

Please cite this paper as follows:

Hashemi Shahraki, S. (2024). Evaluating the Validity of a Socially-situated Assessment: Group Dynamic Assessment of Intermediate EFL Listening Comprehension. *International Journal of Foreign Language Teaching and Research*, 12 (49), 129-150. <http://doi.org/10.30495/IJFL.2023.703378>

Research Paper

Evaluating the Validity of a Socially-situated Assessment: Group Dynamic Assessment of Intermediate EFL Listening Comprehension

Sara Hashemi Shahraki*

Assistant Professor, English Department, Najafabad Branch, Islamic Azad University, Najafabad, Iran

sara_m_hashemi@yahoo.com

Received: January 27, 2024

Accepted: March 17, 2024

Abstract

This study evaluated the validity of group dynamic assessment (G-DA), grounded in Vygotsky's (1987) Sociocultural Theory, implemented in a class of intermediate learners to assess and promote L2 listening comprehension. To navigate the dual goals of assessment and instruction, flexible mediation attuned to the zone of proximal development of the learners was provided within the G-DA interactions. This led to the detection of nine mediational strategies. The validity of these G-DA interactions was explored by extending Poehner's (2011) validation model to classroom setting. Poehner's (2011) model includes two interrelated foci for DA validation: micro and macro-validity. Following Kane's (2021) argument-based approach to validation, evidence-based arguments were developed to explore the appropriateness of each mediational strategy given to the learners (micro-validity) as well as the success of that mediational strategy and the entire G-DA procedure in promoting learners' L2 listening comprehension (macro-validity). Class transcripts were analyzed to gain evidence for the micro- and macro-validity of the G-DA interactions. The findings supported the usefulness of Poehner's validation model in developing validity arguments to determine the appropriateness of the interpretations made about learners' abilities and the effects of the G-DA procedure on their development. Moreover, the study concluded that the analysis of learners' independent performance needed to be added to Poehner's macro-validation model so that it becomes applicable to G-DA.

Keywords: Group dynamic assessment, Micro-validity, Macro-validity, Learner L2 development, validity arguments, and Mediation

این مطالعه روایی ارزیابی پویای گروهی (G-DA) مبتنی بر نظریه اجتماعی و یگوتسکی (1987) را مورد مطالعه قرار داده است که در کلاسی از زبان آموزان متوسط برای ارزیابی و ارتقای درک شنیداری زبان دوم اجرا شد. برای هدایت اهداف دوگانه ارزیابی و آموزش، میانجیگری منعطف هماهنگ با منطقه توسعه نزدیک فراگیران در تعاملات G-DA به زبان آموزان ارائه داده شد. این منجر به شناسایی نه استراتژی میانجیگری شد. در این تحقیق، روایی تعاملات G-DA با بسط مدل سنجش روایی پوهنر (2011) به محیط کلاس درس بررسی شد. مدل پوهنر (2011) شامل دو کانون مرتبط برای سنجش روایی ارزیابی پویا است: روایی خرد و کلان. به دنبال رویکرد مبتنی بر استدلال کین (2021) برای سنجش روایی، استدلال های مبتنی بر شواهد برای بررسی مناسب بودن هر استراتژی میانجی ارائه شده به فراگیران (روایی خرد) و همچنین بررسی موفقیت آن استراتژی میانجی و کل G-DA در ارتقای درک شنیداری زبان دوم (روایی کلان) به کار گرفته شد. برای به دست آوردن شواهدی برای سنجش روایی خرد و کلان، رونوشت تعاملات G-DA انجام شده در کلاس تجزیه و تحلیل شد. یافته ها از سودمندی مدل سنجش روایی پوهنر در توسعه استدلال های سنجش روایی برای تعیین مناسب بودن تفاسیر انجام شده در مورد توانایی های فراگیران و تأثیرات روش G-DA بر توسعه آن ها حمایت می کند. علاوه بر این، این تحقیق نشان داد که تجزیه و تحلیل عملکرد مستقل فراگیران باید به مدل سنجش روایی کلان پوهنر اضافه شود تا برای G-DA قابل استفاده باشد.

Introduction

Unlike standardized testing, the development-oriented assessment, known as Dynamic Assessment (DA), does not seek to describe learners' performance consistencies (Lantolf & Poehner, 2004). Rather, this approach to assessment which originated in Vygotskian Sociocultural Theory (SCT) reflects the dynamism and ongoing nature of development and prescribes examiner-examinee dialogue during the assessment procedure (Davin, Herazo, & Sagre, 2017). To illuminate processes of learner development, DA draws on Vygotsky's (1978) Zone of Proximal Development (ZPD). This zone is commonly 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). Therefore, assessing learners through DA could lead to the diagnosis of not only the developed abilities of the learners but also the abilities that are in the process of forming (Haywood & Lidz, 2007; Sternberg & Grigorenko, 2002).

In DA, mediation is assumed to play a key role in guiding learner development since according to Vygotsky, development occurs within a learner's ZPD under appropriate mediation. The form of the mediation in DA can be pre-scripted, i.e. interventionist approach, or flexible, i.e. interactionist approach (Lantolf & Poehner, 2004). In both approaches, mediation is the support given to learners in a systematic way. This is carefully calibrated to learners' emerging needs and responsiveness (Lantolf & Poehner, 2013).

To avoid purely impressionistic interpretations about learners' abilities diagnosed through DA, as with any assessment, it is imperative to address the matter of validity in this type of assessment (Poehner, 2011). Validity in testing has primarily been conceptualized as discovering whether a test "measures accurately what it is intended to measure" (Hughes, 1989, p. 22). In other words, it determines how much the assessment procedure uncovers what it is intended to uncover about the individuals' knowledge and abilities. Therefore, one aspect of validity investigations is to focus on interpreting the assessment outcomes and the appropriateness of decisions based on the assessment results (Bachman, 2000). Most of the research into validity has addressed standardized tests (e.g. Baird et.al., 2017; Chapelle, Enright, & Jamieson, 2008; Mislevy, 2009;) rather than classroom-based assessments. Validity in DA is likened to validity in classroom-based assessment on the grounds that DA like some of the classroom-based assessment aims to help individuals move beyond their current capabilities (Poehner, 2008b). To illuminate how the concept of validity differs in a psychometric approach to assessment, i.e. standardized testing, and classroom-based assessment, the notion of validity is briefly described within each approach.

Validity in a psychometric approach to assessment

There was no model of validity till about 1920; at the time, the main concern was with the precision and cohesion of the measurements (Kane, 2012). In the 1920s, language testers began to show interest in validating their tests; therefore, the notion of criterion validity was developed and the test was validated against a plausible criterion measure (Cureton, 1951). Although this was a major milestone in the process of test validation, there was, sometimes, no way to validate a test when there was no other more plausible criterion (Ebel, 1961). To overcome this problem, the content validity model was developed. In this model of validation, attempts were made to show the content of the test was a representative from a larger universe of tasks of which the test was assumed to be a sample (Cureton, 1951). However, content-based claims for validity faced criticisms. One criticism was that content-based claims tended to be subjective. Another was that content-related evidence supported claims that went beyond content and provided judgments about the test takers internal processes (Cronbach, 1971).

By early 1950s, the language test developers realized that the criterion and content validity models were not enough to provide interpretations about the test takers psychological processes.



To compensate for the shortcomings of the criterion and content validity models, Cronbach and Meehl (1955) presented the construct validity model as an alternative. Cronbach and Meehl (1955) claimed the construct validity model could be applied “whenever a test is to be interpreted as a measure of some attribute or quality which is not operationally defined” (p. 282), and “for which there is no adequate criterion” (p. 299). Moreover, they went on to say that construct evidence “is desirable for almost any test” (Cronbach & Meehl, 1955, p. 282). Cronbach and Meehl (1955) assumed that measures of the construct could be validated as a result of validating the theory from which some attribute or quality is derived. At the time, none of the mentioned validity models was considered more comprehensive and the choice of model depended mainly on the availability of data (Guion, 1977).

