The Effect of Oral Feedback on Iranian EFL Learners' Complexity and Accuracy in Speaking

Mobina Rahnama[1], Alireza Ahmadi[2*], Seyyed Ayatollah Razmjoo[3], Omid Mazandarani[4]

[1]Department of English Language Teaching, Qeshm Branch, Islamic Azad University, Qeshm, Iran, *mobina.rahnama@gmail.com*

[2*] Associate Professor of TEFL, Department of Foreign Languages & Linguistics, Shiraz University, Iran, *arahmadi@shirazu.ac.ir*

[3] Professor of TEFL, Department of Foreign Languages & Linguistics, Shiraz University, Iran, *arazmjoo@rose.shirazu.ac.ir*

[4] Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran, *omazandarani@gmail.com*

Abstract

The debate continues on the influence of oral feedback on the features of oral performance. The present study tries to provide an answer to this question in an EFL context. To this end the effect of six different modes of oral feedback on the features of oral complexity and accuracy was investigated using data from 66 Iranian EFL learners who were selected conveniently from the Iran Language Institute. The participants were divided into experimental and control groups at two different levels of elementary and pre-intermediate. The experimental groups were presented with six different types of oral feedback modes (recasts, clarification requests, metalinguistic, praising, elicitation, and repetition) and at the end of the research they were tested by an in-class oral test to measure their complexity and accuracy (CA). To compare the participants' oral features, a Wilcoxon Signed Rank Test and a Mann-Whitney U Test were run. The results indicated that complexity and accuracy significantly improved among the groups, moreover; there were significant differences in the post-tests between both elementary and pre-intermediate levels regarding CA. The results further indicated that Iranian learners of English would have fewer errors and would be more accurate when receiving oral feedbacks. The study highlights the complex relationship that exists between features of oral performance. The findings of the present study can have theoretical and practical implications for syllabus designers, teacher trainers, and testing researchers.

Keywords: Accuracy, Complexity, Oral Feedback Modes, Oral Performance

1. Introduction

It has been almost half a century since second language acquisition (SLA) researchers started to explore the best way to teach a second language (Varnosfadrani & Basturkmen, 2009). Once teacher intervention was considered ineffective (Krashen, 1985; Long, 1981); by now, however, there is a consensus that to 'push' learners and, thus, to foster second language development, instruction and overall, teachers' practices, it is essential to take it into account (Lightbown & Spada, 2006). Teachers use different methods to support learners in their development of knowledge and proficiency in the target language and help them use the tools needed to learn the language. One method is feedback, which is used to encourage learners or correct them when making an error. The phenomenon of language learners' errors has attracted the attention of a wide range of researchers from varied fields including psychologists, psycholinguists, first and second language educationalists, applied linguists and teachers (Chaudron, 1977; Corder, 1974; Dekeyeser, 1993; Erdogan, 2005; Selinker, 1974), for a number of decades.

The issue of errors in a second language is one of the most fundamental concepts in applied linguistics; accordingly, it is the subject of ongoing and intense debate. Often, this debate is about competing theories or models on the nature of error, whether it is constructive or destructive, and on the effectiveness of error treatment. Hendrickson's study in 1978 and many others since then, have questioned whether or not errors should be corrected at all. Back in the fifties and early sixties, second language scholars considered errors as 'sins' that should not be tolerated and their occurrences should be prevented (Skinner, 1957). Later, some argued that error correction does not facilitate L2 learning and may even hinder it (Krashen, 1994; Truscott, 1999). With the advent of the communicative approach to language teaching, however, the seventies witnessed an upsurge of interest in studying language learners' errors. Most of the research conducted at that time examined different types of errors and explained their possible causes.

Fanselow's (1977) study, yet, paved the way for another shift of interest, i.e. from analyzing learners' errors to focusing on what teachers do when errors are committed, namely, error treatment. Evaluations of the language programs have highlighted the need for some form-focused instruction to improve learners' accuracy (Lightbown & Spada, 1990). Hence, a more balanced view on the nature of error and error treatment has been taken by language learning specialists in the last two or three decades (Lennon, 1991; Tomasello & Herron, 1989). It is proved now that judicious

error correction as manifested in the form of corrective feedback (CF) is facilitative in fostering second language development.

Among the language skills, speaking development has been an ongoing and challenging process for both teachers and learners. The ability for learners to express their ideas orally in a second or foreign language with reasonable coherence and accuracy is a considerable achievement for foreign learners. However, speaking is not a skill to be acquired naturally, as Li (2014) maintains, speaking is essentially learned and transmitted through experience and practices in formal and informal instructional settings. In the process of speaking, learners commit different errors in expressing their ideas. When L2 learners make linguistic decisions in their speaking, they build linguistic hypotheses based on learning experiences. Thus, corrective feedback is regarded by many scholars to be an important facilitative strategy in error treatment. Several studies have provided strong evidence of its value in the process of language development (see meta-analyses by Mackey & Goo, 2007; Russell & Spada, 2006).

