' Lambda=0.988, $F(2, 140)=0.856$, $p>0.05$] and among the categorical variables of time, study group, and level [Wilks' Lambda=0.989, $F(2, 140)=0.799$, $p>0.05$]. However, an interaction effect was observed for time and study group (experimental or control group) [Wilks' Lambda=0.376, $F(1, 140)=232.413$, $p<0.0005$]. Therefore, follow-up tests were conducted to explore this relationship further. To do this, the results split based on the independent variable of study group had to be analyzed. Thus, a one-way repeated measures ANOVA was run to explore the effect of time separately on experimental and control groups. The results are provided in Table 4.

Table 4

Multivariate tests on the pretest and posttest organized by experimental groups and control groups

Effect (Wilks' Lambda)	Value	F	Hypothesis df	Error df	p	Partial eta squared
Time on experimental groups	0.097	650.984	1	70.00	0.000	0.903
Time on control groups	0.995	0.373	1	74.00	0.543	0.005

The results shows that there was a significant effect for time on the experimental groups [Wilks' Lambda=0.097, $F(1,70)=650.984$, $p<0.0005$] with the partial eta squared of 0.903 which means it had a large effect. In other words, the G-DA procedures resulted in the improvement of the L2 listening comprehension ability in the participants of the experimental groups. However, time did not have a significant effect on the control groups [Wilks' Lambda=0.995, $F(1,74)=0.373$, $p>0.05$]. This observation also attests the effectiveness of G-DA procedures in enhancing L2 listening comprehension ability in learners since the control group only took the tests in the non-dynamic format in this study.

The descriptive data obtained from the pretest and posttest of the experimental groups of all proficiency levels presented in Table 1 shows that although the learners of this group received a significant gain from the G-DA intervention, the standard deviation of the scores achieved in the posttest exhibited considerable variation [Elementary: pretest $M=7.36$ $SD=2.34$, posttest $M=12.86$ $SD=3.35$; Lower intermediate: pretest $M=27.72$ $SD=2.02$, posttest $M=34.04$ $SD=3.10$; and Upper intermediate: pretest $M=36.00$ $SD=1.91$, posttest $M=12.86$ $SD=3.47$]. To determine whether all learners are improving in the same amount first correlation tests were run between the pretest and posttest of the experimental groups of the three proficiency levels. Then the percentage that each individual improved on the posttest was analyzed. Table 5 displays the Pearson product moment correlation coefficients between the pretest and posttest scores of the experimental groups of all proficiency

levels.

Table 5

Pearson Product-Moment Correlations between the pretest and posttest of the experimental groups of all proficiency levels

Level	Comparison	<i>r</i>	<i>N</i>	<i>p</i>
Elementary	pretest experimental × posttest experimental	0.764	22	0.00
Lower intermediate	pretest experimental × posttest experimental	0.831	25	0.00
Upper intermediate	pretest experimental × posttest experimental	0.830	24	0.00

As shown in Table 5, the pretest and posttest scores in all experimental groups are highly correlated [Elementary level: $r=0.76$, $N=24$, $p<0.0005$; Lower intermediate level: $r=0.83$, $N=25$, $p<0.0005$; and Upper intermediate level: $r=0.83$, $N=22$, $p<0.0005$]. This finding indicates that many of the learners indeed benefited from mediation; however, the question still remains that what has caused the standard deviation values of the pretests to be at a high level. In order to have a clearer picture of the learners' improvement, Figures 1, 2 and 3 illustrate learners' improvement in L2 listening comprehension by their pretest scores.