Further Messick (1988) claimed that construct validity could be considered as a unifying framework within which criterion and content validity were embedded. Messick’s (1988) unitary framework of validity made a revolution in the definition of validity. Despite its appealing conceptual framework, Messick’s unified version of validity seemed not to provide a clear guidance on how to validate score interpretations or uses. This caused House (1980) and Cronbach (1988) to address these limitations. They suggested an argument-based approach for validation was required to specify the intended interpretation and use of test scores. Kane (1992) introduced an argument-based approach to validation that included interpretive and validity arguments. This was later refined in Kane (2013). Furthermore, to validate the uses of score interpretation, Bachman (2005) proposed a framework to test validation, i.e. Assessment Use Argument (refined in Bachman & Palmer, 2010). In his framework, Bachman argued that a validity argument could be supplemented by a utilization argument (which was formulated to address the relevance and usefulness of the score meaning).

Validity in Classroom-based Assessment

In recent years, there has been a growing recognition among second language (L2) scholars that transferring the validation processes of standardized tests to the classroom assessment is problematic on the grounds that these two types of assessment differ in fundamental ways (e.g. Fulcher & Davidson, 2007; Kane & Wools, 2019; Moss, 2003, Stobart, 2012). Moss (2003), for example, asserts that the theoretical underpinnings of standardized tests and classroom assessment vary in the way they consider assessment and teaching. Standardized testing postulates assessment as a discrete activity which is distinct from teaching and learning (Cheng, 2005; Cheng, Watanabe, & Curtis, 2004). Drawing on socio-cultural theory and hermeneutics, Moss declares that unlike standardized testing practices, the purpose of classroom assessment is not only to identify learners’ abilities but also to inform better teaching and more efficient learning. Brookhart (2003) also argues assessment and teaching are integrated within the classroom. She sees this in terms of Vygotsky’s ZPD and claims classroom teachers are constantly assessing learners to identify where additional teaching is needed. Likewise, Fulcher and Davidson (2007) contend the main difference between standardized tests and classroom assessment is the context of classroom. They maintain that in the psychometric approach to test validation, context is usually referred to as one part of construct-irrelevant variance whereas in classroom assessment, context, i.e. the learning environment, is part of the construct, and is therefore directly related to the assessment of the learners.

Furthermore, some researchers have warned against transferring statistical analyses and technical approaches of test validation of standardized testing to classroom assessment (e.g. Kane, 2017; Poehner, 2011; Teasdale & Leung, 2000; Torrance, 1995). In standardized tests, the interpretation of test performance often relies on advanced statistical analyses which tend to be highly technical. These analyses are not applicable to classroom assessments which unlike standardized tests do not yield hundreds or thousands of scores. More importantly, standardized



tests and classroom assessment have completely divergent assumptions about assessment. As Moss (2003) points out, standardized testing considers assessment as a discrete activity, administrated to the test-takers in isolation, the result of which is usually reduced to a score or ranking. In contrast, assessment in classroom-based practices, is an ongoing and dynamic activity, typically done collaboratively which culminates in a detailed profile of the abilities of test-takers (Kane & Wools, 2019; Moss 2003).

As noted above, due to the purpose and context of testing, the validation processes of standardized tests and classroom assessment are quite different. However, in both types of assessment, it is important to collect evidence that justifies the appropriateness of the decisions made (Fulcher & Davidson, 2007; Moss 2003; Poehner, 2011). In a psychometric approach, evidence has to be amassed to support the claims about the proposed interpretations and uses of scores gained (Kane, 2012). Likewise, in classroom-based assessment evidence is collected to show this type of assessment has resulted in improved learning. This in turn attests to the usefulness of the assessment and the validity of the interpretation of evidence (Fulcher & Davidson, 2007).

Validity research in classroom-based assessment

Literature of test validation reveals a few studies addressing validity in classroom assessment (e.g. Moss, 2003; Stobart, 2012). In her study, Moss (2003) explores the shortcomings of psychometric validity for classroom assessment practices. By providing additional theoretical resources from the sociocultural theory and hermeneutics, she questions the value of the validity paradigm of psychometric approaches to assessment to classroom practices. In addition, Moss seeks to delineate how teachers should look at their own aims within a particular learning context to determine what constitutes valid assessment for that context. Hence, she suggests classroom assessment validity procedures should be built within a “framework to guide thinking and action” (Moss, 2003; p.15).

Moreover, Stobart (2012) suggests the validity of any classroom-based assessment is demonstrated through how well it meets its purpose of improving learning. Stobart further claims that a strict adherence to determining whether a specific type of classroom assessment has led to further learning is not sufficient for a validity inquiry of classroom assessment. A validity inquiry should also seek to find what hinders the underlying intention of assessment to be realized. Thus, Stobart (2012) suggests in order to examine the validity of classroom assessment, “the cultural and learning context, the quality of classroom interaction, the teacher’s and learners’ clarity about what is being learned, and the effectiveness of feedback” (p. 235) should be analyzed. This, he believes, can help find out what supports or threatens effective classroom assessment.

Validity research in dynamic assessment

Studies on the validity of DA seem also to be scarce (e.g. Poehner, 2008b & 2011). Lantolf and Poehner (2004), in an article in which they outline a theoretical framework for the application of DA procedures to L2 assessment and pedagogy, assert “DA derives its validity not from the assessment instruments but from the procedures followed in the administration of the instrument” (p. 67). They claim the validity of DA is established to the extent that DA is able to achieve its purpose, namely pushing the learner’s language abilities forward. In another paper on fundamentals of DA, Lantolf and Poehner (2007) propose when the validity of the activity of teaching-assessment is in focus, it is necessary to interpret its impact on learner development. Moreover, they assert considering that the process of teaching and assessment in the classroom is ongoing, its validation should have the same nature. It is worth noting that according to Aljaafreh and Lantolf (1994) and Poehner (2008a) a learner’s development is not always manifested in



his/her independent performance, rather subtle changes in a learner's responsiveness to mediation may also indicate development.

Further Poehner (2008b) addresses the issue of validity in DA within a validity framework. He maintains when assessment is reoriented from a measurement activity (as in standardized tests) to one focused on learner development (as in DA) a theoretical framework is needed that fosters cooperation with learners and intervention. He, therefore, proposes that the theoretical framework articulated in Vygotsky's (1987) Sociocultural theory for development-oriented assessment can serve as the kind of "robust validity framework to guide thinking and action" (Moss, 2003, p. 15) in classroom assessment which Moss has called for. More importantly, Poehner (2008b) asserts the core argument to establish validity in DA is to determine the extent to which the assessment procedure can support learner development. Poehner (2008b) also remarks in DA, the constructs or abilities assessed are always in flux; i.e. the focus is on changes in the learner abilities. In addition, he claims the interpretations of evidential basis for learner abilities are more complex in DA since the evidence itself is dynamic rather than stable. This means during the interactions, the mediator should have moment-by-moment interpretation of learner abilities to provide the learners with appropriate types of mediation.

In another study on DA validity, Poehner (2011) refers to Messick's (1989) approach to validity and suggests the argumentation in Messick's model is more likely relevant to validating classroom assessment. Poehner, further, asserts in order to determine the appropriateness of diagnoses of learner abilities and the instruction informed by them, evidence-based arguments seem inevitable. He, therefore, proposes two interrelated foci for validating L2 DA: micro- and macro-validity. Micro-validity, in Poehner's terms, examines specific mediation or mediating strategies given by a mediator during DA. It comprises the stages of developing logical arguments about the appropriateness of the interpretations of learners' abilities and how they end in offering mediational support. Poehner explains that the micro-validation process begins with a learner's initial action. Then based on that action, the mediator makes a provisional interpretation of the learner's ability and hence provides mediational support. However, the interpretation made is tentative and the learner's responsiveness to that mediation will result in either accepting or rejecting that interpretation. Vygotsky (1987) emphasizes the important role of learners' responsiveness to mediation in revealing what is in the learners' ZPD. If the interpretation is rejected then the process of making interpretation is repeated.

Macro-validity, on the other hand, focuses on the entire DA procedure to determine the success of the interactions in revealing and guiding learner development (Poehner, 2001). The macro-validation process includes detecting three forms of evidence: *changes in mediation*, *changes in learner responsiveness*, and *learner verbalization*. By identifying changes in the *quality of mediation required by a learner* as well as changes in *learner responsiveness to mediation* (how learners respond to a mediation), Poehner claims evidence can be accrued to argue for a particular diagnosis of a learner's ZPD. He, further, suggests the commentaries verbalized by the learner ascertain the diagnosis made about his/her abilities. Finally, Poehner recommends his model to classroom teachers, suggesting that it helps them form and verify the interpretations of their interactions with learners and minimizes purely subjective statements about learners' abilities.