Applied linguists have identified three components or features of second language performance: complexity, accuracy, and fluency (CAF) to measure L2 development (Ellis & Barkhuizen, 2005; Housen, Kuiken, & Vedder, 2012). According to Housen and Kuiken (2009), CAF are very beneficial components that can be used as performance descriptors for the oral and written assessment of language learners, as indicators of learners' proficiency underlying their performance, or as units for measuring progress in language learning. While these components are considered valuable in L2 research, they remain under-investigated in corrective feedback research. There are not many studies focusing on the developmental features of corrective feedback through the lens of CAF. The majority of error treatment studies (Ellis, Loewen, & Erlam, 2006; Hunter, 2011; Hyland, 2003; Lee, 2013; Meihami & Rashidi, 2018; Quinn, 2014; Russell & Spada, 2006; Saito & Lyster, 2012) have focused on one or two components of CAF. Isolating one or two aspects of corrective feedback only enables researchers to shed light on one or two pieces of the jigsaw puzzle, overlooking other pieces that could be essential for the completion of the picture. Accordingly, exploration and comparison of oral feedback (OF) modes through the lens of CAF are of value to ascertain more precisely what L2 aspects are influenced by various oral feedback modes. It consequently gives a more holistic picture of the effectiveness of various OF modes in the development of grammatical and lexical complexity, accuracy, and fluency.

In response to the issues raised above, this study intended to investigate the impact of oral feedback modes on foreign language learners' development of CA. Considering the fact that the mode of corrective

feedback provided by the teacher is a pivotal variable in determining the extent of the OF effectiveness, the study specifically examined the comparative effectiveness of a variety of OF modes. Of particular interest were six OF modes that previous studies had suggested may affect L2 performance: recasts, clarification requests, metalinguistic feedback, praising, elicitation, and repetition.

The current study focused on English learners' speaking ability as the construct of interest and the main purpose was to probe how different modes of corrective feedback triggered by different interactional moves affect subsequent development in oral performance. Capturing development of speaking ability, however, is much more difficult than measuring a unitary and static aspect of the learners' utterances. Foreign language and L2 speaking ability is a complex phenomenon requiring multiple constructs to understand its development in order to ultimately understand the differential effects of various oral feedback modes. Researchers in the area of L2 language learning (Ellis, 2003; Skehan 1998) are now in agreement that L2 proficiency, in general, and speaking proficiency, in particular, are multi-componential in nature, and that their principal dimensions can be adequately captured by the notions of complexity, accuracy and fluency (Housen & Kuiken, 2009). Despite decades of research dedicated to error treatment, many questions about the effectiveness of error treatment through oral feedback remain unanswered: Is it possible to develop L2 learners' grammatical accuracy, fluent speech, and sufficient complexity with the help of oral feedback techniques? Do accuracy, fluency, and complexity develop simultaneously or separately? What factors affect the development of these variables?

The present study, henceforth, investigated how the demands of different modes of oral feedback affected the quality of learners' speaking ability as displayed by the complexity, and accuracy of their utterances. In this study, complexity refers, according to Hepford (2017), to learners' ability to incorporate a variety of grammatical structures, diverse vocabulary, and rich descriptive language (adjectives, adverbs, relative clauses, etc.), and accuracy refers to the ability to produce grammatically and semantically correct sentences. These features are measured in different ways which will be explained in detail in the following sections. By measuring these features, the research aimed to observe how features of learners' speaking ability might be affected by the six different OF types mentioned earlier.

Given the above-mentioned objectives, the present study aimed to address the following questions:

1. What is the effect of using oral feedback modes (recasts, clarification requests, metalinguistic, praising, elicitation, and repetition) on Iranian EFL learners' oral complexity?

2. What is the effect of using oral feedback modes (recasts, clarification requests, metalinguistic, praising, elicitation, and repetition) on Iranian EFL learners' oral accuracy?

3. Does the use of oral feedback affect the improvement of complexity and accuracy differently in elementary and pre-intermediate EFL learners?

2. Literature Review

Brown (2007) asserts that although L2 learners vary in the progression of their language acquisition, there are four stages of development, in terms of errors, during the learning process. The first stage is what he calls the random errors stage which indicates the stage of experimentation and inaccurate guessing. At this stage the learners are only vaguely aware that there is some rule and order to a particular class of items and try to experiment their wild guesses. The second, or emergent, stage is when the learners have begun to internalize certain rules, although the rules may not necessarily be correct by target language standards. Brown (2007) believes that at this stage the learner is still unable to correct errors when they are pointed out by someone else. The third stage is the systematic stage where the rules internalized by learners are more present in their minds, and even though they might still not be completely correct, they are closer in accuracy to the target language than before. The learners are, at this stage, able to correct their errors, even when they are pointed out subtly. The stabilization stage is the final stage in which the learners' fluency and intended meanings are not a problem and they can get their points across without difficulty. This stage is characterized by the learners' ability to self-correct, and it is in this stage the language can stabilize so fast that fossilization can occur.

Since L2 researchers and practitioners are of the view that learner errors should be treated at specific moments during the classroom experience, they are now concerned with the question of which specific errors to be corrected. Should some, or all of the errors, be corrected? Cathcart and Olsen's (1976) findings indicate that 91% of the learners preferred continuous correction, i.e. to be corrected all or most of the time. Its shortcoming, however, was that learners complained that it was relatively impossible for their L2 speech to be coherent, due to the continuous interruption and error correction during their productions. Hendrickson's (1978) stance is to deal with only some of the errors. He argues that when the teacher overlooks some errors and corrects others, learners feel more

comfortable in speaking than if the teacher were to correct every error. Accordingly, he concludes that while errors should be corrected, the correction of all errors was undesirable, or at least unfeasible.

Recently, some studies started to focus on the effects of instruction and error correction on L2 speech production (Lyster, Saito, & Sato, 2013; Saito, 2013; Saito & Lyster, 2012). Havranek (2002) recommends corrective feedback to grammar errors, arguing that the correction of grammatical errors led to better results in subsequent tests. Her findings suggest that with the correction of a grammatical error, a learner is both informed or reminded of the applicable language rules and provided an example of the correct use of the rules. Katayama's (2007) survey research focuses on the benefits of correcting pragmatic errors. He found that Japanese learners pursuing English as L2 strongly preferred their pragmatic errors, and those errors that inhibited their communication, to be corrected. In conclusion, the research supports what Hendrickson (1978) established: the errors that should receive priority are those that occur frequently, those that inhibit coherent communication, and those that put a learner at risk of social stigma. In other words, systematic errors should be addressed.