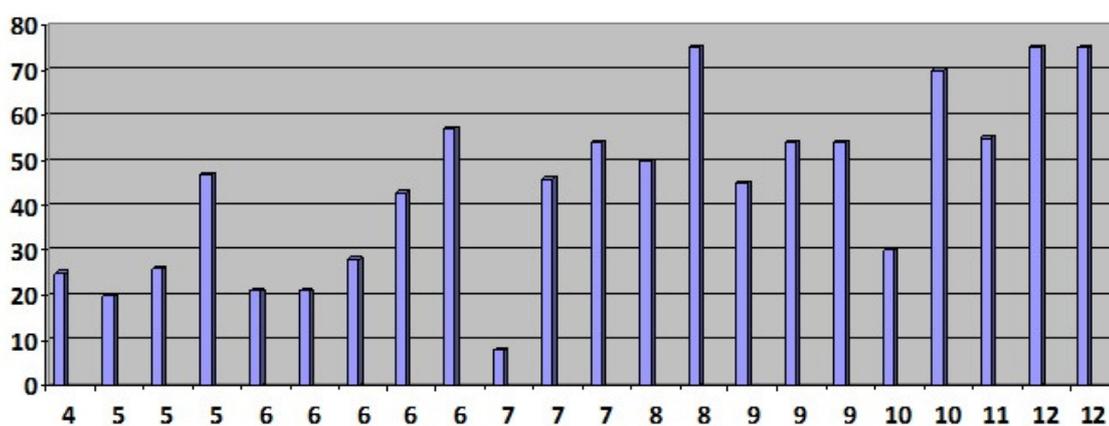


Figure 1. Development of the experimental group of Elementary level by their pretest scores

Note: X-axis=pretest scores; Y-axis= Percentage of development

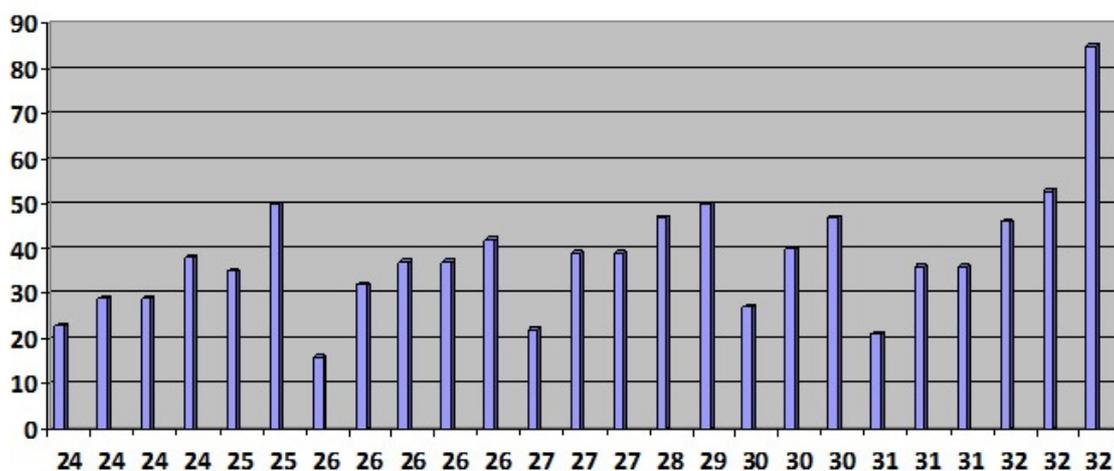


Figure 2. Development of the experimental group of Lower intermediate level by their pretest scores

Note: X-axis=pretest scores; Y-axis= Percentage of development

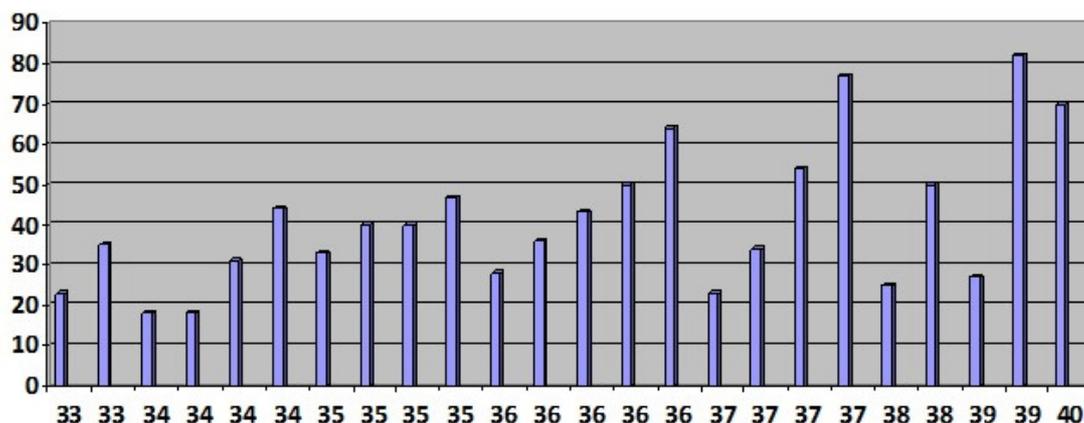


Figure 3. Development of the experimental group of Upper intermediate level by their pretest scores

