



Journal of Teaching Language Skills
(JTLS)

39(3.1), Fall 2020, pp. 217-252

DOI: 10.22099/jtls.2020.38116.2865

Online ISSN: 2717-1604

Print ISSN: 2008-8191

Research Paper

Improving EFL Learners' Performance on Receptive-Response and Productive-Response Listening Comprehension Items through Diagnostic vs. Dynamic Assessment

Ghazaleh Zandi *

Kobra Tavassoli **

Fateme Nikmard ***

Department of ELT, Karaj Branch, Islamic Azad University, Karaj, Iran

Abstract

Listening comprehension is a challenging skill in EFL contexts and it is necessary to research various aspects of this skill. Therefore, this study tried to investigate EFL learners' progress on two types of listening comprehension items, receptive-response and productive-response, through diagnostic assessment (DIA) and dynamic assessment (DA). To do this, a Nelson proficiency test was administered among 120 EFL students in six classes, out of which, 90 students whose scores were in the acceptable range were selected. The classes were then randomly divided into three groups of control, DIA, and DA. As the pretest, a KET listening test in two sections of receptive-response and productive-response items was used to check the learners' initial performance. Next, as for the treatment, in the DIA group, the learners took four listening tests with receptive-response and productive-response items in the form of test-feedback; in the DA group, the learners took the same four tests in the form of test-mediation-retest, and in the control group, the learners practiced receptive-response and productive-response items after each listening activity. At the end of the study, another KET listening test in two sections of receptive-response

Received: 11/09/2020 Accepted: 12/11/2020

* M.A, Email: ghazalehzandi89@gmail.com

** Assistant Professor, Email: kobra.tavassoli@kiaou.ac.ir, Corresponding Author

*** Ph.D. Candidate, Email: fateme.nikmard@kiaou.ac.ir

and productive-response items was administered to the participants to investigate the three groups' progress from pretest to posttest. Finally, two repeated-measures two-way ANOVAs and one MANOVA were run, and the results showed that both DIA and DA resulted in significant improvements in the learners' performance on both receptive-response and productive-response listening items; however, the difference between the two types of assessment was not significant.

Keywords: Diagnostic Assessment, Dynamic Assessment, Listening Comprehension, Productive-Response Items, Receptive-Response Items

Listening comprehension has a critical role in learning a second/foreign language, and many EFL learners consider it as the most demanding language skill to develop (Li, 2019). A major problem most EFL learners come across in listening to the new language is understanding the input produced by other speakers whether in face-to-face interactions or when listening to audio files. Not being proficient in the listening skill might be one of the main reasons why many EFL learners do not participate orally in language classes (Hayati & Jalilifar, 2009). One of the major reasons for this lack of proficiency is that listening comprehension is not receiving due attention in language classes. Therefore, it is crucial to identify certain techniques that can be used willingly by both teachers and learners when dealing with listening comprehension in EFL classes to overcome this problem.

Listening comprehension has also been ignored when considering its assessment since it presents unique challenges to both teachers and researchers (Wagner, 2014). However, due to the importance of listening comprehension in providing input to language learners, both its teaching and testing need to be investigated more. In fact, instruction and assessment are two sides of the same coin and closely interconnected with each other (Lantolf & Poehner, 2008), thus, it is not possible to draw a demarcation line between them. As Brown and Abeywickrama (2010) mentioned assessment is an

ongoing process and a good teacher never stops assessing learners in the classroom. They further argue that assessment is part of teaching and covers a great deal of it. In fact, through assessment, teachers can identify how to provide better instruction to students to improve their language ability.

Although listening comprehension was the subject of some recent studies (e.g., Cheng & Matthews, 2018; Dai & Roever, 2019; Li, 2019; Rukthong & Brunfaut, 2020), to date, the interconnection of instruction and assessment of this skill and how its assessment can enhance its instruction in EFL classes was not investigated enough. Thus, this study was an attempt to make a connection between teaching and testing listening by using two popular assessment techniques, diagnostic assessment (DIA) and dynamic assessment (DA), to enhance EFL learners' listening comprehension ability since both of these assessment techniques focus on the close connection between teachers and learners. While, the main goal of DIA is to find out the strengths and weaknesses in the learners' knowledge of the language and providing them with appropriate feedback to help them improve their abilities (Jang & Wagner, 2014), the main purpose of DA is to ensure the learners can successfully react to the used mediations to overcome their weaknesses (Poehner, 2014).

Considering the importance of listening comprehension in learning a new language (Rost, 1994), the purpose of this study was to investigate how EFL learners' performance on two types of listening comprehension items, that is, receptive-response and productive-response items, could be improved through using diagnostic vs. dynamic assessment in the classroom. The study findings can provide insights for EFL teachers about how to link instruction and assessment in EFL classes to improve one of the most neglected aspects of language instruction, which is the learners' listening comprehension ability. Accordingly, the following research questions were posed:

1. Does diagnostic assessment have any significant effect on EFL learners' progress on receptive-response listening comprehension items?
2. Does diagnostic assessment have any significant effect on EFL learners' progress on productive-response listening comprehension items?
3. Does dynamic assessment (its interventionist model) have any significant effect on EFL learners' progress on receptive-response listening comprehension items?
4. Does dynamic assessment (its interventionist model) have any significant effect on EFL learners' progress on productive-response listening comprehension items?
5. Is there any significant difference between the effect of diagnostic vs. dynamic assessment on EFL learners' progress on receptive-response and productive-response listening comprehension items?

Review of the Related Literature

Assessment is an essential part of the teaching and learning process in which the performance and progress of learners are measured; however, this measure should not be expressed only by a single grade or mark (Brown & Abeywickrama, 2010). Assessment is an ongoing process in which a variety of techniques can be used to evaluate the learners' progress throughout the course (Brown & Abeywickrama, 2010). In fact, assessment is being used in language classes every day. Some of the popular assessment types used in different educational contexts are diagnostic, dynamic, performance, portfolio, task-based, and self and peer-assessment. Out of these assessment types, two types with strong theoretical backgrounds in connecting teachers and learners more in the classroom, that is diagnostic and dynamic assessment, were selected for more investigation in this study.