There are many different ways in which error treatment can be addressed. Based on their descriptive study of teacher-student interaction in French immersion classrooms, Lyster and Ranta (1997) developed six kinds of CF types. The six types they identified were: 1) recasts, 2) elicitation, 3) metalinguistic feedback, 4) clarification requests, 5) repetition, and 6) explicit correction. They subsequently regrouped these teacher feedback types into two broad categories: reformulations and prompts (Ranta & Lyster, 2007). Reformulations embrace recasts and explicit correction, because both of these moves supply learners with either implicit or explicit reformulations of their non-target output and are thus input-providing. The Prompts category, in turn, includes a variety of output-promoting corrective techniques that are designed to push learners to self-repair by recognizing the corrective intent of CF and helping learners to fix the error on their own, and subsequently correcting it. Accordingly, techniques like elicitation, metalinguistic clues, clarification requests, and repetition come under this category.

In addition to Ranta and Lyster's (2007) classification, some other theorists also classified teacher feedback types into some broad categories. Loewen and Nabei (2007) regrouped the six feedback types according to two criteria: other-repair or self-repair, and the explicitness of corrective feedback. Figure 1 illustrates their classification of oral feedback. When the correct form is not provided by the teacher or interlocutor, and the learner is prompted to correct the error by him/herself it is regarded as self-repair. Furthermore, when the correct form is given to the learner by the teacher or interlocutor, the

feedback is considered as other-repair. In other words, self-repair feedback types require learners to reformulate problematic utterances after receiving teachers' feedback, which includes the whole range from explicit to implicit:
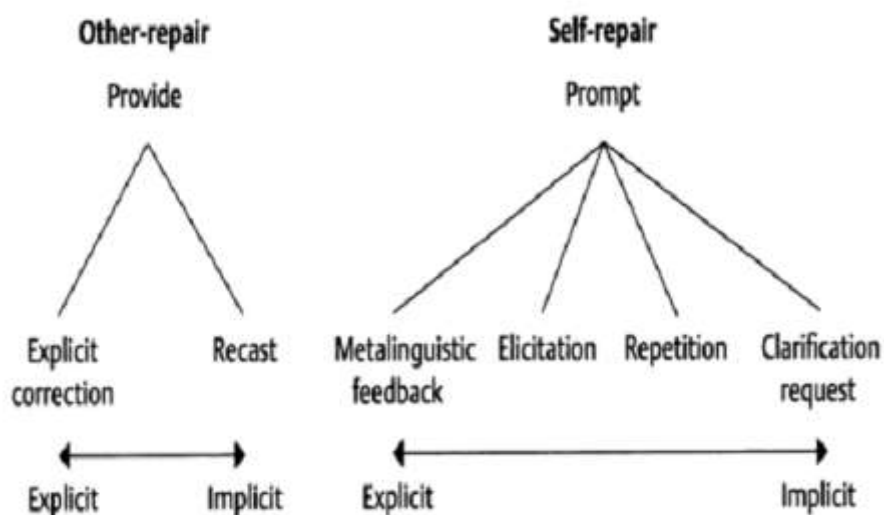
metalinguistic feedback, elicitation, repetition and clarification requests. Other-repair, which includes recasts and explicit correction, refers to a feedback type that provides learners with correct forms in either an explicit or implicit way (Loewen & Nabei, 2007).

Figure 1. Feedback Types Classified by Loewen and Nabei (2007, p. 326)

Ellis and Barkhuizen (2005) define complexity as a variable determining the extent to which the language produced by the L2 learner is elaborate and varied, and the learner's ability to use a wide range of different structures. Due to its multifaceted nature, complexity of produced language has drawn the most controversy in the field. In measuring the level of complexity, SLA scholars commonly differentiate between two types of complexity: According to Ortega (2003), the first, structural or syntactic complexity, refers to the variety in the range of syntactic forms, such as types of clauses, descriptive modifiers, and verb forms and the degree of sophistication of such forms. In theory, as L2 learners develop their language proficiency, they will incorporate more complex structures in their speech.

Ortega (2012) indicates that the metrics of syntactic complexity exhibit a strong reliable relationship with proficiency which, therefore, can be used as a useful shortcut to measure general proficiency. Although research into variables that reliably indicate syntactic complexity goes on (Vyatkina, 2012), most recent research suggests two measures of syntactic complexity to assess development: subordination which is defined as the number of

dependent clauses per unit of analysis and elaboration which is indicated in terms of the total number of words per unit of analysis (Ellis & Barkhuizen, 2005).

Accuracy is the most easily defined of the three in that there is more agreement in the goal, which is matching the target language. Pallotti (2009) defines accuracy as the degree of conformity to certain standards. According to Housen and Kuiken (2009), accuracy simply refers to "error-free" speech, i.e. the ability to produce grammatically and semantically correct sentences. However, the issue of prescriptive versus descriptive language has led to ambiguity and controversies in defining accuracy (Ellis & Barkhuizen, 2005). It is unclear from which dialect the accuracy standards should come. For instance, when analyzing the output of L2 English learners, which version of English is to be the "standard" – e.g., British, American, Australian, or Singaporean? Additionally, which dialect within each of those versions is the "standard"? Although most researchers acknowledge the flaws, SLA studies tend to define grammatical standards by adopting the researcher's native-speaker intuition and a few additional raters (Polat & Kim, 2014).