Note: X-axis=pretest scores; Y-axis= Percentage of development

As it is evident in these figures, while some learners shared the same score on the pretest, their respective performances differed significantly on the posttest, with some showing slight improvement (e.g., 8% in the elementary level; 16% in the lower intermediate; and 23% in the upper intermediate) following the enrichment phase and others manifesting a marked increase in their posttest scores (e.g., 75% in the elementary level; 42% in the lower intermediate; and 77% in the upper intermediate). This can be an indication that although the learners of each level had almost the same actual proficiency level, it was possible that their development to be different.

In addition, in order to determine the effectiveness of the G-DA procedures on the L2 listening ability of the learners, the non-dynamic performance of the participants on LTs was also analyzed. As mentioned before, both the experimental and control groups of all proficiency levels took 7 LTs related to their proficiency level during the enrichment phase. Table 6 depicts the mean scores of the 21 LTs taken in this study along with their standard deviations.

Table 6

Descriptive statistics of the LTs of the three proficiency levels

Test Type	Elementary		Lower Intermediate		Upper Intermediate	
	Experimental N=22	Control N=25	Experimental N=25	Control N=24	Experimental N=24	Control N=26
	Mean	Mean	Mean	Mean	Mean	Mean
	SD	SD	SD	SD	SD	SD
LT 1	2.45	1.88	2.52	1.79	2.17	1.54
	0.80	0.78	0.92	0.72	0.96	0.76
LT 2	2.72	1.93	2.60	1.80	2.29	1.47
	0.96	0.92	0.97	0.97	0.96	0.44
LT 3	3.01	1.95	2.80	1.87	2.71	1.20
	0.73	1.02	0.90	0.96	0.80	0.56
LT 4	3.18	1.83	2.99	1.80	2.81	1.41
	0.83	0.96	0.93	0.97	0.68	0.46
LT 5	3.13	1.90	3.15	1.82	2.86	1.25
	0.84	0.81	0.82	0.82	0.93	0.61
LT 6	3.22	1.88	3.40	1.83	3.00	1.31
	0.81	0.83	0.91	0.81	0.88	0.67
LT 7	3.81	1.72	3.88	1.75	3.50	1.31
	0.79	0.67	0.83	0.74	1.10	0.61

Note. LT maximum possible score=5; The participants of each proficiency level took the LTs that were relevant to their own proficiency level

To compare the performance of the participants on the LTs taken in each session of the enrichment phase, t-tests were run between the non-dynamic performance of the experimental and control groups of each level on their relevant LTs. The comparison of their performance, presented in Table 7, also attests the benefit of the enrichment phase in promoting the L2 listening comprehension ability of the participants of the three proficiency levels.

Table 7

Comparison tests between the experimental and control groups on the 7 LTTs

Independent t-test	Elementary			Lower Intermediate			Upper Intermediate		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>
LT1 exp. × LT1 con.	2.47	45	0.017	3.07	47	0.003	2.57	48	0.013
LT2 exp. × LT2 con.	2.87	45	0.006	2.86	47	0.006	2.97	48	0.005
LT3 exp. × LT3 con.	4.02	45	0.000	3.44	47	0.001	7.69	48	0.000
LT4 exp. × LT4 con.	5.08	45	0.000	4.54	47	0.000	8.42	48	0.000
LT5 exp. × LT5 con.	5.06	45	0.000	5.66	47	0.000	7.14	48	0.000
LT6 exp. × LT6 con.	5.59	45	0.000	6.32	47	0.000	7.62	48	0.000
LT7 exp. × LT7 con.	9.76	45	0.000	9.46	47	0.000	8.57	48	0.000

Note. The participants of each proficiency level took the LTs that were relevant to their own proficiency level; exp.=experimental; con.=control

As it is apparent in this table, the results of independent t-tests, show that in the three proficiency levels, there was a significant difference in the performance of the experimental and control groups on LT1 through LT7 ($p < 0.05$ in all). The quantitative data analyses done in this section provide evidence for the effectiveness of the implementation of DA procedures in the classroom context for improving the L2 listening comprehension ability of the participants. Therefore, part a of the hypothesis made, which claimed that G-DA would enhance the development of the listening comprehension ability of EFL elementary, lower intermediate and upper intermediate learners, was confirmed. Section 3.2 investigates the second part of the research question and determines whether the proficiency level of the learners has any impact on the learners' gains in the G-DA procedures.