It needs to be mentioned that Poehner's (2011) model, detects validity evidence for the interactions made between one mediator and one learner, and hence supports an individual's development through his/her ZPD. The question whether this model of validation could also be extended to Poehner's (2009) group dynamic assessment (G-DA) motivated the present study. In GDA, the focus is not on one individual but on the entire class. Poehner (2009) bases G-DA on Vygotsky's (1998) description of ZPD as "the optimum time for teaching both the group and each individual" (Poehner, 2009, p. 204). He believes through offering mediation to a group of



learners, it is possible to construct a group's ZPD. Therefore, in G-DA, by attuning the mediational support to the group's ZPD, the group could become a psychological entity within which the development of the group and each group member would be interrelated. Poehner (2009) introduces cumulative and concurrent approaches in G-DA. In the cumulative approach, when a student produces an incorrect answer, mediation prompts are provided to that particular student until he or she finds the correct answer. Thus, in cumulative G-DA, through co-constructing ZPDs with individuals, the aim is to move the group forward in its ZPD. On the contrary, in the concurrent approach, the mediator opens dialog with the entire group and provides mediation upon realizing that a learner has faced a problem. However, that same learner is not required to provide an answer in response to the mediation received. As a result, in concurrent G-DA, the goal is to support the development of each individual by working within the group's ZPD. It seems possible to determine the validity of cumulative and concurrent G-DA through Poehner's (2011) validity model.

Current study

In this study, attempts were made to extend Poehner's (2011) validity model to G-DA. In so doing, the aim is to develop validity arguments, following Kane (2021), to explore the validity of the G-DA procedures carried out to enhance L2 listening comprehension ability through utilizing Poehner's (2011) validation model. In other words, the purpose has been to identify validity evidence for the teacher interpretations about learners' abilities and thereby the mediation which addresses learner development in G-DA. To create evidence-based validity arguments, according to Kane (2021) the inferences and interpretations made by the teacher should be supported by *warrants*. These *warrants* themselves are based on assumptions which require *backing* or *support*. Thus, the aim is to provide an explicit statement concerning the teacher interpretations and supporting assumption that gets us from the observed performances to the claims made about these performances. The study, therefore, will address the following questions:

Q1. What evidence is observed in determining the micro-validity of G-DA of intermediate EFL listening comprehension?

Q2. What evidence is observed in determining the macro-validity of G-DA of intermediate EFL listening comprehension?

Method

The current study has revisited the finding of prior project (Hashemi Shahraki, Ketabi, & Barati, 2015) examining the applicability of G-DA to L2 listening comprehension. The project aimed to organize classroom interactions in terms of the learners' ZPD. The principal purpose of employing G-DA was to diagnose the sources of L2 listening comprehension difficulty in intermediate learners and to promote their emerging capacities in this skill.

Participants

The sample under study was EFL learners at a language institute in Iran. Two upper-intermediate (based on the Oxford Placement Test, Geranpayeh, 2003) intact classes were selected to participate in this study. One of the classes was randomly assigned as the experimental group (N=24) while the other as the control group (N=26). The 50 participants were Persian females (14-18yrs.) who had taken English courses for three years in that institute. Students from both groups were asked to avoid any additional English input during the course of this study. All the participants were asked to sign a consent form for taking part in this study as well as for being audio-visually recorded.



Materials

This study included a pretest/posttest and seven listening tests (LT). Given the level of the participants and the expert judges' analyses, the listening material and comprehension questions of the tests were chosen. The pretest/posttest was extracted from Interaction/Mosaic placement test (Tarver Chase, Hanreddy, & Whalley, 2013). The test included fifty four-option multiple choice (MC) items. The test was to measure test takers' listening ability in extracting the gist of what they heard, getting particular details, identifying speaker's opinion, and making inferences. The reliability of the test was estimated 0.93 (Cronbach's Alpha). The listening material of the pretest/posttest consisted of question and statement items as well as dialogic and monologic texts.

Based on the learners' independent and mediated performance on the pretest, the listening material of the seven LTs along with their comprehension questions was selected from Mosaic 1 Listening/Speaking (Hanreddy & Whalley, 2008). Each LT had 6-8 four-option MC questions. The listening material of each LT was either a part of a longer conversation or a number of short conversations with a focus on implied meanings all at the same difficulty level. The listening texts of the pretest/posttest and the LTs were at normal speech rate, i.e. 140 words per minute (Buck, 2001), all in standard American accent. Necessary modifications were made to the items and choices of the pretest/posttest and LTs as a result of their piloting.

The rationale for using MC items as comprehension questions in the present study was that they could easily be scored. More importantly, MC items could assess a variety of listening sub-skills. The time required for completing the non-dynamic pretest/posttest and each non-dynamic LT was 40 minutes and 10-15 minutes, respectively. However, the dynamic implementation (see 'procedure' below) of the tests took longer: the dynamic pretest=80 minutes, the dynamic posttest=60 minutes, and each dynamic LT= 15-25 minutes. An example of one of the question items of the pretest/posttest is provided below.

Narrator: Couldn't you have arrived an hour later?

- a) I'm sorry I was late.*
- b) I couldn't have come earlier.*
- c) Would you like me to come back in a while?*
- d) Sorry we left so late.*

(Source: Item10, p. T39, the Interactions/Mosaic Listening Placement Test, Tarver Chase et al., 2013).

Procedure

The G-DA procedures in this study comprised non-dynamic and dynamic pretest- enrichment phase- non-dynamic and dynamic posttest (following Ableeva, 2010) for a time span of 10 weeks as shown below:

- Week 1 → The Oxfords' Placement Test
- Week 2 → Non-dynamic and dynamic pretest
- Weeks 3-9 → Enrichment phase (7 Listening Tests: Non-dynamic and dynamic assessments)
- Week 10 → Non-dynamic and dynamic posttest

From week2 to week10, initially the learners of the experimental and control groups took the test relevant to that session in a non-dynamic format. The non-dynamic administration of the tests was the same for both groups. It means they listened to the text or texts once and answered comprehension questions. Then the test sheets were collected, and only the experimental group took each test dynamically. In the dynamic administration of each test, the mediator, who was one of the researchers, replayed the listening material portion by portion for the class to provide their recalls either in their L1, Persian, or English. Upon the learners' failure to provide an acceptable recall, the mediator intervened and offered mediation. The mediation provided was not pre-specified but emerged from the mediator's ongoing collaborations with learners, i.e. the



interactionist approach (Lantolf & Poehner, 2004). Moreover, to mediate a learner's problem, the mediator opened dialog with the entire group and provided mediational support, i.e. the concurrent approach (Poehner, 2009). In this approach, the interaction shifts rapidly between the primary and secondary interactants since a learner's question or comment set the stage for other learners' contribution. To ensure the active participation of the learners', before each dynamic test, the learners were told their silence would be construed as their lack of understanding of the listening texts.

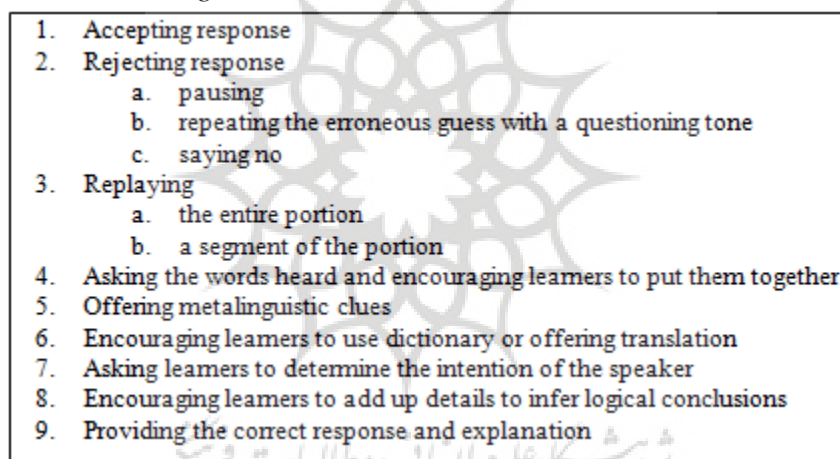
Data Analysis

The audio- and video-recorded mediator-learner interactions were transcribed. Thematic analysis (Guest, 2012) was used to analyze the transcript data. This would help to identify the mediational strategies offered and to support evidence for micro- and macro-validation.