Remaining work needs to be done in order to establish how teachers use the various feedback modes for features of oral performance, and which modes are likely to be the most successful and effective for them. Further research is also needed to determine the types of feedback that would best succeed in promoting progress in the target language. Thus, it is crucial to carry out further investigation in this study with the purpose of clarifying and providing more empirical evidence about oral feedback modes and their impacts on L2 complexity and accuracy.

3. Method

3.1. Participants

A total number of 66 participants who comprised four groups of Iranian female EFL learners from Iran Language Institute of Gorgan in the North of Iran participated in this study. They were selected through convenient sampling. The participants ranged in age from 14 to 35 years. They were selected from four intact classes from two levels of proficiency. The first experimental group consisted of 17 elementary learners. The second group of participants, as a control group, consisted of 16 elementary learners studying at the same institute. The third group of participants in the present study was a sample of 18 pre-intermediate learners studying at the same institute; moreover, the fourth group was a control group which was made up of 15 learners. All the participants passed the same placement test at ILI and they had the same English educational backgrounds. Some participants were

absent for the pre-test; therefore, the final number of learners who actively participated in this study dropped from 71 to 66.

3.2. Materials and Instruments

To collect data the following instruments were employed. In order to ascertain the participants' speaking complexity, accuracy as well as the influence of the corrective feedback modes on CA, two in-class oral interaction tests from IELTS Speaking Tests (Iravani, 2007) were administered to the learners. Two parallel versions of an oral test were developed, with one serving as the pretest and the other as the posttest. They were asked a question to elaborate on and were supposed to talk about it for approximately two minutes while their voices were recorded. Each version of the test would measure productive knowledge that initially contained one main topic question with four related sub-questions. The participants were asked a question from a task card on a particular topic, and it included key points that they talked about. They gave a short talk, of about 1-2 minutes, on a simple topic. The questions did not require any specific knowledge, but were based on personal experience.

3.3. Procedure

The following procedures were followed to achieve the purpose of the study.

Due to the participants' educational level, they were placed into different groups based on their educational backgrounds. As mentioned previously, first a total of 71 Iranian female elementary and pre-intermediate learners from four intact classes were selected through availability sampling, and they sat a speaking pre-test in the second week of the autumn semester in 2016. The experimental groups were assigned to two groups (elementary and pre-intermediate) each receiving six modes of oral feedback whenever needed: recasts, clarification requests, metalinguistic, praising, elicitation, and repetition. The learners had a 15 minute warm-up speaking time before the dialog was taught and a 15 minute post speaking task after the dialog was taught. While the participants took part in the conversations the teacher would provide the oral feedbacks. For instance, the teacher would model the correct form only of the portion that the learner said incorrectly (recasts), ask what the learner meant to say (clarification requests), point out a helpful grammar point (metalinguistic), praise learners to function as reinforcement (praising), prompt or ask a question to draw the correct form from the learner without actually demonstrating it such as "What is the (x) form of (y)?" to

simply direct the learner to try again (elicitation), and repeat verbatim the learners' errors with a rising intonation at the end as in a question to indicate that there is a problem (repetition).

Moreover, the two control groups consisted of elementary and pre-intermediate learners who received no oral feedback modes. In the control groups the participants took part in the speaking tasks and practiced speaking without receiving any feedback on their CAF, only when they asked for the correct form the teacher would provide it. Therefore, during 15 treatment sessions, which were two sessions a week, oral feedback modes were provided to the experimental groups. The ILI textbooks that helped the learners in their speaking and provided them with sufficient information were used in this study. Out of ILI elementary and pre-intermediate textbooks the dialogs and speaking tasks were practiced as the base materials of the study during 15 sessions. These tasks involved (a) Talk about yourself, (b) Describe a photo or picture, (c) Discussion, (d) Information gap activity, (e) Tell a story or personal anecdote, and (f) Oral presentation (British Council, n.d.).

In the present study, first an oral pre-test was administered and the learners' oral performances were audio-taped. At the end of the course learners both the experimental and control groups were tested using an in-class oral test as the post-test, they were to answer four questions related to the topics in approximately one to two minutes as mentioned above. On the basis of the participants' shared cultural background and gender, and the researcher's teaching experiences with such language learners, it was attempted to select two topics for the pre and post-tests that were of general interest and easily understood. For the topics to be at a similar level of difficulty and generality, the two topics revolved around the same concept of friendship and tourism, which is usually an interesting topic in the Iranian culture. The topics were (a) a lovable and inspiring friend administered as the pre-test, and (b) a city you visited as the post-test. The participants felt at ease to express their ideas on these topics and the researcher who ran all the tests and functioned as the interviewer was also their teacher. All of this was due to reduce the topic effect, which has been indicated to have a significant effect on test-takers' production (Leaper & Riazi, 2014).

The participants were to answer four questions related to the topics of the pre-test and post-test in approximately two minutes. During the testing, all speeches were audio-recorded using a digital recorder placed on the desk. At first, the audio-recorded data were transcribed by the first researcher. This study included two cooperative and eager teachers who rated the transcripts after the researcher (inter-raters) while all the oral feedback modes were provided by the researcher. Then 10% of the transcripts from the pre and

post-tests were randomly selected and checked by the second researcher. The agreement was 97.5%. Second, the transcripts were coded for clauses, words, errors, and self-corrections by the first researcher. About 10% of the data were recorded by another teacher, and the intercoder agreement was calculated to be 80% based on "the number of agreements divided by the total number of coding decisions" (Brooks, 2009, p. 350). The problematic cases were discussed and resolved. Third, language production measures were calculated on the basis of the coded features. All the coding and calculations were done by hand and finally, the obtained data were submitted to statistical analysis.