Diagnostic Assessment (DIA)

Diagnostic assessment refers to assessing the learners' strengths and weaknesses (Alderson, Haapakangas, Huhta, Nieminen, & Ullakonoja, 2014); however, it focuses more on weaknesses rather than strengths. The purpose is to use such information to guide future instruction and consequently improve the students' learning (Jang & Wagner, 2014). DIA is done either at the beginning of the course to provide information about further instruction or during the course based on the instruction which was used in the classroom. Through DIA, it is tried to identify those areas in which a student needs more help. This diagnosis can be made relatively generally by showing, for example, whether a student needs help with a main language skill, or it can be more specific by looking for identifying weaknesses in a student's use of grammar (Alderson, Brunfaut, & Harding, 2015). As Jang and Wagner (2014) stated DIA focuses on assessing the differences between a student's actual ability and short- and long-term potential development through guiding them and providing them with various types of feedback. In fact, one of the determining features of diagnostic assessment is to utilize different types and delivery modes of feedback to enhance students' learning. "Feedback, in general, is conceptualized as information provided for learners following an assessment task regarding positive aspects and areas for improvement in their performance or understanding" (Jang & Wagner, 2014, p. 158).

There are numerous types of feedback that can be used in DIA. For example, teachers might offer students oral or written comments, or student-teacher interactions might be used. The feedback can be provided in direct or indirect form. In direct feedback, the correct form of a mistake is provided to the learner. On the other hand, in indirect feedback, the correct form is not presented and teachers regularly use codes to highlight the linguistic errors in the students' writing or speaking, hoping that their comments will assist the

students in correcting their own mistakes (Harding, Alderson, & Brunfaut, 2015; Jang & Wagner, 2014). In other words, in DIA, teachers offer feedback to students about what is correct or incorrect in their responses, and about their strengths and weaknesses. This information will then be used by the students to make the necessary changes (Nicol & Macfarlane-Dick, 2006).

Because of its usefulness in educational contexts, DIA has been investigated in recent years. For instance, Granfeldt and Ågren (2014) investigated the Direkt Profil as a diagnostic assessment tool to check its effectiveness. It was indicated that Direkt Profil provided fast and detailed feedback about how particular types of correct or incorrect linguistic structures were related to various stages of the learners' development. Such a tool can be quite useful for teachers in different classes to identify their students' present levels and to help them move to the next level. Ranjbaran and Alavi (2016) conducted research on cognitive diagnostic assessment where they followed a cognitive framework to develop a test of reading comprehension, which could be used for diagnostic purposes. Their test was validated with 1986 general English students and they used the results to provide exact learners' skill mastery profiles. The test turned out to be a useful tool for diagnostic purposes, which can be used in different EFL classes to boost the students' current level of performance. In a study similar to this research, Nikmard and Tavassoli (2020) investigated the impact of diagnostic assessment on selective and productive reading comprehension tasks among EFL learners. Parallel with the literature on DIA, they found that diagnostic assessment improved the learners' performance on both kinds of reading tasks, showing the positive impact of DIA in EFL contexts. The findings of these studies show the positive impact DIA can have on the learners' progress throughout a course.

Dynamic Assessment (DA)

Dynamic assessment is a cooperative approach in assessment that concentrates on the learners' ability to respond to intervention (Haywood & Lidz, 2006). According to Elliott (2003), the strength of DA is that in contrast to traditional static tests, it leads to better predictions of students' future educational performance. DA was proposed as a supplement to static assessment because it measures developing skills (Jeltova, Birney, Fredine, Jarvin, Sternberg, & Grigorenko, 2007). It is believed that DA can increase fairness and accuracy in the assessment of talents because it focuses on the ability to analyze the language. The theoretical background of DA is rooted in Vygotsky's sociocultural theory and his focus on the concept of the zone of proximal development (ZPD). ZPD refers to the gap between the learners' ability to do a task individually and their ability to do a task with the help of a more competent person (Gibbons, 2003), which shows the significant role of teachers in this technique.

“In DA, assessment and instruction are a single activity that seeks to simultaneously diagnose and promote learner development by offering learners mediation, a qualitatively different form of support from feedback. Mediation is provided during the assessment procedure and is intended to bring to light underlying problems and help learners overcome them” (Lantolf & Poehner, 2008, p. 1). There are different models of DA, but the most popular one is the classification of DA into interventionist and interactionist models. According to Lantolf and Poehner (2004), the interventionist model generally works with a quantitative approach and is more related to a psychometric orientation. In fact, in interventionist DA, mediation is standardized, therefore, it allows more use of quantification and inferential statistics (Lantolf & Poehner, 2008). On the other hand, the interactionist model is based on the qualitative assessment of mental processes, which was

avored by Vygotsky (Lantolf & Poehner, 2004). Since the interventionist DA model entails using standard administration procedures to reach desired results and is used to compare groups and individuals, it was selected to be used in this study. In this model of DA, mediation is not offered during the assessment, rather, it is introduced in a training phase between a traditional pre-test and post-test (Lantolf & Poehner, 2008), which is referred to as test-mediation-retest.

Due to its strong theoretical background in educational psychology and its claim to improve students' learning successfully, DA has attracted the attention of many researchers. Poehner, Zhang, and Lu (2015) studied computerized dynamic assessment believing that mediation in DA is an integral part of the learners' language development. In their study, they provided the learners with mediated scores, through which the amount of support the learners needed was also provided. These mediated scores, which were automatically produced by the tests, were an influential diagnosis of the procedure of the learners' second language development, which offered information for subsequent teaching and learning. In Iran too, DA has been studied in recent years. The impact of DA on improving EFL learners' vocabulary knowledge was studied by Sarani and Izadi (2016). Their major finding was that mediated scores, which were gained by helping students through DA, should be used to find out where learners have problems and need help to improve their abilities. Further, the effect of interactionist and interventionist group dynamic assessment (GDA) on EFL learners' listening comprehension was examined by Ahmadi Safa and Beheshti (2018), who found the more significant influence of interactionist GDA on enhancing EFL learners' listening comprehension. In recent research on DA, Tavassoli and Nikmard (2019) also investigated the influence of dynamic assessment on selective and productive reading comprehension tasks among EFL learners.

They concluded that the learners' performance was significantly improved in both kinds of tasks, showing the usefulness of DA. Similar to what is stated in the literature about the significant role of DA on improving different areas of students' knowledge, these results showed the significance of DA in EFL classes.