3.2 The second part of the research question

The second part of the research question intended to investigate which proficiency level would benefit the most from the DA procedures carried out in EFL classrooms with the aim of assessing and promoting learners' L2 listening comprehension ability. To answer this part of the question, it is necessary to take a closer look at the between-subject variables of this study which were study group and the level of the participants. In section 3.1, as it was observed, due to the presence of an interaction effect between the variables of time and study group, the effect of time had to be separately investigated for the two levels of study group, i.e. control group and experimental group. As noted before, the analysis of the split results of the independent variable obtained from running a one-way repeated measure ANOVA revealed that time only had a significant main effect on the experimental groups. As a result of this, only the effect of the proficiency level on the learners' gains in the experimental groups had to be scrutinized. Table 8 displays the results of the tests of Between-Subjects Effects split based on the study group (experimental group) obtained from running a 2×3×2 mixed between-within subjects ANOVA.

Table 8

Tests of Between-Subjects Effects organized by experimental groups

Source	<i>df</i>	<i>F</i>	<i>p</i>	Partial eta squared
Level	2	1.44	0.24	0.005
Error	68			

The results of this table reveal that the main effect for level did not reach statistical significance [$F(2, 68)=1.44, p=0.24$]. This means that the learners of the elementary, lower intermediate, and upper intermediate

levels benefitted the same amount from the G-DA procedures conducted in this study. Figure 4 provides evidence in support of this finding. In this figure, there are three vectors that show the amount of development the learners of each proficiency level had from pretest to posttest. This growth in the L2 listening comprehension indicate that the mediation the participants of the experimental groups received in the dynamic administration of the pretest plus the mediated support offered to them during the enrichment phase were effective in promoting this skill in them. More importantly, as it is displayed in this figure, the slopes of the three vectors of all proficiency levels are approximately the same value. This could be accounted as a proof that the proficiency level of the learners is not a determining factor in the effectiveness of G-DA procedures in the enhancement of learners' L2 listening comprehension ability. The findings of this section result in the confirmation of part b of the hypothesis of this study, which claimed that the three proficiency level would benefit in the same amount from the DA procedures in the classroom context.

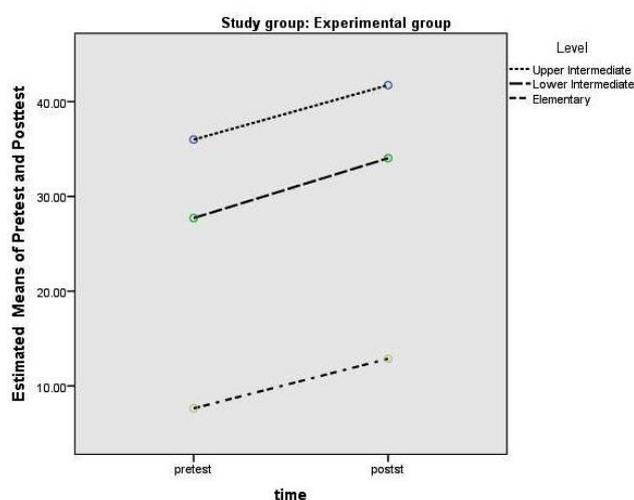


Figure 4. The amount of development of the experimental groups of all proficiency levels from pretest to posttest

4. Discussions and Conclusions

The present study provided a quantitative analysis of implementing DA in EFL classrooms to assess and promote the listening comprehension ability of learners. The research question posed in this study investigated whether G-DA can enhance the development of L2 listening comprehension ability in elementary, lower intermediate, and upper intermediate language learners. The data obtained from the pretest, posttest, and the 21LTs given to the participants of the study during the enrichment phase indicated that, over time, mediated support attuned to learners' ZPD resulted in significant changes in the learners' listening comprehension ability in the three proficiency levels. This finding shows that the implementation of G-DA provided the opportunity for the mediator to engage participants in mediated dialogue about the linguistic features that hindered the learners' comprehension of L2 texts. By doing so, the mediator was able to create a group ZPD through which she attempted to push individual learners' forward in their ZPD. These findings are in line with the results of previous DA studies on listening comprehension for example Ableeva (2010), Hidri (2014) and Hashemi Shahraki et al. (2015). Moreover, they accord with the results of DA studies done in a group format for example Lantolf and Poehner (2011), Davin (2013), and Gibbons (2003). This observation can be explained through Vygotsky's developmental theory (1998) which is the foundation upon which dynamic assessment rests. According to this theory, Haywood and Lidz, (2007) claim that any mediational strategy that is attuned to the learners' ZPD has the potential to uncover those abilities that have matured as well as those which are on the verge of maturing. When these abilities of the learners are revealed, opportunities for the maturation of new abilities could be provided.