3.4. Data Analysis

The following measures were considered for all the coding and calculations of the features:

Syntactic complexity measures: average number of clauses per AS-unit (Foster, Tonkyn, & Wigglesworth, 2000). To compare measures of the produced discourse on the pre and post-tests, nonparametric statistics (Wilcoxon Signed Rank Tests and Mann-Whitney U Tests) were used. Nonparametric statistics were used in this study because of the small sample size, the nature of the data (some of the data were of frequency type rather than interval data), and violations of parametric statistics.

Accuracy measures: ratio of error-free clauses (Foster & Skehan, 1996). Following Nitta and Nakatsuhara (2014), all the errors of syntax, morphology, and lexical choice were considered, but errors of discourse and phonology were ignored.

Finally, the learners' oral performances were transcribed, then the clauses and errors were calculated by hand and the obtained data were then submitted to statistical analysis with the aid of SPSS 24.0.

4. Results and Discussion

4.1. Results

This study specifically investigated the effect of oral feedback modes; recasts, clarification requests, metalinguistic, praising, elicitation, and repetition, on Iranian EFL learners' oral complexity, and accuracy.

The results are provided in two sections. At first the results are provided about the effect of oral feedback modes on the learners' performance in terms of oral complexity. Then the results for accuracy are provided. It should be noted that all the results are provided at two different levels of proficiency: namely, elementary and pre-intermediate. At first the participants were compared based on their performance on the pre-tests to

make sure that they were not statistically different before treatment. The summary of the descriptive statistics are yielded in the following tables.

Table 1

Descriptive Statistics on Elementary learners' Oral Complexity on the Pre-tests

Complexity Pretest El	
Mann-Whitney U	127.000
Wilcoxon W	263.000
Z	-.338
Asymp. Sig. (2-tailed)	.736
Exact Sig. [2*(1-tailed Sig.)]	.763[a]

[a.] Not corrected for ties.

Table 2

Mann-Whitney U Test Results of Elementary Learners' Oral Complexity on the Pre-tests

Complexity Pretest El	N	Mean Rank	Sum of Ranks
Experimental	17	17.53	298.00
Control	16	16.44	263.00
Total	33		

A Mann-Whitney U Test, Table 1, did not reveal a significant difference regarding participants' pre-test performance on complexity in the elementary experimental group (Mean Rank= 17.53, N=17) and control group (Mean Rank= 16.44, N= 16), p= 0.736 >0.05. This indicates before treatment the two groups were similar in their oral performance.

The participants took part in a post test at the end of the instructional program to see how their oral performance had changed due to the effect of oral feedback. Tables 3 and 4 below depict the results.

Table 3

Descriptive Statistics on Elementary learners' Oral Complexity on the Post-tests

Complexity Post-test El	
Mann-Whitney U	75.000
Wilcoxon W	211.000
Z	-2.215
Asymp. Sig. (2-tailed)	.027
Exact Sig. [2*(1-tailed Sig.)]	.028[a]

[a] Not corrected for ties.

A Mann-Whitney U Test based on Table 3 revealed a significant difference regarding participants' post-test performance on Complexity in the

elementary experimental group (Mean Rank= 20.59, N=17) and control group (Mean Rank=13.19, N=16), p= 0.027<0.05. In other words the groups were different after the instructional program meaning that the oral feedbacks were influential. As depicted in Table 4 the mean rank for the experimental group is much higher than the control group depicting the influence of oral feedbacks.

Table 4

Mann-Whitney U Test Results of Elementary Learners' Oral Complexity on the Post-tests

Complexity Post-test El	N	Mean Rank	Sum of Ranks
Experimental		20.59	350.00
Control	17		
Total		13.19	211.00
	16		
	33		

Table 5

Descriptive Statistics on Pre-intermediate learners' Oral Complexity on the Pre-tests

Complexity Pretest Pre-Inter	
Mann-Whitney U	123.500
F Wilcoxon W	243.500
Z	-.418
Asymp. Sig. (2-tailed)	.676
Exact Sig. [2*(1-tailed Sig.)]	.682[a]

[a.] Not corrected for ties.

Table 6

Mann-Whitney U Test Results of Pre-intermediate Learners' Oral Complexity on the Pre-tests

Complexity Pretest Pre-Inter	N	Mean Rank	Sum of Ranks
Experimental	18	17.64	317.50
Control	15	16.23	243.50
Total	33		

A Mann-Whitney U Test as shown in Table 5 did not reveal a significant difference regarding participants' pre-test performance on complexity in the pre-intermediate experimental group (Mean Rank= 17.64, N=18) and control group (Mean Rank=16.23, N= 15), p= 0.676>0.05. This indicates before treatment the two groups were similar in their oral performance.

Another Mann-Whitney U test was used to compare the groups based on the post-tests. As depicted in Table 8 the experimental group has a higher mean rank than the control group. Table 7 also depicts that this difference is significant meaning that providing oral feedbacks has been effective.

Table 7

Descriptive Statistics on Pre-intermediate learners' Oral Complexity on the Post-tests

Complexity Post-test Pre-Inter	
Mann-Whitney U	78.000
Wilcoxon W	198.000
Z	-2.075
Asymp. Sig. (2-tailed)	.038
Exact Sig. [2*(1-tailed Sig.)]	.040[a]

[a.] Not corrected for ties.