Listening

Normally, people do more listening than speaking in their everyday life; listening also provides the input people need to have a successful language acquisition (Brown & Abeywickrama, 2010). Therefore, more attention should be paid to listening in teaching/learning contexts. Listening comprehension is viewed theoretically as a lively process wherein each person focuses on selected factors of aural input and discourse meaning, and relate what they hear to existing knowledge. Listening is an important language skill that is unfortunately unnoticed in many EFL classes due to the practical complexities and challenges of providing spoken language to students. One way to encourage teachers to focus on listening in their classes is through assessing it regularly (Buck, 2001).

However, assessment of L2 listening presents unique challenges to teachers and even test developers. In fact, assessing listening is technically more complicated than assessing other language skills, and that is most probably why it has been somewhat neglected in the literature (Wagner, 2014). Nevertheless, because of the unique aspects of listening and the information assessment of this skill provides to teachers and testers about the students' ability, it should not be overlooked. Perhaps, the most noticeable problem in assessing listening is that listening is an internal process, which cannot be observed. Therefore, in the assessment of listening, appropriate tasks and activities should be used for students to respond to. Based on the

students' responses, teachers would be able to make inferences about their level of listening ability and make any necessary adaptations in their instruction (Wagner, 2014).

In line with the need to do research on listening comprehension to get more insight into its various aspects, different studies have been conducted on how listening comprehension can be enhanced in EFL contexts. For example, Winke and Lim (2017) studied the impact of listening test preparation on listening test performance. The results showed that neither receiving explicit test-taking strategies nor taking practice tests had any significant impact on the participants' test performance. In fact, the only use of the test preparation sessions was the participants' familiarity with the test. This shows that to improve the learners' listening comprehension ability, it should actually be taught and practiced in language classes instead of only focusing on the test. Consistent with this, Saito and Akiyama (2018) in a longitudinal study found out the positive impact of using video-based interactions between native and non-native speakers on Japanese learners' development of listening comprehension ability. In addition, the results of Cheng and Matthews' (2018) study showed that performance on listening tests was correlated with and could be predicted from test-takers' performance on different vocabulary tests. In fact, it was found that the two constructs of listening and vocabulary were closely inter-related with each other and it is not possible to ignore either of them in EFL classes. In other words, to improve the learners' listening comprehension, their vocabulary knowledge needs to be enhanced too. The results of the studies mentioned here show the usefulness of a variety of techniques to enhance learners' listening ability in L2 classes. This signifies the vital role of doing more research in this area.

As it was mentioned above, teachers should assess their students' listening in their classes to attain information about their actual level and the

problems they might have. To use assessment successfully for instructional purposes, it is crucial to use a variety of assessment activities in the class. There are numerous assessment activities mentioned in the literature (e.g., Alderson, 2000; Brown, 2005; Brown & Abeywickrama, 2010; Buck, 2001; Luoma, 2004; Weigle, 2002), which can be used for such a purpose. Two of the most popular types of assessment activities are receptive-response and productive-response items, which were introduced by Brown (2005), and focused on in different studies, such as Esfandiari and Tavassoli (forthcoming), Nikmard and Tavassoli (2020), Zareinajad, Rezaei, and Shokrpour (2015), and Zhang (2017). In receptive-response items, students are required to scan the message for specific information and to choose the answer from the available options. Various types of receptive-response items are true-false, multiple-choice, matching, ordering, listening cloze, and information transfer (Brown, 2005; Brown & Abeywickrama, 2010). On the other hand, in productive-response items, students are required to produce the answer in either spoken or written form. The popular types of productive-response items are fill-in-the-blanks and short-responses (Brown, 2005). Further, in receptive-response items, the focus is more on the formal aspects of the language including the lexicon and grammar, whereas, in productive-response items, the focus is often on both form and meaning with more emphasis on meaning (Brown & Abeywickrama, 2010). Since this classification of assessment items is similar to the classification of language skills into receptive and productive skills, is quite practical to be used in language classes, and has been studied by different researchers, it was selected for further investigation in this study to examine how learners' progress on such items would differ when they engage in diagnostic or dynamic assessment during the semester. To measure EFL learners' performance on receptive-response and productive-response items, such items corresponding

to the above-mentioned definitions and types were identified and selected from the listening section of a standardized international test (i.e., KET) to be used in this study.

Method

Participants

To start the research, initially, 120 pre-intermediate female and male EFL students with the age range of 15-25 years old in six intact classes (three female and three male classes) were selected through convenience sampling. They were all studying English in a language institute in Karaj, Iran, and their L1 was Persian. To ensure having homogeneous participants, they participated on the Nelson proficiency test and 90 individuals who scored between one standard deviation above and below the mean score on the test were selected as the final participants of the study.

Instruments

Four different instruments were used in this research. These are explained below.

A Nelson Proficiency Test. Due to the importance of having homogeneous participants in a quantitative study, a Nelson test was used to check the proficiency level of the students to have homogeneous groups. The purpose of using a proficiency test is to reduce the variability among the participants' performance to make the results more reliable and comparable (Haywood & Lidz, 2006). The 150 version of the Nelson test was used in this study, which is appropriate to check the level of the pre-intermediate EFL learners, who are supposed to have 150 hours of instruction. The test consists of 50 multiple choice items.

Key English Test (KET). The listening sections of two standardized KET tests were used as the pretest (KET, 2015) and posttest (KET, 2014) in this study. KET is appropriate for the learners at the pre-intermediate level as they are supposed to be able to listen and realize dialogs and monologs in formal and informal settings about everyday life topics. In addition to being appropriate for the pre-intermediate level, KET was selected because it includes both receptive-response and productive-response items. The test consisted of five parts and 25 items, out of which 15 were receptive-response items and 10 were productive-response items. There were 10 multiple-choice and 5 matching items in three parts (representing receptive-response items), and 10 fill-in-the-blanks items in two parts (representing productive-response items). The number of receptive-response and productive-response items were the same in both the pretest and posttest.

Practice Tests. Four listening practice tests with both receptive-response and productive-response listening items were used during the term in the DIA and DA experimental groups to implement the required treatments. They are explained in details in the procedure section.