To detect the mediational strategies, the transcribed data were analyzed to identify instances of mediation support given to the learners. The data were then coded and categorized by two expert judges (inter-rater reliability=0.79). A taxonomy of nine mediational strategies emerged from the analyzed data (Figure 1). Following Aljaafreh and Lantolf (1994) and Ableeva (2010), the strategies were arranged from the most implicit to the most explicit.

Figure 1

Typology of mediational strategies

- 
1. Accepting response
 2. Rejecting response
 - a. pausing
 - b. repeating the erroneous guess with a questioning tone
 - c. saying no
 3. Replaying
 - a. the entire portion
 - b. a segment of the portion
 4. Asking the words heard and encouraging learners to put them together
 5. Offering metalinguistic clues
 6. Encouraging learners to use dictionary or offering translation
 7. Asking learners to determine the intention of the speaker
 8. Encouraging learners to add up details to infer logical conclusions
 9. Providing the correct response and explanation

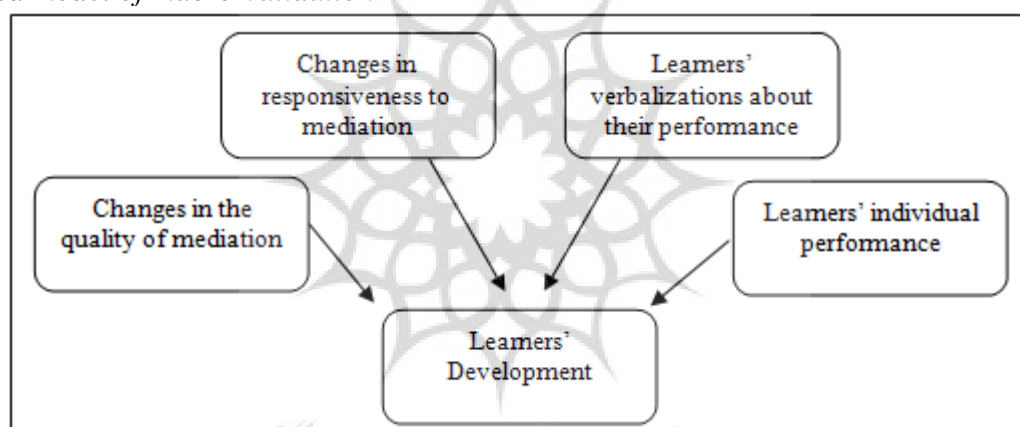
To address the first research question, we applied Poehner's (2011) model for micro-validation to the data. During every G-DA session, the mediator deployed the micro-validation process and searched for evidence to support the interpretations on learner abilities. The mediator began the process by formulating a provisional interpretation about what the action might indicate about the learners' abilities. For instance, a provisional interpretation about the learners' inability in providing a recall could mean they did not have access to adequate attentional resources to decode the words heard, or had difficulties in word recognition. Then based on her interpretation, the mediator provided a mediational strategy. The learners' responses to the mediation provided *warrant* for the mediator to either accept or reject her provisional interpretation (Kane, 2021). The acceptance of the provisional interpretation, in turn meant and provided *support* that the diagnosis made about the learners' abilities and the mediational strategy provided were appropriate. On the other hand, the rejection of the provisional interpretation signaled the inappropriateness of the diagnosis made and the inadequacy of the mediational strategy offered. In this case, the mediator continued the process of observing learners' actions,

developing tentative interpretations of learners' abilities, providing the required mediational strategy and analyzing learners' responses to the mediation offered. Such a process continued until the mediator found evidence for the appropriateness of the diagnosis made and the mediational strategy offered.

As for the second research question, unlike what Poehner's (2011) macro-validation model suggests, *changes in mediation*, *changes in learner responsiveness*, and *learner verbalization* did not seem to provide adequate evidence for the success of G-DA in promoting each individual learner's L2 listening comprehension. In other words, by analyzing the overall pattern of mediation required by the learners and their responsiveness to the mediation received, it was only possible to determine whether the group had any progress in L2 listening comprehension. Thus, individual learner's development within the group would be left unnoticed. In concert with that, Poehner (2009) states, "both responsiveness to support as well as independent performance (emphasis added)" (Poehner, 2009, p. 489) of the learners need to be analyzed in order to discern whether each individual member in a group is also developing. Consequently, the present study added learners' independent performance to the macro-validation model (Poehner, 2011). The resulting model was then implemented to find evidence for the macro-validation of G-DA. Figure 2 shows the way the above pieces of evidence function in relation to each other.

Figure 2

Expanded model of macro-validation



To gain evidence for changes in mediation, after the mediating strategies were coded and categorized, they were tallied and a frequency count was reported for each G-DA session. Moreover, the transcript data was analyzed further to find instances where the mediation offered resulted in eliciting the desired response (changes in learners' responsiveness) and to detect learners' verbalization about their own performance. The changes detected in learners' responsiveness were coded and categorized; their frequency count was then recorded for each G-DA session. Finally, to analyze the independent performance of the learners in both groups, their answers on the pre and post-tests were scored, and then several t-tests were run on the resulting data.

Results and Discussion

The excerpts that follow are randomly selected from among the G-DA interactions of the experimental group during the enrichment phase (weeks 3-9, see Procedure). They are only to demonstrate examples of L2 G-DA interactions in classroom setting and to provide a point of reference for the G-DA validity discussion presented below.

Micro-Validation of the G-DA Procedures

The first research question concerned the detection of evidence for evaluating the micro-validity of the G-DA interactions provided to assess and/or promote intermediate learners' L2 listening comprehension ability. Excerpt 1, taken from LT3, enrichment phase, week5, illustrates the process of the micro-validation of the mediational strategies of replaying, accepting response, rejecting response, and offering metalinguistic clues (Figure 3). Put differently, it depicts what interpretations of learner abilities motivated the mediator to provide these mediational strategies and brings evidence to support these interpretations.

Figure 3

Excerpt 1(Extracted from LT3, Enrichment Phase, week 5), Note: M=Mediator, S=Student

- [.....*Wife*: It's nice that banks are beginning to make more loans to people with low incomes. On the other hand, that money comes with a lot of strings attached....]
1. M: What did you hear?
 2. Ss: [Silent.]
 3. M: Yes,.... Nothing?
 4. Ss: [Silent.]
 5. M: Let's listen again. [The entire portion is replayed-Strategy 3, Sub-strategy a]
 6. S1: It's about banks.
 7. M: Aha... [Strategy 1]
 8. Ss: [Silent.]
 9. M: OK. [The entire portion is replayed once more-Strategy 3, Sub-strategy a]
 10. S2: It's nice that the banks give loans.
 11. M: Good..., and? [Strategy 1]
 12. S3: to everybody?
 13. M: Mmmm. [Strategy 2, Sub-strategy a]
 14. S4: No, the lady said, "It's good that they give loans to poor people."
 15. M: Great, what else? [Strategy 1]
 16. Ss: [Silent.]
 17. M: Yes?
 18. Ss: [Silent.]
 19. M: Let's listen again then. [The second sentence of the woman is replayed two times-Strategy 3, Sub-strategy a]
 20. Ss: [Silent.]
 21. M: How does she start her sentence? [Strategy 5]
 22. Ss: On the other hand, [Some students say it together].
 23. M: When do we use 'on the other hand', 'in contrast', 'on the contrary'...?
 24. S6: When we want to say something opposite, something bad. [Strategy 5]
 25. M: Very good. Now, let's listen once more. [The second sentence of the woman is replayed-Strategy 3, Sub-strategy a]
 26. S7: She says that money has a lot of strings.
 27. M: That money has a lot of strings? [Strategy 3, Sub-strategy b]
 28. S8: The lady says, "But that money comes with a lot of strings attached." So in this sentence she wants to say something bad about loans, but I don't know....em what....em bad thing.
 29. M: That's it. Ok, what does "a lot of strings attached" mean? [Strategy 1]

Replaying

As demonstrated in this interaction, *the learners' initial action*, i.e. being silent (lines 2 and 4), was their inability to provide a recall of the content of the portion heard. According to Vygotsky (1998) and as stressed by Lantolf and Poehner (2013) the purpose of providing mediation in DA is to determine the minimum level of support learners require to perform successfully. Therefore, the mediator's *provisional interpretation* about the learners' listening comprehension was that the

learners were not able to provide a correct recall since they did not have access to adequate attentional resources to decode the words heard. She assumed multiple hearings of the text could free up required attentional resources (Tyler, 2001) to notice the aspects they had failed to catch during the first listening. The mediator, thus, decided to offer the *replaying* strategy using the sub-strategy of *replaying the entire portion*. S1's response (line 6) illustrates the incompleteness of this mediational strategy, resulting in the *rejection* of the provisional interpretation of the learners' abilities.