In addition, the mediator-learners' dialoging within the G-DA procedures also enabled the mediator to have a diagnostic analysis of the sources of poor performance of the learners in each proficiency level in L2 listening comprehension. It is worthy of notice that a thorough qualitative analysis of the mediator-learners' interactions is required to delineate the causes of learners' poor performance; However, this is not the focus of this study.

Another aim of this study was to find out which proficiency level took the greatest advantage of the mediation received. The results of the relevant statistical measures revealed there was no significant difference among different proficiency levels. In other words, the L2 listening comprehension ability of the learners of elementary, lower intermediate, and upper intermediate proficiency levels improved almost equally, and the proficiency level doesn't affect the learners gain from group dynamic assessment. This finding provides empirical support in the context of L2 learning for Vygotsky's (1978) claim about the actual proficiency level of individuals, that asserts the size of an individual's ZPD is a determining factor in his/her learning development and not the individual's actual proficiency level. Moreover, this finding is in line with Ajideh and Nourdad (2012) who found the learners' proficiency level did not have any effect on intermediate EFL learners DA performance on reading comprehension.

A close analysis of the data obtained from the pretest and posttest of the experimental groups of all proficiency levels showed that although the learners of these groups showed improved ability from pretest to posttest, the standard deviations across the assessments exhibited considerable variation. This shows that some learners were improving more, and some less than others. Another point that deserves considerable attention is that there were some participants in the experimental groups of all proficiency levels that had the same score on the pre-test; however, they performed differently on the posttest. This means after receiving mediational support, some learners had marginal improvements while others exhibited significant developments. The analysis of correlation between the pretest scores and posttest scores of all experimental groups showed that they were all highly correlated. However, this high correlation did not allow the prediction of the development of individual learners on the posttest on the basis of their pretest performance. This finding which runs in accordance with the results of Ableeva and Lantolf (2011), Kozulin and Garb (2002), and Poehner and Lantolf (2013) signifies that the learners were not developmentally at the same ability level. Moreover, this observation is in line with Vygotsky's (1978) claim about the size of a learner's ZPD and his/her actual proficiency level, which maintains the former can be more revealing about the learner's development in learning than the latter.

The present study showed that the mediational support the learners of the three proficiency levels received throughout the G-DA procedures resulted in the improvement of their later independent performance on the posttest. Therefore, it can be concluded no matter what proficiency level a learner is, if correct mediation appropriated to his/her ZPD is provided, that learner could take advantage of the procedure and move forward as much as his/her ZPD allows. Moreover, the results of this study can provide us with a deeper understanding of the use of Vygotsky's ZPD construct to guide teacher-learners mediated dialoging within the classroom context.

The findings of this study can have important implications for all classroom teachers. Regardless of the level of instruction and depending on their needs, classroom teachers can implement G-DA in their classes in a strategic manner to either introduce new concepts or to provide additional instruction on content previously covered. In addition, the G-DA procedures used in the current study have the potential to inform the instructional process regarding specific areas where learners need improvement in L2 listening comprehension.

It is essential to note that there were some limitations within which the results of this study must be interpreted. First, about the sample size; had the study been implemented on a larger sample, the quantitative results gained in this study could have been more reliable and generalizable to other contexts. However, the findings of this study could induce some insight concerning how DA may be organized to function within a class's ZPD. Secondly, about the learners' ability to transfer the abilities gained to other similar or harder contexts; had this study included both non-dynamic and dynamic transfer assessments; claims could be made whether the mediational support provided in this study would result in the transfer of the newly matured

abilities of the learners to other more difficult contexts.