Top Notch 1B Textbook. The textbook which was taught in all the three groups throughout the semester was Top Notch 1B (Saslow & Ascher, 2011). This is a skills-based textbook to teach English to EFL students and it focuses on all the skills and components of language.

Procedure

The final participants were 90 pre-intermediate female and male EFL students in six intact classes who were selected through convenience sampling and their performance on the Nelson proficiency test. The classes were randomly divided into three groups of DIA, DA, and control. Each group comprised two classes (one female and one male class) to have 30 participants

in each group. The three groups sat for the listening section of KET (2015) as the pretest to identify their ability to perform on receptive-response and productive-response listening items. During the treatment, the same textbook (Top Notch 1B) was used in all the three groups to improve their knowledge of English. However, listening comprehension was practiced differently in each group depending on the nature of the required treatment.

In the DIA group, in addition to studying the textbook, the participants were given a listening practice test to find out their main strengths and weaknesses. The results of the test gave the teacher some ideas about the students' problems in listening comprehension, which were then checked with the students in a conference to know their opinions about the probable reasons for the difficulties. Some of the DIA strategies mentioned by Alderson et al. (2015) and Nicol and Macfarlane-Dick (2006) were used at this stage as:

- The teacher asked the learners to talk about the kinds of support they would receive after each test.
- The teacher asked the learners to specify exactly the parts of the test they had difficulty doing.
- The teacher pointed out the incorrect answers along with a clear explanation of the reason why they were wrong.
- The teacher helped the learners to be able to mention why they were certain about the correctness of some of their answers.

The necessary feedback and follow-up support were provided to the students in both group and individual formats. After giving the students the appropriate feedback and assistance they needed during two successive sessions following the practice test, they were given a second practice test with two main aims, to check whether their previous problems have been solved and to investigate their further problems to provide more support to them. This process was repeated with four practice tests during the semester.

On the other hand, in the DA group, next to teaching the required textbook, the teacher went through the test-mediation-retest model of dynamic assessment and used the same set of four practice tests similar to those used in the DIA group during the semester. However, each test was administered twice, once before the mediation (as a test) and once after it (as a retest). After the test, the teacher helped the learners overcome their problematic areas through the 'mediation' technique, which is unique to DA, without providing the learners with exact answers. In the mediation phase, the teacher used some strategies related to DA mentioned by Aljaafreh and Lantolf (1994, p. 471), by:

- Asking the learner to focus and read the sentence that contains the error.
- Indicating that something is wrong in a sentence or clause.
- Narrowing down the location of the error.
- Indicating the nature of the error without identifying it.
- Providing clues to help the learner find the right answer.

Then, the retest was implemented in which the same test was administered among the students to see how much they have improved their listening comprehension during the mediation phase. This process was also repeated with the same four practice tests during the semester.

Conversely, the control group went through the usual routine of EFL classes in studying the textbook and practicing listening comprehension. In this group, listening techniques were taught regularly and the students performed receptive-response and productive-response items after each listening in their textbook and after each additional listening activity they had in the class. However, no assessment techniques were used in this group since this group acted as a criterion to measure the degree of development of the participants in the experimental groups.

All the groups were taught by the same teacher, who was one of the researchers. The treatment sessions lasted for a whole semester, 16 sessions, and at the end of the semester, the listening section of another KET (2014), which was similar in content, format, and number of receptive-response and productive-response items to the pretest, was administered to the three groups as the posttest. The purpose was to examine the effectiveness of the treatments provided on receptive-response and productive-response listening comprehension items of the participants.

The design of this study was pretest/posttest non-equivalent groups, which is a quasi-experimental design, with convenience (non-random) sampling of the participants (Best & Kahn, 2006).

Results

Preliminary Investigation

To begin, the data obtained on the Nelson proficiency test and the KET pretest and posttest from the three groups were checked to see if they were normally distributed and whether to use parametric or non-parametric analyses. The one-sample Kolmogorov-Smirnov test was used to check the normality of the data, and the results are reported in Tables 1 and 2.

Table 1.

One-sample Kolmogorov-Smirnov Test of the Proficiency Test of the Three Groups

		Proficiency test of the control group	Proficiency test of the DA group	Proficiency test of the DIA group
N		30	30	30
Normal Parameters	Mean	28.27	26.73	28.23
	SD	4.89	5.11	5.31
Asymp. Sig. (2-tailed)		.33	.75	.67

Table 2.

One-sample Kolmogorov-Smirnov Test of the KET Pretest and Posttest of the Three Groups

		Pretest of the control group	Posttest of the control group	Pretest of the DA group	Posttest of the DA group	Pretest of the DIA group	Posttest of the DIA group
N		30	30	30	30	30	30
Normal Parameters	Mean	15.30	16.60	16.53	19.80	14.63	18.40
	SD	4.69	5.23	3.47	2.59	4.15	2.83
Asymp. Sig. (2-tailed)		.47	.02*	.59	.73	.74	.45

As it is shown in Table 1, the significance value of the proficiency test of the control group is .33, that of the DA group is .75, and that of the DIA group is .67, and they are all considered normal since they are all above the critical .05 level of significance ($\alpha = .05; p > \alpha$). Also, it can be seen in Table 2 that the KET pretest scores of the three groups are normally distributed since all their p values are higher than the critical .05 level (p of control group = .47, p of DA group = .59, and p of DIA group = .74; $\alpha = .05; p > \alpha$). Moreover, the KET posttest scores of DA and DIA groups are also normally distributed since their significance values are above the .05 level of significance (p of DA group = .73, and p of DIA group = .45; $\alpha = .05; p > \alpha$), whereas the significance value for the posttest of the control group is less than .05 ($p = .02; \alpha = .05; p < \alpha$), which shows it is not normally distributed. However, since only the data on one of the tests in this study is not normally distributed and because of the robustness of parametric measures in analyzing the data and generalizing the results, parametric analyses were used throughout this study.

Next, it was necessary to check the participants' homogeneity in terms of their knowledge of English at the beginning of the study. This was done by

using a one-way ANOVA to compare the mean scores of the three groups on the Nelson proficiency test. The descriptive statistics of the three groups on the Nelson proficiency test and the one-way ANOVA on their mean scores are presented in Tables 3 and 4, respectively.

Table 3.