Accepting response

The *learner's response* in line 6 in turn directed the mediator to make another provisional interpretation (see Figure 2). Her provisional interpretation was the learners were uncertain about the correctness or appropriateness of their response and needed caring support to clear this doubt. Based on this tentative interpretation, the mediator set up the next mediating move. She provided the *accepting response* strategy (line 7) reflecting Vygotsky's affective-volitional aspect of learning (Warford, 2010). Line 8 shows the *learners' response* to this mediation, i.e. further silence, which led the mediator to reject her second *provisional interpretation* about learners' abilities. At this point, the mediator, however, decided not to provide a more explicit mediational move and resorted to the replaying strategy used before. She assumed another exposure to the listening text could help the learners regulate their thoughts. The partial recall (*learner response*) given by S2 (line 10) testified her conjecture to some extent and provided evidence for her interpretation of the learner's abilities. Upon observing the learner's *responsiveness to the mediation* offered, the mediator, therefore, made use of the accepting response strategy again.

Rejecting response

S3's incorrect guess in line 12 showed the futility of the accepting response strategy. The mediator interpreted the learner's response as a demand for a more explicit type of mediation. Accordingly, she provided the strategy of *rejecting response* applying the sub-strategy of *pausing* (line 13) to clearly send a message to the learners that something was amiss with their performance. As depicted in line 14, this strategy encouraged S4 to make an attempt to overcome the difficulty and to provide a recall of the first sentence heard. The learner responsiveness to the mediational move could offer important evidence in support of the mediator's *provisional interpretation* of learners' abilities which could in turn lead to the acceptance of this interpretation. The repeated process of making interpretation of learner's abilities and its subsequent refutation or confirmation (based on observing the learners' responsiveness to the mediation received) had the potential to refine the mediator's understanding of the learners' deficiencies. The learners' responses provided up to line 14 suggested to the mediator that the learners were experiencing difficulties with word recognition. Thus, the recognition of words was facilitated by less explicit forms of mediation.

In lines 16 and 18, the mediator observed the learners' problems in producing a correct recall of the second sentence. Moreover, the implicit sub-strategy of replaying the entire portion seemed not to be sufficient. The mediator interpreted the learner's response as calling for further assistance. She, therefore, decided to use the sub-strategy of replaying a segment... (line 19). The purpose of this sub-strategy was to narrow down the scope of the focus and to draw the learners' attention to the problematic area (a particular lexical phrase or a grammatical structure). Nevertheless, the silence of students (line 20) helped the mediator realize her interpretation about the learners' abilities was incorrect. The inability of the learners to provide an acceptable recall showed that the learners' problem was perhaps far more than word recognition.



Offering metalinguistic clues

In lines 16 and 18, the mediator observed the learners' problems in producing a correct recall of the second sentence. Moreover, the implicit sub-strategy of replaying the entire portion seemed not to be sufficient. The mediator interpreted the learner's response as calling for further assistance. She, therefore, decided to use the sub-strategy of replaying a segment... (line 19). The purpose of this sub-strategy was to narrow down the scope of the focus and to draw the learners' attention to the problematic area (a particular lexical phrase or a grammatical structure). Nevertheless, the silence of students (line 20) helped the mediator realize her interpretation about the learners' abilities was incorrect. The inability of the learners to provide an acceptable recall showed that the learners' problem was perhaps far more than word recognition.

Figure 4

Excerpt 2 (Extracted from LT4, Enrichment Phase, week 6) Note: M=Mediator, S=Student

- [.....the equity of creativity can be stifled by a particular context or environment.....]
1. M: What did you hear?
 2. Ss: [Silent.]
 3. M: Let's listen again. [The sentence is replayed twice.]
 4. Ss: [Silent.]
 5. M: OK. Let's listen once more. [The first and second portion of the sentence is replayed]
 6. Ss: [Silent.]
 7. M: What words did you hear?
 8. S1: Creativity
 9. S2: Environment
 10. M: Great, what is the verb of the sentence? Let's hear the verb. [The sentence is replayed]
 11. S3: rifle
 12. M: rifle?
 13. S4: No, it was stifled.
 14. M: What does it mean?
 15. S4: To stop something from happening.
 16. M: Perfect, but what stifled what? What other words did you hear?
 17. Ss: [Silent.]
 18. M: OK, let's listen again then. [The sentence is replayed]
 19. S5: The verb is "can be stifled"; it's passive.
 20. M: Very good, what other words did you hear?
 21. S6: equity of creativity, a particular context
 22. Ss: [Silent.]
 23. M: Good. Now, who can tell us what can be stifled and by what?
 24. S1: the equity of creativity can be stifled by a particular context
 25. M: Aha, and..?
 28. S6: and by the environment
 29. M: That's it.

Excerpt 2, taken from LT4, enrichment phase (week 6), shows the micro-validation process of the strategy of asking the words heard and encouraging learners to put them together (Figure 4). An explanation of what interpretation of learner abilities led the mediator to provide this strategy, and the evidence she gained through learners' responsiveness to mediation is given in this interaction.

Asking the words and encouraging learners to put them together

Based on the *initial response of the learners*, i.e. not producing any recalls, the mediator offered them one of the most implicit strategies on the list, namely the *replaying* strategy (line 3). The learners' response to this strategy (lines 4 and 6) revealed their inability to produce even a partial recall after multiple exposures, indicating the inadequacy of this mediational move. Following her observation, the mediator made the *provisional interpretation* that the inability on the part of the learners was possibly due to the semantic density or syntactic complexity of the

sentence. Consequently, she deemed appropriate to offer the mediational strategy of *asking the words and putting them together*. The mediator believed that by breaking down the entire sentence into manageable portions, she could lower the comprehension load for the learners and encourage them to take part in a joint intellectual activity. This strategy supports the principle that purposeful group work results in sharing of knowledge and abilities which moves the group forward in its ZPD while also benefitting individuals (Petrovsky, 1985). The collaborative work done among the mediator and learners (displayed in lines 7 through 23) aided S1, who had earlier provided only a partial segment, to put together the decoded words and, thus, uttered most of the content (line 24). Her recall was then completed by S6 (line 26). The learners' responses in this excerpt suggested that their inability in providing a recall was due to the semantic density of the sentence heard and was not related to its syntactic structure. Evidence in support of the mediator's *provisional interpretation* of learners' abilities is gained through observing the effectiveness of the mediational strategy offered, and, thus, the acceptance of this interpretation.

Figure 5

Excerpt 3 (Extracted from LT2, Enrichment Phase, week 4) Note: M=Mediator, S=Student

- [.....Husband: You can say that again! I couldn't agree with you more. It's definitely better to pay cash than to pay for years and years.
Wife: Yes, but don't forget there are exceptions to that rule.....]
1. M: What did you understand from this conversation?
 2. Ss: [Silent.]
 3. M: Let's listen again. [The conversation is replayed twice.]
 4. S1: First, the man says, "you can say that again." But I don't know what the woman emm said before.
 5. M: Oh. You're thinking about what the woman had said before?
 6. Ss: [Silent.]
 7. M: Maybe, it has another meaning.
 8. S2: You mean it's an expression?
 9. M: Yes. What does it mean? Who knows?
 10. Ss: [Silent.]
 11. M: OK, let's check it. Refer to your dictionaries.
 12. S1: I have a problem. I can't find it in my dictionary.
 13. M: Under what headword are you looking?
 14. S1: I'm searching for "you"?
 15. M: No, no. This is an idiom. Do you know what an idiom is?
 16. S3: [In Persian, she says, "It's a group of words that you can't get its meaning from the words that make it."]
 17. M: Exactly, so you have to look under the headword "say".
 18. S4: It's an idiom. It means "I agree with you completely."
 19. M: Great, what else did you understand?
.....lines deleted.....
 31. M: Very good, what did you understand from this conversation?
 32. Ss: [Silent.]
 33. M: Let me narrow it down, what is the woman implying? What did she mean?
 34. Ss: [Silent.]
 35. M: OK, what word does the woman start her sentence with?
 36. S9: with "but"
 37. M: That's right. What does this word show?
 38. S9: It shows contrast.
 39. S10: Oh, it shows that the woman disagrees with the man.
 40. S11: I know this, but what rule are they talking about?
 41. Ss: [Silent.]
 42. M: emmm. It's about banks, can't you infer what rule they're talking about..em from what you heard?
 43. Ss: [Silent.]
 44. M: OK, as you remember the man said, "It's definitely better to pay cash than to pay for years and years."
What do we pay back little by little and usually for years?
 45. S11: a loan?
 46. M: Yes, so the rule refers to...?
 47. S11: Getting a loan?
 48. M: Mmm, so what can we understand from this conversation, overall?
 49. S9: They are disagreeing about something.
 50. M: Great, what else?
 51. Ss: [Silent.]
 52. M: The rule refers to buying something in cash and not getting a loan. Is it always possible to buy something cash?
 53. S12: No, sometimes something is very expensive, like buying a house.
 54. M: Exactly. So, the woman is implying in her speech the man has to think through what she had said and not agree with her without thinking. Sometimes you cannot save enough money to buy something cash, and you do not have any other choice than getting a loan.