Table 8

Mann-Whitney U Test Results of Pre-intermediate Learners' Oral Complexity on the Post-tests

Complexity Post-test Pre-Inter	N	Mean Rank	Sum of Ranks
Experimental	18	20.17	363.00
Control	15	13.20	198.00
Total	33		

A Mann-Whitney U Test as shown in Table 7 revealed a significant difference regarding participants' post-test performance on complexity in the pre-intermediate experimental group (Mean Rank=20.17, N=18) and control group (Mean Rank=13.20, N=15), p= 0.038<0.05. In other words the groups were different after the instructional program meaning that the oral feedbacks were influential.

Table 9

Wilcoxon Signed Rank Test on Elementary Learners' Oral Complexity (Experimental Group)

Complexity Pre/Post-test Experimental EL	N	Mean Rank	Sum of Ranks
Negative Ranks	4	7.50	30.00
Positive Ranks	13	9.46	123.00
Ties	0		
Total	17		

Wilcoxon Signed Rank Test revealed that a two-month, twice weekly oral feedback treatment elicited a statistically significant change in participants' performance regarding complexity (clause per AS) (p = 0.028 <0.05). The Ranks Table 9 provides some interesting data on the comparison

of participants' performance on the pre-test and post-test in the experimental group regarding complexity. It can be seen from the table's legend that 13 participants had a higher score after receiving treatment. However, four participants had a higher pre-test score than their post-test score.

Table 10

Wilcoxon Signed Rank Test on Elementary Learners' Oral Complexity (Control Group)

Complexity Pre/Post-test Control EL	N	Mean Rank	Sum of Ranks
Negative Ranks	5	7.90	39.50
Positive Ranks	8	6.44	51.50
Ties	3		
Total	16		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly speaking practice without oral feedbacks did not elicit a statistically significant change in participants' performance regarding complexity (p=0.675>0.05). The Ranks table 10 provides some interesting data on the comparison of participants' performance on the pre-test and post-test in the control group regarding complexity. It can be seen from the table's legend that only eight participants had a higher score after receiving no treatment whereas five participants had a higher pre-test score than their post-test score. Furthermore, three participants saw no change in their performance.

Table 11

Wilcoxon Signed Rank Test on Pre-intermediate Learners' Oral Complexity (Experimental Group)

Complexity Pre/Post-test Experimental Pre-Inter	N	Mean Rank	Sum of Ranks
Negative Ranks	1	14.00	14.00
Positive Ranks	15	8.13	122.00
Ties	2		
Total	18		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly oral feedback treatment modes elicited a statistically significant change in pre-intermediate participants' performance regarding complexity (p=0.005<0.05). The Ranks Table 11 provides some interesting data on the comparison of participants' performance on before and after the treatment of oral feedback regarding complexity. It can be seen from the table's legend that 15 participants had a higher score after treatment whereas one participant had a higher pre-test score than after their treatment. Furthermore, two participants saw no change in their performance.

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly speaking practice without oral feedback modes elicited a statistically significant change in pre-intermediate participants' performance regarding complexity (p=0.042<0.05). The Ranks Table 12 provides some interesting data on the comparison of participants' performance on the pre-test and post-test in the pre-intermediate control group regarding complexity. It can be seen from the table's legend that five participants had a higher score after receiving no treatment whereas none of the participants had a higher pre-test score than their post-test score. Furthermore, there was observed no change in their performance of 10 participants.

Table 12

Wilcoxon Signed Rank Test on Pre-intermediate Learners' Oral Complexity (Control Group)

Complexity Pre/Post-test Control Pre-Inter	N	Mean Rank	Sum of Ranks
Negative Ranks	0	.00	.00
Positive Ranks	5	3.00	15.00
Ties	10		
Total	15		

The results of this study indicated that the experimental groups at both elementary and pre-intermediate levels had a significantly higher performance on complexity after receiving oral feedback modes whereas the control group did not improve much. The significant difference between the experimental and control groups at both levels of proficiency after the treatment provided evidence for the effect of oral feedback modes on oral complexity.

Concerning accuracy of oral proficiency, the participants were compared based on their performance on the pre-tests to make sure that they were not statistically different before treatment on oral accuracy. The summary of the descriptive statistics is yielded in Tables 13 to 20.

Table 13

Descriptive Statistics on Elementary learners' Oral Accuracy on the Pre-tests

Accuracy Pretest El	
Mann–Whitney U	125.000
Wilcoxon W	261.000
Z	-.398
Asymp. Sig. (2-tailed)	.691
Exact Sig. [2*(1-tailed Sig.)]	.709[a]

[a.] Not corrected for ties.

Table 14

Mann-Whitney U Test Results of Elementary Learners' Oral Accuracy on the Pre-tests

Accuracy Pretest El	N	Mean Rank	Sum of Ranks
Experimental	17	17.65	300.00
Control	16	16.31	261.00
Total	33		

A Mann-Whitney U Test Table 13 did not reveal a significant difference regarding participants' pre-test performance on accuracy in the elementary experimental group (Mean Rank=17.65, N=17) and control group (Mean Rank=16.31, N=16), p= 0.691>0.05. This indicates before treatment the two groups were similar in their oral performance.

The participants took part in a post-test at the end of the instructional program to see how their oral performance (accuracy) had changed due to the effect of oral feedback. Tables 15 and 16 below depict the results.

A Mann-Whitney U Test in Table 15 revealed a significant difference regarding participants' post-test performance on accuracy in the elementary experimental group (Mean Rank=20.59, N=17) and control group (Mean Rank=13.19, N=16), p= 0.028<0.05. In other words, the groups were different after the instructional program meaning that the oral feedbacks were influential. As depicted in Table 16 the mean rank for the experimental group is much higher than the control group depicting the influence of oral feedbacks.