Descriptive Statistics of the Nelson Proficiency Test of the Three Groups

Group membership	N	Mean	SD
Control Group	30	28.27	4.89
DA Group	30	26.73	5.11
DIA Group	30	28.23	5.31

Table 4.

One-way ANOVA on the Nelson Proficiency Test of the Three Groups

		Sum of Squares	df	Mean Square	F	Sig .
Nelson Test	Between Groups	46.2	2	23.01	.88	.41
	Within Groups	2271.10	87	26.10		
	Total	71595.00	90			

Table 3 displays the descriptive statistics of the scores of the three groups on the Nelson proficiency test. Having a look at this table reveals that the mean scores of the three groups (28.27 for the control group, 26.73 for the DA group, and 28.23 for the DIA group) on the proficiency test are slightly different from each other. However, to check whether this difference in the mean scores is significant or not, a one-way ANOVA was needed. As it can be seen in Table 4, there was no significant difference between the three groups' performance on the Nelson test at the beginning of the study since the related significance value is .41, which is above the critical value ($F = .88$; p

= .41; $\alpha = .05$; $p > \alpha$). In other words, the participants had similar knowledge of English at the beginning of the study; therefore, the three groups belonged to the same population, and the results of the groups would be comparable.

The next important point was checking the reliability and validity of the KET pretests and posttests in this study (Tables 5-7).

Table 5.

Reliability of Receptive-Response and Productive-Response Listening Items in the KET Pretest and Posttest

	Receptive-Response Items in Pretest	Receptive-Response Items in Posttest	Productive-Response Items in Pretest	Productive-Response Items in Posttest
N of Items	15	15	10	10
Cronbach's Alpha	.75	.66	.72	.71

As presented in Table 5, the reliability of the pretests and posttests were checked through Cronbach's Alpha, which is a popular way of calculating reliability. The reliability of receptive-response items in the pretest was .75, that of productive-response items in the pretest was .72, and that of productive-response items in the posttest was .71, which were all acceptable. However, the reliability of receptive-response items in the posttest was .66, which was a little lower than the acceptable .7 value. The not very high reliability indexes in this study might be because of the few number of receptive-response and productive-response items in the pretest and posttest.

To check the validity of the receptive-response and productive-response listening items in the pretest and posttest, their correlations with the total scores on KET, which is a standardized and already validated test, were calculated and the results are reported in Tables 6 and 7.

Table 6.

Correlations between KET Total and Receptive-Response and Productive-Response Listening Items in the Pretest

		Receptive-Response Items in Pretest	Productive-Response Items in Pretest
KET Total in Pretest	Pearson Correlation	.88	.89
	Sig. (2-tailed)	.00**	.00**

Table 7.

Correlations between KET Total and Receptive-Response and Productive-Response Listening Items in the Posttest

		Receptive-Response Items in Posttest	Productive-Response Items in Posttest
KET Total in Posttest	Pearson Correlation	.93	.80
	Sig. (2-tailed)	.00**	.00**

As the results of Tables 6 and 7 show, the receptive-response and productive-response listening items in both the pretest and posttest were highly and significantly correlated with total KET scores, which is a good manifestation of the criterion-related validity of the pretests and posttests in this study.

Addressing the Research Questions

To answer the research questions of the study and to check if the improvements in the performance of the groups on receptive-response and productive-response listening items were significant or not, it was necessary to run two repeated-measures two-way ANOVAs as well as one MANOVA. Before running these analyses, first, the descriptive statistics of receptive-response and productive-response listening items in the pretest and posttest of the three groups are reported in Table 8.

Table 8.

Descriptive Statistics of Receptive-Response and Productive-Response Items in the Pretest and Posttest of the Three Groups

		Receptive-Response Items in Pretest	Receptive-Response Items in Posttest	Productive-Response Items in Pretest	Productive-Response Items in Posttest
Control Group (N=30)	Mean	10.30	10.33	5.00	6.27
	SD	2.16	3.66	2.94	2.11
DA Group (N=30)	Mean	11.70	12.07	4.83	7.70
	SD	1.86	1.50	2.27	1.53
DIA Group (N=30)	Mean	10.57	11.30	4.07	7.13
	SD	2.64	1.93	1.79	1.35

According to the information presented in Table 8, in the control group, the mean score of receptive-response items changed from 10.30 in the pretest to 10.33 in the posttest. Also, the mean score of productive-response items changed from 5.00 in the pretest to 6.27 in the posttest. On the other hand, in the DA group, the mean score of receptive-response items changed from 11.70 in the pretest to 12.07 in the posttest, and the mean score of productive-response items changed from 4.83 in the pretest to 7.70 in the posttest. In the DIA group too, the participants' performance changed in both receptive-response and productive-response listening comprehension items from the pretest to the posttest. The mean score of receptive-response items changed from 10.57 in the pretest to 11.30 in the posttest, and the mean score of productive-response listening comprehension items changed from 4.07 in the pretest to 7.13 in the posttest.

Addressing Research Questions 1 and 3

Because there were three groups in this study whose performances were repeatedly measured on receptive-response listening items, the most appropriate data analysis technique to be run was repeated-measures two-way ANOVA (Hinton, Brownlow, McMurray, & Cozens, 2008). The result of the related repeated-measures two-way ANOVA is reported in Table 9.

Table 9.

Repeated-Measures Two-way ANOVA of Receptive-Response Listening Items in the Pretest and Posttest of the Three Groups

Effect	Value	F	Sig.	Partial Eta Squared
Time	.07	7.43	.00*	.07
Group		2.99	.05*	.06
Time * Group	.06	3.08	.05*	.06

As it is represented in Table 9, the level of significance of within-subject effect (time) is .00 which is smaller than the critical .05 level ($F = 7.43$; $p = .00$; $\alpha = .05$; $p < \alpha$); therefore, it was concluded that the performance of the three groups changed significantly from the pretest to the posttest. The second row of Table 9 reports the between-subject effect (group), which is .05 and equals the critical level. According to Pallant (2011), for a value to be considered significant, it should be smaller than or equal to .05. In other words, there was a significant difference among the three groups in doing receptive-response items ($F = 2.99$; $p = .05$; $\alpha = .05$; $p = \alpha$). Finally, considering the interaction of time and group for receptive-response items in Table 9, the reported level of significance and the critical level of significance are again the same ($F = 3.08$; $p = .05$; $\alpha = .05$; $p = \alpha$). Therefore, the conclusion was that there was a significant difference in the amount of progress in the

participants' performance on receptive-response items from the pretest to the posttest in the three groups.