The micro-validation process of the mediational strategies of using dictionary or offering translation, determining the intention of the speaker, adding up details to infer logical conclusions and providing the correct response and explanation is depicted in Excerpt 3 extracted from LT2, enrichment phase (week4) (Figure 5). It shows what interpretation of learners' abilities motivated the mediator to make use of these strategies so as to diagnose the sources of difficulties learners encountered in comprehending L2 listening texts and hence to promote their understanding.

Encouraging learners to use dictionary or offering translation

The learner's silence in line 2 (*initial response*) led the mediator to offer the strategy of *replaying*, beginning the procedure of offering support with implicit mediation as prescribed in DA. With the aid of this mediation, the learners could recall a part of the sentence heard verbatim, which was the idiom "you can say that again" (line 4). The learner's response here showed that although this mediation helped learners to provide a correct recall of the idiom heard, they were oblivious of its figurative meaning. This response signified that a more explicit prompt was required on the part of the learners. Upon observing this, the mediator made a *tentative interpretation* that this lexical item was absent in the lexical repertoire of the learners, and therefore chose to offer the strategy of *dictionary use* (line 11). The mediator assumed using dictionaries could serve as a symbolic artifact that had the capability of triggering the learning process of the learners which might bring about the expansion of their ZPDs (Vygotsky, 1978). S1's remarks in lines 12 and 14 revealed she had problems both with the meaning of the word and with dictionary-use. Following the provision of a dictionary-use skill on how to find idioms in dictionaries (lines 15 and 17), one of the students looked up the idiom in her dictionary and informed the class of its meaning (line 18). This *responsiveness to mediation* brings in evidence in support of the *acceptance* of the *tentative interpretation* the mediator made about this ability in the learners.

Asking learners to determine the intention of the speaker

The silence of the learners (line 32) revealed for the mediator once more that they could only grasp the literal meaning of the material heard and were not cognizant of the hidden intention of the speaker. Making the *provisional interpretation* that the learners had difficulty in extracting meaning beyond the literal meaning of words and sentences, she implemented the strategy of *determining the real intention of the speaker* (line 33). This strategy aimed to help learners discover what the speaker was saying beyond her words. The *learners' response* in line 34 reveals this strategy, however, was not sufficient to elicit a correct response from the learners.

Encouraging learners to add up details to infer logical conclusions

Upon observing the insufficiency of the mediation provided, the mediator deemed to offer the strategy of *adding up details to infer logical conclusions* which is more explicit. This strategy encourages learners to search for clues, whether linguistic or extralinguistic (Buck, 2001; Nunan, 2002; Vandergrift, 2004), and to piece them up in order to find a logical conclusion not expressed in words. Through this strategy and with the help of the strategies of offering metalinguistic clues and accepting response (lines 35 to 48), the mediator helped the learners unearth some facts about the conversation, e.g. "a loan," "a disagreement." However, she was once again faced with the learners' partial understanding and inability to grasp the intended meaning (line 49 and 51).

Providing the correct response and explanation

The inability on the part of the learners (line 49 and 51) led the mediator to the provisional interpretation that the required skills for making an understanding of that implied meaning were



beyond the learners' ZPD. The mediator, thus, decided to make use of the most explicit and the last strategy on the list, *providing correct response*.... This strategy reflects Vygotsky's (1987) notion about providing instruction. Vygotsky believes this could be the driving motor of intellectual development when it is attuned to learners' abilities and can help them move up to a higher level of ZPD. To implement this strategy, the mediator clearly explained to the learners how to arrive at the speaker's implicit intent. In so doing, she clarified the learners should focus on important details they heard and combine them with their own background knowledge and experiences to infer hidden meanings (lines 52 and 54). This interaction showed that learners' responses to each mediation were carefully screened by the mediator with the aim of determining the next appropriate mediational support. Moreover, the continuous process of providing mediational strategies to the learners gave the mediator a clearer picture of the learners' development. As observed in Excerpts 1, 2, and 3, the mediational strategies provided did not always result in eliciting the intended response and therefore did not always yield a positive outcome. Observing this, one might conclude these strategies are not effective in promoting the learners' listening comprehension ability, or the learners are far from independently controlling that L2 feature. However, as mentioned above, to determine the appropriateness of a mediational strategy not just a particular exchange but the entire G-DA procedure should be analyzed. The analysis, which is explained in the next part, pertains to validation at a macro-level. This scrutinizes the learners' performance over the entire G-DA procedure to evaluate the claims regarding the learners' development and their ZPDs.

Macro-Validation of the G-DA Procedures

The second research question aimed to find evidence for the macro-validity of the G-DA interactions. To address the question, the expanded model for macro-validation was implemented. This model demands four forms of evidence: (i) *changes in the quality of mediation required*, (ii) *changes in learners' responsiveness to mediation*, (iii) *learners' verbalization about their performance*, and (iv) *learners' independent performance*.

Changes in the quality of mediation

From an SCT perspective, one way to track learners' progress in the ZPD is to refer to the number and the degree of the explicitness of the mediation offered to the learners over time (Aljaafreh & Lantolf, 1994; Poehner, 2005 & 2011). Consequently, to provide evidence whether changes in the quality and quantity of the mediation offered occurred, the frequency and degree of explicitness of the mediational strategies offered over time were analyzed. Table 1 presents the frequency of mediational strategies offered to the learners in the dynamic administration of the pretest and posttest.

Table 1

Frequency and the degree of explicitness of mediational strategies

Type of mediational strategy	Degree of explicitness	Pretest	Post test
Accepting response	least explicit	98	62
Rejecting response		84	32
Replaying → the entire portion		46	50
↙ a segment of the portion			
Asking the words and putting them together		#2 19 #3 13 #4 9 #5 5	#2 30 #3 18 #4 2 #5 0
Offering metalinguistic clues		98	23
Encouraging learners to use dictionary or		#2 44 #3 25 #4 16 #5 13	#2 12 #3 11 #4 0 #5 0
Asking learners to determine the intention of		47	8
Encouraging learners to add up details to		55	9
Providing the correct response and explanation	most explicit	37	10
		25	6
		18	4
		15	3
		Total 532	Total 207

Note: #2= second replay, #3= third replay, #4= fourth replay, and #5= fifth replay and more



A comparison of the frequency of the mediational strategies offered in the pretest and posttest shown in Table 1 reveals that at the end of the G-DA procedure (posttest) the learners needed fewer explicit mediational strategies and demanded implicit mediation. This change in the type of mediation required is an indication of the learners' growing autonomy and self-regulation functioning (Poehner, 2008a). Moreover, this observed change offers evidence in support of the macro-validation of the strategies used in this G-DA procedure as well as the G-DA interactions taken place. Another form of evidence that had to be detected in the mediator-learners interaction to evaluate the macro-validation of the G-DA procedure implemented in this study was to observe the overall pattern of *learners' responsiveness to mediation* and how it may have changed during this procedure.

Changes in learners' responsiveness to mediation

In G-DA, like DA, learners' development may also manifest as changes in responsiveness to mediation (Aljaafreh & Lantolf, 1994, Lantolf & Poehner, 2013). Put differently, learners' lack of responsiveness to a type of mediational strategy may be interpreted to mean the learners' understanding of that relevant feature of the L2 was far from where it needed to be for successful independent performance. However, over time during the G-DA interaction this same mediational strategy might result in eliciting the desired response from the learners. This change in responsiveness to mediation on the part of the learners signals the relevant ability is in the process of maturing in the learners (Poehner, 2011).