Table 15

Descriptive Statistics on Elementary learners' Oral Accuracy on the Post-tests

Accuracy Post-test El	
Mann-Whitney U	75.000
Wilcoxon W	211.000
Z	-2.202
Asymp. Sig. (2-tailed)	.028
Exact Sig. [2*(1-tailed Sig.)]	.028[a]

[a] Not corrected for ties.

Table 16

Mann-Whitney U Test Results of Elementary Learners' Oral Accuracy on the Post-tests

Accuracy Post-test El	N	Mean Rank	Sum of Ranks
Experimental	17	20.59	350.00
Control	16	13.19	211.00
Total	33		

Table 17

Descriptive Statistics on Pre-intermediate learners' Oral Accuracy on the Pre-tests

Accuracy Pretest Pre-Inter

Mann-Whitney U	131.000
Wilcoxon W	302.000
Z	-.146
Asymp. Sig. (2-tailed)	.884
Exact Sig. [2*(1-tailed Sig.)]	.901[a]

[a.] Not corrected for ties.

Table 18

Mann-Whitney U Test Results of Pre-intermediate Learners' Oral Accuracy on the Pre-tests

Accuracy Pretest Pre-Inter	N	Mean Rank	Sum of Ranks
Experimental	18	16.78	302.00
Control	15	17.27	259.00
Total	33		

A Mann-Whitney U Test in Table 17 did not reveal a significant difference regarding participants' pre-test performance on accuracy in the pre-intermediate experimental group (Mean Rank=16.78, N=18) and control group (Mean Rank=17.27, N=15), p= 0.884>0.05. This indicates before treatment the two groups were similar in their oral performance.

Another Mann-Whitney U test was used to compare the groups in pre-intermediate classes based on the post-tests. Table 19 also depicts that this difference is significant meaning that providing oral feedbacks has been effective on learners' oral accuracy. Moreover, as depicted in Table 20 the experimental group has a higher mean rank than the control group on oral accuracy.

Table 19

Descriptive Statistics on Pre-intermediate learners' Oral Accuracy on the Post-tests

Accuracy Post-test Pre-Inter	
Mann-Whitney U	55.000
Wilcoxon W	175.000
Z	-2.896
Asymp. Sig. (2-tailed)	.004
Exact Sig. [2*(1-tailed Sig.)]	.003[a]

[a.] Not corrected for ties.

Table 20

Mann-Whitney U Test Results of Pre-intermediate Learners' Oral Accuracy on the Post-tests

Accuracy Post-test Pre-Inter	N	Mean Rank	Sum of Ranks
Experimental	18	21.44	386.00

Control	15	11.67	175.00
Total	33		

A Mann-Whitney U Test in Table 19 revealed a significant difference regarding participants' post-test performance on accuracy in the pre-intermediate experimental group (Mean Rank=21.44, N=18) and control group (Mean Rank=11.67, N= 15), p= 0.004<0.05. In other words the groups were different after the instructional program meaning that the oral feedbacks were influential. As depicted in Table 20 the mean rank for the experimental group is much higher than the control group depicting the influence of oral feedbacks.

Table 21

Wilcoxon Signed Rank Test on Elementary Learners' Oral Accuracy (Experimental Group)

Accuracy Pre/Post-test Experimental EL	N	Mean Rank	Sum of Ranks
Negative Ranks	5	6.90	34.50
Positive Ranks	12	9.88	118.50
Ties	0		
Total	17		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly oral feedback treatment modes elicited a statistically significant change in participants' performance regarding accuracy (p=0.047<0.05). The ranks Table 21 provides some interesting data on the comparison of participants' performance on before and after the treatment of oral feedback regarding accuracy (error-free clauses). It can be seen from the table's legend that 12 participants had a higher score after the treatment whereas five participants had a higher pre-test score than after their treatment.

Table 22

Wilcoxon Signed Rank Test on Elementary Learners' Oral Accuracy (Control Group)

Accuracy Pre/Post-test Control EL	N	Mean Rank	Sum of Ranks
Negative Ranks	4	8.13	32.50
Positive Ranks	9	6.50	58.50
Ties	3		
Total	16		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly speaking practice without oral feedback modes did not elicit a statistically significant change in participants' performance regarding accuracy (p=0.363>0.05). The ranks table in Table 22 provides some interesting data on the comparison of participants' performance on the pre-test and post-test in the control group regarding accuracy. It can be seen from the table's legend that nine participants had a higher score after receiving no treatment whereas four participants had a higher pre-test score than their

post-test score. Moreover, three participants saw no change in their performance.

Table 23

Wilcoxon Signed Rank Test on Pre-intermediate Learners' Oral Accuracy (Experimental Group)

Accuracy Pre/Post-test Experimental Pre-Inter	N	Mean Rank	Sum of Ranks
Negative Ranks	0	.00	.00
Positive Ranks	16	8.50	136.00
Ties	2		
Total	18		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly oral feedback treatment modes elicited a statistically significant change in pre-intermediate participants' performance regarding accuracy (p = 0.000<0.05). The ranks table in Table 23 provides some interesting data on the comparison of participants' performance on before and after the treatment of oral feedback regarding accuracy. It can be seen from the table's legend that 16 participants had a higher score after treatment, and there was observed no change in the performance of two participants.