To find out which group had significant progress from the pretest to the posttest on receptive-response listening items, Table 10 is provided, which shows the Scheffe post-hoc test on these items in the posttest of the three groups.

Table 10.

Scheffe Post-Hoc Test on Receptive-Response Listening Items of the Three Groups in the Posttest

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
DA	Control	1.57	.54	.01*
	DIA	.95	.54	.21
DIA	Control	.62	.54	.52

According to the information in Table 10, the significance value of the difference between the DA and control groups is .01, which is less than the critical value ($p = .01$; $\alpha = .05$; $p < \alpha$), meaning that the learners' performance on receptive-response listening items in these two groups was significantly different from each other. Looking back at their mean scores in Table 8, it becomes clear that the participants of the DA group had a far better improvement in their posttest. On the other hand, the difference between the DA and DIA groups' posttest scores on receptive-response listening items was not significant due to the significance value which is .21 and above the critical level ($p = .21$; $\alpha = .05$; $p > \alpha$). Also, the difference between the DIA and control groups' posttest scores on receptive-response listening items was not significant either ($p = .52$; $\alpha = .05$; $p > \alpha$). Overall, the DA group performed the best on the posttest of receptive-response listening items, followed by the DIA group, and then the control group.

Based on the information presented in Tables 8, 9, and 10, the participants' performance improved significantly from the pretest to the posttest on receptive-response listening items in both DIA and DA groups.

Addressing Research Questions 2 and 4

Next, another repeated-measures two-way ANOVA was run to check the differences in the performance of the three groups on productive-response listening items (Table 11).

Table 11.

Repeated-Measures Two-way ANOVA of Productive-Response Listening Items in the Pretest and Posttest of the Three Groups

Effect	Value	F	Sig.	Partial Eta Squared
Time	.39	57.57	.00*	.39
Group		1.70	.18	.03
Time * Group	.03	1.49	.22	.03

As it is shown in Table 11, time as the within-subject effect has the significance value of .00, which is smaller than the critical level of significance ($F = 57.57; p = .00; \alpha = .05; p < \alpha$); this shows a significant difference in the three groups' performance on productive-response listening items from the pretest to the posttest. On the other hand, the value for the between-subject effect, group, is .18 and larger than the critical level ($F = 1.70; p = .18; \alpha = .05; p > \alpha$), which means there was no significant difference in answering productive-response listening items among the three groups. More importantly, it can be seen in Table 11 that the level of significance of the interaction of time and group is .22 and larger than the critical level ($F = 1.49; p = .22; \alpha = .05; p > \alpha$), which means there was no significant difference

in the three groups' amount of progress on productive-response listening items from the pretest to the posttest. Overall, Table 11 shows that the three groups performed significantly better on productive-response items in the posttest.

Once more, to find out which group had significant progress from the pretest to the posttest, the Scheffe post-hoc test was run on productive-response listening items whose results are provided in Table 12.

Table 12.

Scheffe Post-Hoc Test on Productive-Response Listening Items of the Three Groups in the Posttest

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
DA	Control	.63	.47	.40
	DIA	.67	.47	.37
DIA	Control	.03	.47	.99

As it is represented in Table 12, there is no significant difference among the three groups' performance on productive-response listening items in the posttest since all the three significance values are larger than the critical level (p of DA and control = .40; p of DA and DIA = .37; p of DIA and control = .99; $\alpha = .05$; $p > \alpha$). However, the DA group performed the best on the posttest of productive-response listening items, followed by the DIA group, and then the control group.

The information in Tables 8, 11, and 12 also show that the participants' performance improved significantly from the pretest to the posttest on productive-response listening items in both DIA and DA groups.

Addressing Research Question 5

To check the interaction between all the independent and dependent variables simultaneously, one MANOVA was run. MANOVA is used to

compare different groups on a number of dependent variables (Hinton, et al., 2008; Pallant, 2011). Therefore, Table 13 provides the results of MANOVA, which was run to check the interaction effect of both DA and DIA on both receptive-response and productive-response listening items. Using this analysis made it possible to check the three group's improvement on both receptive-response and productive-response listening items from the pretest to the posttest simultaneously.

Table 13.

MANOVA on Receptive-Response and Productive-Response Items in the Pretest and Posttest of the Three Groups

Source	Measure	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Receptive-Response	252.050	1	29.606	3.496	.06	.02
	Productive-Response	352.800	1	149.422	29.871	.00*	.14
Group	Receptive-Response	125.000	2	38.756	4.576	.01*	.05
	Productive-Response	296.450	2	12.600	2.519	.08	.02
Time * Group	Receptive-Response	54.450	2	12.289	1.451	.23	.01
	Productive-Response	88.200	2	3.889	.777	.46	.00

As reported in Table 13, the significance values for receptive-response and productive-response types of items in the case of time (within-subject effect) for the three groups are .06 ($p = .06$; $\alpha = .05$; $p > \alpha$) and .00 respectively ($p = .00$; $\alpha = .05$; $p < \alpha$). Therefore, it was concluded that although there was not significant progress on receptive-response items from the pretest to the

posttest, considerable progress was seen on productive-response items. That is, the three groups did not have a considerable amount of improvement from the pretest to the posttest on receptive-response items, but significant progress was seen on productive-response items.

On the other hand, regarding the grouping of the participants (between-subject effect), the significance value reported for receptive-response items is .01, which is below the critical level ($p = .01$; $\alpha = .05$; $p < \alpha$). That is, there was a significant difference among the three groups' performance on receptive-response items. However, the results are different regarding the learners' performance on productive-response items. That is, based on the significance value reported for productive-response items, which is .08 ($p = .08$; $\alpha = .05$; $p > \alpha$), there was not a significant difference in the performance of the three groups on productive-response items.