To illustrate how learners' responsiveness to a specific mediation may change during an interaction, let us return to Excerpt 2 (Figure 5). As shown in lines 2 and 4 of the excerpt, at the beginning of the interaction, the learners were not responsive to the mediational strategy of *replaying*. However, later on in this interaction and after a continued course of providing appropriate mediational moves, as line 19 of the excerpt depicts the learners were responsive to this strategy. It is worthy of notice that on the basis of this single occurrence of change in the learners' responsiveness to the mediation observed in this interaction, it seems untimely to claim this interaction resulted in the development of learners' listening comprehension ability. However, each single change in learners' responsiveness to mediation is assumed to have a small contribution to the maturation of their intended abilities (Poehner, 2008a).

In this paper, the length constraints preclude the transcription of all interactions in which changes in learners' responsiveness to mediational support were observed. However, in Table 2, the effectiveness of each mediational strategy in eliciting the desired response in the pretest and posttest during this G-DA procedure is provided in percentages.

Table 2

The effectiveness of mediational strategies in eliciting a correct response

Type of mediational strategy	Pretest			Posttest		
	Total <i>f</i>	efficient %	inefficient %	Total <i>f</i>	efficient %	inefficient %
Accepting response	98	40	60	62	63	37
Rejecting response	84	24	76	32	84	16
Replaying	144	35	65	73	82	18
Asking the words and putting them together	47	55	45	8	87	13
Offering metalinguistic clues	55	48	52	9	78	22
Encouraging learners to use dictionary or offering...	37	70	30	10	80	20
Asking learners to determine the intention of the...	25	72	28	6	67	33
Encouraging learners to add up details to infer logical...	18	73	27	4	75	25
Providing the correct response and explanation	15	100	0	3	100	0
Total	532			Total	207	



As Table 2 shows, in the pretest the strategies of *accepting response*, *rejecting response*, and *replaying* were found to be less useful to the learners in eliciting the correct response. However, in the posttest these three mediational strategies seemed to be more useful to the learners. As noted, these strategies are more implicit as compared to the other mediational strategies used in this study. This change in the learners' responsiveness, i.e. benefiting from implicit mediation could be interpreted as an indication that they are relatively close to autonomy and self-regulation (Lantolf & Poehner, 2013), this could provide evidence towards the macro-validation of the G-DA interactions of this study. The third piece of evidence which could contribute to the macro-validation of DA is *learners' verbalization about their performance*.

Learners' verbalization about their performance

The learners' reflection on their performance and their reasons for choosing particular linguistic items to produce or comprehend material in L2 is known as the *learners' verbalization about their performance*. According to Poehner (2011) the information a mediator gains through this source is useful in both ascertaining the appropriateness of learners' understanding of the language and how they are using this understanding to produce and comprehend materials in that language. In DA, learners' verbalizations of their understanding of a particular linguistic feature seem to be crucial to the mediator's diagnosis of the type of mediation suitable for the need of the learners as well as an aid to determine the abilities of the learners (Poehner, 2008b). Throughout the G-DA procedure the verbalization of the learners about what was impeding their understanding and what led them to have a particular understanding of the material heard provided the mediator with evidence to either refute or accept the interpretation she had made about a particular diagnosis of the learners' ZPD. The learners' verbalization data collected in this study was the verbalization occurred incidentally in class.

Excerpt 3, (Figure 5) captures the importance of *learners' verbalization* in mediator's diagnosis of learners' abilities. S1's verbalization about her problem clarified for the mediator that not only was the idiom of "you can say that again" absent in the lexical repertoire of the learners, they also did not know how to search for an idiom in a dictionary revealing their weaknesses in dictionary use. Another example is in line 40 of this excerpt; had the learner not mentioned her lack of understanding of the "rule" stated in the text heard, the mediator would have assumed the learners had grasped the meaning of the portion. Although the excerpt indicates the learners could eventually recite what they heard in the portion and were aware the interlocutors were disagreeing about something, they could not discern the main intent of the interlocutors. The Learners' verbalizations about their performance created a clearer picture of the learners' abilities during G-DA.

Learners' independent performance

Another piece of evidence that can reveal whether the entire G-DA procedure could improve learners' L2 listening comprehension could be shown through analyzing their independent performance on the pretest and posttest.

Table 3

Descriptive statistics of the experimental and control groups on the pretest and posttest

		Mean	SD	95% confidence interval for mean		Kolmogorov-Smirnov		
				Lower	Upper	statistic	df	p
Experimental (N=24)	Pretest	36.00	1.91	36.19	36.81	0.12	24	0.20
	Posttest	41.75	3.08	40.45	43.05	0.14	24	0.20
Control (N=26)	Pretest	36.07	1.92	35.27	36.89	0.11	24	0.20
	Posttest	36.46	1.92	35.77	37.15	0.11	24	0.20

Note: pretest and posttest maximum possible score=30



Descriptive statistics regarding the non-dynamic pretest and posttest of the experimental and control groups are presented in Table 3. The preliminary assumption testing was conducted to check for normality (non-significant results of Kolmogorov-Smirnov statistic) and homogeneity of variance with no serious violations was noted.

Table 4

T-test results on the experimental and control groups' performance on the pretest and posttest

	<i>t</i>	<i>df</i>	<i>p</i>	95% confidence interval of the difference		<i>Eta squared</i>
				Lower	Upper	
Pretest Experimental × Pretest Control	0.14	48	0.89	-1.01	-1.17	-
Pretest Experimental × Posttest Experimental	-13.90	23	0.00	-6.60	-4.89	0.86
Pretest Control × Posttest Control	-1.79	25	0.09	-0.83	0.06	-
Posttest Experimental × Posttest Control	-7.45	48	0.00	-6.73	-3.85	0.54

As depicted in Table 4, there was no significant difference in the listening comprehension ability of both groups at the outset of the tests [$t(48)=0.14$, $p>0.05$]. The listening comprehension performance of experimental group increased from the pretest to posttest [$t(23)=-13.90$, $p<0.05$; and $t(48)=-7.45$, $p<0.05$, both with large effect sizes, $\eta^2>0.14$], whereas, from the pretest to posttest, the performance of the control group did not exhibit any change [$t(25)=-1.79$, $p>0.05$]. The improved scores of the learners of the experimental group on the posttest could be attributed to the mediation offered during the G-DA procedure. Therefore, it appears the mediator was successful in co-constructing a group ZPD in the classroom settings and calibrated the mediational support to the ZPD of the group as a whole. Moreover, it seems that by providing mediation within the group, the mediator could bring about the L2 listening comprehension development of each individual learner.

Finally, further analysis of the mediator-learner dialogue indicated the causes of L2 listening comprehension difficulty in this study were mostly due to lack of L2 lexical knowledge and issues related to implied meanings. The L2 idiomatic expressions were often unknown to the learners. This might not be surprising since perhaps the learners are still in the process of building their lexical knowledge. Moreover, sometimes phonologically and grammatically-rooted problems were observed to impede the learners' listening comprehension. Towards the end of the G-DA procedure and especially on the dynamic posttest it was seen the strategies of using dictionary (aiming at remediating L2 lexical deficiency) and the strategies of determining the intention.... and adding up details to... (encouraging learners to extract meanings beyond literal ones) were demanded much less by the learners (see Table 1). This was interpreted by the mediator as the sign of success of the G-DA procedure in remediating learners' L2 listening comprehension difficulties which brings in further evidence for the macro-validation of the G-DA procedures.

Conclusion

This paper aimed to evaluate the validity of G-DA, grounded in SCT which holds the psychological abilities arise as a result of participation in activities in which mediation is provided by others and by artifacts made available through culture. As noted earlier, L2 DA unlike standardized tests is less interested in measuring learners' consistencies in L2 performance; rather it seeks to measure learners' developments. Nevertheless, this does not obfuscate the need to validate the claims made through DA about learner abilities and their development. Two research questions were posed in this study to explore the validity of the G-DA procedure carried out to enhance the listening comprehension ability of EFL intermediate



learners. In order to provide a response for the first research question which concerned the detection of evidence for the micro-validity of the G-DA procedure implemented in this study, Poehner's (2011) micro-validation model was used. Based on this model, through a careful analysis of the mediator-learners interactions, evidence supporting the appropriateness of the interpretations made about learners' abilities and mediational strategies offered to enhance these abilities were detected and gathered. The evidence detected for micro-validation revealed the abilities that were assessed in G-DA were always changing. The mediator, thus, had to have non-stop interpretations of learner abilities in order to offer the appropriate type of mediational strategy to the learners. This finding runs in accordance with Poehner's (2008b) claims about validity in DA.