5. References

- Abdolrezapour, P., Tavakoli, M., & Ketabi, S. (2014). Qualitative analysis of mediational strategies in emotionalized dynamic assessment of L2 reading comprehension. *International Journal of Research Studies in Language Learning*, 1(3), 51-66.
- Ableeva, R. (2010). *Dynamic Assessment of listening comprehension in second language learning*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.
- Ableeva, R., & Lantolf, J. P. (2011). Mediated dialogue and the microgenesis of second language listening comprehension. *Assessment in Education: Principles, Policy & Practice*, 18(2), 133 – 149. <http://dx.doi.org/10.1080/0969594X.2011.555330>
- Ajideh, P., & Nourdad, N. (2012). The effect of dynamic assessment on EFL reading comprehension in different proficiency levels. *Language Testing in Asia*, 4(2)
- Alavi, S. M., & Taghizadeh, M. (2014). Dynamic assessment of writing: The impact of implicit/explicit mediations on L2 learners' internalization of writing skills and strategies. *Educational Assessment*, 19(1), 1-16. <http://dx.doi.org/10.1080/10627197.2014.869446>
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465-483. <http://dx.doi.org/10.1111/j.1540-4781.1994.tb02064.x>
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576-598. <http://dx.doi.org/10.1111/j.1944-9720.2009.01030.x>
- Beeston, S. (2000). The UCLES EFL item banking system, *Research Notes*, 2, 8-9.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511732959>
- Budoff, M. (1987). The validity of learning potential assessment. In C.S. Lidz (Ed.), *Dynamic assessment: An interactive approach to evaluating learning potential* (pp. 52–81). New York: Guilford.
- Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom, *Language Teaching Research*, 17(3), 303–322. <http://dx.doi.org/10.1177/1362168813482934>
- Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment*. Baltimore, MD: University Park Press.
- Geranpayeh, A. (2003). A quick review of the English Quick Placement Test, *Research Notes Quarterly*, 12, 8-10.
- Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in a content-based classroom. *TESOL Quarterly*, 37, 247–273. <http://dx.doi.org/10.2307/3588504>
- Hashemi Shahraki, S., Ketabi, S., & Barati, H. (2015). Group dynamic assessment of EFL listening comprehension: Conversational implicatures in focus. *International Journal of Research Studies in Language Learning*, 4(3), 73-89. <http://dx.doi.org/10.5861/ijrsl.2014.955>
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge: Cambridge University Press.
- Hidri, S. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context. *Language Testing in Asia*, 4(4). <http://dx.doi.org/10.1186/2229-0443-4-4>
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23(1), 112-127. <http://dx.doi.org/10.1177/0143034302023001733>
- Lantolf, J. P., & Poehner, M. (2011). Dynamic assessment in the classroom: Vygotskian praxis for L2 development. *Language Teaching Research*, 15(11), 11-33. <http://dx.doi.org/10.1177/1362168810383328>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

- Lidz, C. (1991). Issues in the assessment of preschool children. In B.A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp. 18-31). Boston: Allyn and Bacon.
- Mackey, A., & Gass, S. (2005). *Second language research methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Panahi, P., Birjandi, P., & Azabdaftari, B. (2013). Toward a sociocultural approach to feedback provision in L2 writing classrooms: the alignment of dynamic assessment and teacher error feedback. *Language Testing in Asia*, 3(13). <http://dx.doi.org/10.1186/2229-0443-3-13>
- Poehner, M. E. (2005). *Dynamic assessment of advanced L2 learners of French*. Unpublished Ph.D. dissertation. Penn State University.
- Poehner, M. E. (2008). *Dynamic assessment. A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer. <http://dx.doi.org/10.1007/978-0-387-75775-9>
- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, 43, 471–491.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic Assessment in the language classroom. *Language Teaching Research*, 9(3), 233–265. <http://dx.doi.org/10.1191/1362168805lr166oa>
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research*, 17(3), 323–342. <http://dx.doi.org/10.1177/1362168813482935>
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic Testing*. New York: Cambridge University Press.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191–210. <http://dx.doi.org/10.1017/S0261444807004338>
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. In M. Cole, V. John-Steiner, S. Scribner & E. Souberman, (Eds. and Trans.) Cambridge, MA: Harvard University Press. (Original work published in 1955).
- Vygotsky, L. S. (1998). *The problem of age*. In R. W. Rieber (Ed.), *Child psychology. Collected works of L. S. Vygotsky* (Vol. 5, pp. 187– 205). New York: Plenum.