The last and the most important row presents the information about the interaction of time and group, which shows the amount of progress of the participants of the three groups in the two kinds of items. Here, the significance value reported for receptive-response items is .23 and that of productive-response items is .46, which are both higher than the critical level (p of receptive-response = .23; p of productive-response = .46; $\alpha = .05$; $p > \alpha$). Therefore, the conclusion was that the interaction of time and group did not show a significantly different performance in the participants of the three groups on either receptive-response or productive-response items. In other words, the treatments did not have different degrees of effect on the participants' progress from the pretest to the posttest on either type of item. It is worth remembering that in the case of receptive-response items, the DA group had more improvement than the DIA group, while in the case of productive-response items, it was the reverse; the DIA group outperformed the DA group. Besides, the participants of the control group had the least

amount of improvement on both receptive-response and productive-response items.

Finally, to check which group had a significant performance in the posttest of both receptive-response and productive-response items, the Scheffe post-hoc test was run and is reported in Table 14.

Table 14.

Scheffe Post-Hoc Test on Receptive-Response and Productive-Response Listening Items of the Three Groups in the Posttest

	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Receptive-Response	DA	Control	1.60	.53	.01*
		DIA	.93	.53	.21
	DIA	Control	.67	.53	.45
Productive-Response	DA	Control	.30	.40	.76
		DIA	.90	.40	.09
	DIA	Control	.60	.40	.34

As the results of the Scheffe post-hoc test in Table 14 show, the only significant difference is between the scores of DA and control groups on the posttest of receptive-response listening items ($p = .01$; $\alpha = .05$; $p < \alpha$). On the other hand, the differences between DA and DIA groups and between DIA and control groups on the posttest of receptive-response listening items were not significant (p of DA and DIA groups = .21; p of DIA and control groups = .45; $\alpha = .05$; $p > \alpha$). Further, no significant differences were seen among the three groups on the posttest of productive-response listening items (p of DA and control groups = .76; p of DA and DIA groups = .09; p of DIA and control groups = .34; $\alpha = .05$; $p > \alpha$), which were in line with the related non-significant result of the group for productive-response items in Table 13.

However, the important point in Table 14 is the non-significant difference between the two experimental groups (DA and DIA) in their performance on both receptive-response and productive-response items. Consequently, the five research questions of the study could be answered as the following:

1. Diagnostic assessment has a significant effect on EFL learners' progress on receptive-response listening comprehension items.
2. Diagnostic assessment has a significant effect on EFL learners' progress on productive-response listening comprehension items.
3. Dynamic assessment (its interventionist model) has a significant effect on EFL learners' progress on receptive-response listening comprehension items.
4. Dynamic assessment (its interventionist model) has a significant effect on EFL learners' progress on productive-response listening comprehension items.
5. There is no significant difference between the effect of diagnostic vs. dynamic assessment on EFL learners' progress on receptive-response and productive-response listening comprehension items.

Discussion

In this study, the effect of DIA vs. DA on EFL learners' progress on receptive-response and productive-response listening comprehension items was examined. The results showed that all three groups made progress on both receptive-response and productive-response listening comprehension items to different degrees. The findings revealed some interesting points as well. An important issue was that all the learners had difficulty in productive-response items in the pretest as the three groups' mean scores were rather low. This was most probably because productive-response items are more challenging as students are required to write what they have heard, which is a demanding

mental process. However, both DIA and DA helped learners to improve both their receptive-response and productive-response listening items, especially the productive-response items, which shows the effectiveness of these two assessment types for instructional purposes. Besides, even though all the participants' performance changed from the pretest to the posttest, in the control group, the improvement was the least on both types of listening items. In contrast, the improvement on both types of items was significant in the two experimental groups. In other words, it could be said that the feedback and support provided in the DIA group and the mediation provided in the DA group had significantly affected the learners' progress on both receptive-response and productive-response listening comprehension items.

The results of this study are relatively similar to other studies conducted in similar domains. Regarding DA, Sarani, and Izadi (2016) investigated the effect of DA on predicting and improving EFL learners' knowledge of receptive vocabulary. They found that taking the actual scores as the representation of the learners' ability does not give a complete picture of their capabilities. Instead, mediated scores (scores obtained when students are guided through DA) should be reported if the aim is to identify the areas learners have difficulty with to help them improve their abilities. Also, Ahmadi Safa and Beheshti (2018) studied the influence of two types of group dynamic assessment (GDA) – interactionist and interventionist – on EFL learners' listening comprehension. They concluded that interactionist GDA had a more significant impact on the improvement of EFL learners' listening comprehension, most probably because of its more interactive nature. Further, Shabani (2018) in his research found the effectiveness of using GDA to enhance EFL learners' writing ability. In general, the results of most studies on DA, in addition to those of this study, showed the effectiveness of this strategy in EFL classes, which support the claims made by pioneers of DA on

its usefulness in educational contexts (Lantolf & Poehner, 2004; Poehner, 2014).

On the other hand, regarding DIA, Nikmard and Tavassoli (2020) also identified the positive impact of DIA on improving both selective and productive reading comprehension tasks among EFL learners, which was very close to the findings of this study in showing the significance of DIA in EFL classrooms. Furthermore, Granfeldt and Ågren (2014), who used Direkt Profil as a diagnostic assessment tool, found the usefulness of such a tool as it provided the learners with fast and comprehensive feedback that indicated how some specific kinds of linguistic structures related to their diverse phases of development. Even though DIA has not been researched as much as DA in EFL classes, the results of previous studies on DIA, as well as this study, showed its positive impact on various aspects of language knowledge. These findings are in line with what is stated about the effectiveness of DIA in the literature (Jang & Wagner, 2014).

As the findings of this study and similar studies indicated, DIA and DA are beneficial techniques to be used in EFL classes since both of these techniques help learners to know their strengths and weaknesses, and enable them to find a way to resolve their problems through either the feedback and support provided in DIA or the mediation provided in DA. The point which has not yet been focused in the literature was comparing and contrasting these two types of assessment for instructional purposes and enhancing different types of listening items. This study showed that both kinds of assessment are advantageous ways of helping EFL learners improve their listening comprehension ability. More specifically, it was found that DA is a more helpful technique when the focus is on improving performance on receptive-response listening items whereas DIA is a better technique to improve performance on productive-response listening items.