To answer the second research question, aiming at detecting evidence to evaluate macro-validation of the G-DA executed in this study, Poehner's (2011) macro-validation model was utilized. The analysis of the G-DA interactions of this study demonstrated the learners required less explicit type of mediational support (i.e. changes in the quality of mediation required) towards the end of the G-DA procedure. Moreover, the more implicit mediational strategies which were not effective to elicit a correct response from the learners at the beginning of the G-DA procedure were found to be useful for the learners to provide a correct response (i.e. changes in learners' responsiveness) towards the end of this procedure. The observation of these two types of changes in learners' G-DA performance over time could signify the G-DA procedure resulted in enhancing L2 listening comprehension in the learners, which in turn offers evidence in support of the macro-validation of the G-DA procedure. Furthermore, the analysis of the mediator-learners interactions occurred in this study revealed learners' verbalisations about their L2 performance helped the mediator to have a more accurate diagnosis of the abilities of the learners, and thus providing evidence in support of the macro-validation. The independent performance of the learners on the posttest brought in further evidence that each individual group member as well as the group had progress in her L2 listening comprehension most probably as a result of the G-DA interactions.

In general, this study demonstrated the usefulness of Poehner's (2011) evidence-based arguments in validating G-DA procedure. This evidentiary reasoning process made it possible, in this study, to examine assumptions pertaining to different types of evidence and to address various aspects related to the validity of G-DA procedure. The present study suggests teachers or mediators can apply this coherent set of procedures to guide them in formulating and justifying interpretations about learners' abilities and their possible development through DA procedures. Moreover, as illustrated in this paper, by sensitizing teachers to learner development, G-DA yields more systematic interactions between the teacher and learners which are calibrated to the group of learners' emergent abilities. Therefore, by recognizing the learning process and needs of the learners through G-DA, teachers can more effectively design mediational strategies to remove learners' L2 problems and to improve their instruction. Nevertheless, further research on the implementation of DA in the classroom context is required to determine how a profile about the developmental changes of each individual in a group can be created. Studies addressing the question of what impact the mediation provided to primary interactants could have on the secondary interactants who happen to be the primary ones in later interactions in a group setting seems also to be necessary. Finally, there is also a need for G-DA validation studies analyzing micro- and macro-validity in interventionist DA across multiple instructors and groups.

References

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



- 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.
- Bachman, L. F. (2000). Learner-directed assessment in ESL. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. ix-xii). New Jersey: Lawrence Erlbaum Associates, Inc.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Baird, J., Andrich, D., Hopfenbeck, T., & Stobart, C. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24, 317-350.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
- Buck, G., (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Chapelle, C. A., Enright, M. E., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. London: Routledge.
- Cheng, L. (2005). *Changing Language Teaching through Language Testing: A Washback Study*. Cambridge: Cambridge University Press.
- Cheng, L., Watanabe, Y., & Curtis A. (Eds.) (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Lawrence Erlbaum.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement*, 2nd ed. (pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, DC: American Council on Education.
- Davin, K. J., Herazo, J. D., & Sagre, A. (2017). Learning to mediate: Teacher appropriation of dynamic assessment. *Language Teaching Research*, 21(5), 632-651.
- Ebel, R. (1961). Must all tests be valid? *American Psychologist*, 16, 640-647.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. London & New York: Routledge.
- Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *Research Notes Quarterly*, 12, 8-10.
- Guion, R. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement*, 1, 1-10.
- Hanreddy, J. & Whalley, E. (2008). *Mosaic 1: Listening and speaking*. Maidenhead, UK: McGraw Hill.
- Hashemi Shahraki, S., Ketabi, S., & Barati, H. (2015). Dynamic assessment in EFL classrooms: Assessing listening comprehension in three proficiency levels. *International Journal of Research Studies in Education*, 4(3), 73- 89
- Haywood, H.C., & Lidz, C.S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge: Cambridge University Press.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage Publications.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.



- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527–535.
- Kane M. (2012). Validating score interpretations and uses. *Language Testing*, 29(1) 3–17.
- Kane, M. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50 (1), 1-73.
- Kane, M. (2017). Loosening psychometric constraints on educational assessments. *Assessment in Education: Principles, Policy & Practice*, 24, 447–453.
- Kane, M. T. (2021). Articulating a validity argument. In *The Routledge handbook of language testing* (pp. 32-47). Routledge.
- Kane, M. T., and Wools., S. (2019). “Perspectives on the validity of classroom assessments,” in *Classroom Assessment and educational measurement*. Editors S. M. Brookhart, and J. H. McMillan (New York, NY: Routledge), p. 11–26.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: bringing the past into the future. *Journal of Applied Linguistics*, 1(2), 49-72.
- Lantolf, J. P., & Poehner, M. E. (2007). Dynamic Assessment. In *Encyclopedia of Language and Education Volume 7. Language Testing and Assessment* (E. Shohamy, Ed., N. Hornberger, Gen. Ed.). Springer Publishing.
- Lantolf, J. P., & Poehner, M. E. (2013). The unfairness of equal treatment: objectivity in L2 testing and dynamic assessment, *Educational Research and Evaluation: An International Journal on Theory and Practice*, 19:2-3, 141-157
- Mackey, A., Gass, S. (2005). *Second language research methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S. A. (1989). Validity. In R. L. Linn (ed.), *Educational measurement*. 3rd Ed. New York: American Council on Education. 13-103.
- Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R. W. Lissitz (Ed.), *The concept of validity: Revisions new directions and applications* (pp. 83–108). Charlotte, NC: Information Age.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice* 22 (4) 13-25.
- Nunan, D. (2002). Listening in language learning. In J. C. Richards & W. A. Renandya, (Eds.). *Methodology in language teaching: An anthology of current practice* (pp. 238-241). Cambridge: Cambridge University Press
- Petrovsky, A. V. (1985). *Studies in psychology: The collective and the Individual*. Moscow: Progress.
- Poehner, M. E. (2005) Dynamic assessment of advanced L2 learners of French. Ph.D. dissertation. Penn State University.
- Poehner, M. E. (2008a). Dynamic assessment. A Vygotskian approach to understanding and promoting L2 development. Berlin, Germany: Springer.
- Poehner, M. E. (2008b). *Dynamic Assessment and the Problem of Validity in the L2 Classroom*. (CALPER Working Paper Series, No. 10). The Pennsylvania State University: Center for Advanced Language Proficiency Education and Research.
- Poehner, M.E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, 43, 471–491.
- Poehner, M. E. (2011). Validity and interaction in the ZPD: Interpreting learner development through L2 Dynamic assessment. *International Journal of Applied Linguistics*, 21, 244–26



- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic Testing*. New York: Cambridge University Press.
- Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and Learning, 2nd edition*. London: Sage.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing* 17 (2), 163–84.
- Torrance, H. (1995). Teacher involvement in new approaches to assessment. In H. Torrance (ed.), *Evaluating authentic assessment*. Buckingham: Open University Press. 44-56.
- Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, 51(2), 257–280.
- Vandergrift, L. (2004). Learning to listen or listening to learn. *Annual review of applied linguistics*, 24, 3-25.
- Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. (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. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (eds.), *The collected works of L. S. Vygotsky*. Vol. 1. Problems of general psychology (pp.39-285). New York: Plenum.
- Vygotsky, L. S. (1998). The problem of age. In R. W. Rieber (Ed.), *Child psychology, Vol. 5. Collected works of L. S. Vygotsky* (pp. 187– 205). New York: Plenum. (Original work published 1932–1934).
- Warford, M. K. (2010). The zone of proximal teacher development. *Teaching and Teacher Education*, 27 (2), 252-258.

Biodata

Dr. Sara Hashemi Shahraki, is an assistant professor at Islamic Azad University, Najafabad Branch (IAUN). She earned her PhD and Master's degrees in TEFL at University of Isfahan with receiving the top student certificate and award for both degrees from the university chancellor. As the faculty member of IAUN, many MA students have received their degrees under her supervision. Dr. Sara Hashemi Shahraki has presented many articles in international and national conference and has published many articles in domestic and foreign scientific journals. Her research interest fields are in teaching, testing, and classroom assessment.
Email: sara_m_hashemi@yahoo.com