Conclusion

The results of the study, which showed the positive impact of both DIA and DA on improving EFL learners' progress on receptive-response and productive-response listening comprehension items, have some implications for EFL learners, teachers, and testers, and can help them do their job more efficiently. EFL learners following DA can become aware of their problems and get help from their teachers through mediation in a friendly manner to solve their problems since the mediation provided by teachers is one of the significant features of DA. Also, EFL learners following DIA can diagnose their weaknesses with the teacher's help during the course and try to cope with them as best as they can. Similarly, EFL teachers following DA can find the main problematic areas of their students at each level through continuous assessment and then provide the necessary mediations to them to overcome those problems. Besides, through DIA during the course, teachers can identify their students' weak points and create rapport with them to provide the necessary feedback and to help them improve different aspects of their language knowledge. Also, language testers can use DA and DIA instead of traditional one-shot tests to let teachers identify their students' weak points and help them overcome their problems throughout the course, instead of merely deciding whether learners can pass the course or not.

As the final point, no research is devoid of limitations, and this study was no exception. The participants' gender, age, educational background, and some other individual features were not taken into account in this study, which might have changed the results and may provide areas for further research to investigate their role on EFL learners' listening comprehension ability. Some other suggestions are also recommended to interested researchers who are keen on working on different types of assessment, especially DIA and DA for instructional purposes. A similar study can be done on other types of

items/activities mentioned in the literature. Also, the effect of group DA and group DIA on EFL learners' progress on any of the language skills can be investigated. Finally, qualitative studies can be done on how using DIA or DA influences what teachers and students do in EFL classes and what they think about the usefulness of these assessment types in enhancing the learners' language knowledge.

References

- Ahmadi Safa, M., & Beheshti, S. (2018). Interactionist and interventionist group dynamic assessment (GDA) and EFL learners' listening comprehension development. *Iranian Journal of Language Teaching Research*, 6(3), 37-56.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), 236-260.
- Alderson, J. C., Haapakangas, E. L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2014). *The diagnosis of reading in a second or foreign language*. Routledge.
- 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.
- Best, J. W., & Kahn, J. V. (2006). *Research in education (3rd ed.)*. Boston: Pearson Education Inc.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices (2nd ed.)*. NY: Pearson Longman.
- Brown, J. D. (2005). *Testing in language programs (2nd ed.)*. New York: McGraw-Hill Companies.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3-25.
- Dai, D. W., & Roever, C. (2019). Including L2-English varieties in listening tests for adolescent ESL learners: L1 effects and learner perceptions. *Language Assessment Quarterly*, 16(1), 64-86.

- Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review*, 55(1), 15-32.
- Esfandiari, S., & Tavassoli, K. (forthcoming). The comparative effect of self-assessment vs. peer-assessment on young EFL learners' performance on selective and productive reading tasks. *Iranian Journal of Applied Linguistics (IJAL)*.
- Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in a content-based classroom. *TESOL Quarterly*, 37(2), 247-273.
- Granfeldt, J., & Ågren, M. (2014). SLA developmental stages and teachers' assessment of written French: Exploring direkt profil as a diagnostic assessment tool. *Language Testing*, 31(3), 285-305.
- Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 32(3), 317-336.
- Hayati, A. M., & Jalilifar, A. (2009). The impact of note-taking strategies on listening comprehension of EFL learners. *English Language Teaching*, 2(1), 101-111.
- Haywood, H. C., & Lidz, C. S. (2006). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge: Cambridge University Press.
- Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2008). *SPSS explained*. New York, NY, 10016.
- Jang, E. E., & Wagner, M. (2014). Diagnostic feedback in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development, Volume II* (pp. 157-175). John Wiley & Sons, Inc.
- Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech Language Pathology*, 9(4), 273-285.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49-72.
- Lantolf, J. P., & Poehner, M. E. (2008). Dynamic assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.)*, Volume 7: *Language testing and assessment* (pp. 1-12). Springer Science + Business Media LLC.
- Li, C. H. (2019). Using a listening vocabulary levels test to explore the effect of vocabulary knowledge on GEPT listening comprehension performance. *Language Assessment Quarterly*, 16(3), 328-344.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.
- Nikmard, F., & Tavassoli, K. (2020). The effect of diagnostic assessment on EFL learners' performance on selective and productive reading tasks. *Journal of Modern Research in English Language Studies, 7*(1), 79-104.
- Pallant, J. (2011). *SPSS survival manual*: McGraw-Hill Education.
- Poehner, M. E. (2014). Dynamic assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development, Volume II* (pp. 194-210). John Wiley & Sons, Inc.
- Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (CDA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing, 32*(3), 337-357.
- Ranjbaran, F., & Alavi, S. M. (2016). Cognitive diagnostic assessment of a reading comprehension test battery for formative diagnostic feedback. *Journal of Foreign Language Research (JFLR), 6*(2), 321-342.
- Rost, M. (1994). *Introducing listening*. London: Penguin.
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing, 37*(1), 31-53.
- Saito, K., & Akiyama, Y. (2018). Effects of video-based interaction on the development of second language listening comprehension ability: A longitudinal study. *TESOL Quarterly, 52*(1), 163-176.
- Sarani, A., & Izadi, M. (2016). Diagnosing L2 receptive vocabulary development using dynamic assessment: A microgenetic study. *Journal of Teaching Language Skills, 35*(2), 161-189.
- Saslow, J., & Ascher, A. (2011). *Top Notch 1B (2nd ed.)*. Pearson Education, Inc.
- Shabani, K. (2018). Group dynamic assessment of L2 learners' writing abilities. *Iranian Journal of Language Teaching Research, 6*(1), 129-149.
- Tavassoli, K., & Nikmard, F. (2019). The effect of dynamic assessment on EFL learners' performance on selective and productive reading comprehension tasks. *Journal of Foreign Language Research (JFLR), 9*(2), 445-478.
- Wagner, E. (2014). Assessing listening. In A. J. Kunnan, (Ed.). *The companion to language assessment: Approaches and development, Volume I* (pp. 47-63). John Wiley & Sons, Inc.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

- Winke, P., & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, 14(4), 380-397.
- Zareinajad, M., Rezaei, M., & Shokrpour, N. (2015). The effects of receptive and productive task-based listening activities on the listening ability of Iranian EFL learners at different proficiency levels. *Pertanika Journal of Social Sciences and Humanities*, 23(2), 537-552.
- Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System*, 66, 156-167.