Table 24

Wilcoxon Signed Rank Test on Pre-intermediate Learners' Oral Accuracy (Control Group)

Accuracy Pre/Post-test Control Pre-Inter	N	Mean Rank	Sum of Ranks
Negative Ranks	1	5.00	5.00
Positive Ranks	11		
Ties	3	6.64	73.00
Total	15		

A Wilcoxon Signed Rank Test revealed that a-two-month, twice weekly speaking practice without oral feedback modes elicited a statistically significant change in pre-intermediate participants' performance regarding accuracy (p=0.007<0.05). The ranks Table 24 provides some interesting data on the comparison of participants' performance on the pre-test and post-test in the control group regarding accuracy. It can be seen from the table's legend that 11 participants had a higher score after receiving no treatment

whereas one participant had a higher pre-test score than their post-test score. Moreover, there was observed no change in the performance of three participants.

The results of this study indicated that the experimental groups at both elementary and pre-intermediate levels had a significantly higher performance on accuracy after receiving oral feedback modes whereas the control group did not improve much. The significant difference between the experimental and control groups at both levels of proficiency after the treatment provided evidence for the effect of oral feedback modes on oral accuracy.

4.2. Discussion

The present study was an attempt to investigate the effectiveness of the oral feedbacks on Iranian EFL learners' oral complexity (clause per AS) and oral accuracy (error-free clauses). On the basis of the experimental results, participants performed well on their post-tests regarding oral complexity and accuracy. The significant difference depicts that providing oral feedbacks has been effective on learners' oral complexity and accuracy.

The experimental groups in both levels, elementary and pre-intermediate, outperformed the control groups. The participants improved on their post-tests in both levels not only on oral complexity but also on oral accuracy. Thus, regarding the first and second research questions using oral feedback modes (recasts, clarification requests, metalinguistic, praising, elicitation, and repetition) in classroom interactions is a significant contributor to the learners' oral complexity and accuracy.

Another point is that although using oral feedback was found to significantly affect the improvement of both oral complexity and accuracy positively at both elementary and pre-intermediate levels, the influence was more noticeable at the elementary level. Consequently, regarding the third research question it can be stated that lower levels can be more susceptible to providing feedback.

The quality of a learner's performance cannot necessarily be represented by focusing on a single score because it may inevitably fail to indicate the complexity of test performance (Douglas, 1994; Iwashita, Brown, McNamara, Hagan, 2008). Hence, if this complexity is more particularly brought into account by bringing features of the produced language into consideration, the validity of the interpretations of test scores could be improved. Moreover, the results revealed significant differences between groups on the post-tests at both levels. However, the analysis of the discourse produced pointed to an interesting picture and highlighted that learners could produce complex structures and clauses based on their level of

proficiency. On the other hand, the linguistic complexity of L2 production can be negatively affected by interactive tasks because of communicative pressure (Gan, 2012). Test takers are supposed to be ready to respond to their teachers, classmates and incept their own turns at the right moment in interactive tasks. Greater listening, understanding, monitoring, and decision making is needed in interactive and oral tasks. Thus, less complex language will tend to be produced (Robinson, 2001; 2005).

This study indicated that the oral feedback modes were helpful in producing correct and well-organized clauses. Participants tried not to repeat their mistakes that they had during the term and had improved their oral accuracy in their post-tests. Accuracy was the criterion that had a high and significant improvement. The feature of accuracy is more salient than other language features such as fluency and complexity (Ahmadi & Sadeghi, 2016). Accuracy measures may also be explained by the idea of communicative pressure. The production of a less complex language can be due to communicative pressure, and more accuracy probably comes along with less complexity (Skehan, 2009). As participants try to generate more complex utterances, the possibility of making errors may increase as well. The present study indicated that oral interaction favors both accuracy and complexity.

5. Conclusion and Implications

The present study was an attempt to investigate and explain experimentally the possible effects of the oral feedback modes; recasts, clarification requests, metalinguistic, praising, elicitation, and repetition, on Iranian learners' oral complexity (clause per AS) on the one hand, and oral accuracy (error-free clauses), on the other hand. The analysis of the related data resulted in significant findings. They indicated that the feedback modes significantly influenced the learners' oral complexity (clause per AS). Comparing the oral complexity scores, the drawing card was that the learners in the elementary group preformed slightly better on their post-tests. In addition, from the means of the accuracy (error-free clauses) scores, it can be concluded that there were significant differences between post-tests in the experimental and control groups at both elementary and pre-intermediate levels. As a result, based on the participants' oral performance, learners tend to use less complex clauses in order to be more accurate. It seems that in the Iranian context, learners tend to be more concerned about being correct and using the right language rather than using complex clauses with errors. Therefore, teachers can adapt their oral feedback modes to their learners' learning styles. In the light of the results of the present research, teachers should be aware that feedback has different effects on the learners based on their proficiency level. The awareness of the effectiveness of oral feedback

on learners' oral complexity and accuracy can help teachers make more informed decisions as to how to provide oral feedback and how to assess learners' oral performance. The fact that the two features of complexity and accuracy develop at the same time can be beneficial for teachers and language testing, in such a way that when teachers asses learners accuracy, they can understand which complex structures the learners do not use and need more practice on. On the other hand, in language testing complexity should be assessed separately in order to pinpoint learners' reasons of avoidance and help them conquer their fear of using complex structures. It is believed that teachers should treat learners' errors seriously and carefully, so should learners themselves.

Choices need to be made in accordance with linguistic targets, interactional contexts, learners' age and proficiency, and the classroom's communicative orientation and curricular objectives. As Ammar and Spada (2006) concluded, "one size does not fit all" (p. 566).

However, the challenges entailed in designing classroom studies that would account for the multiplicity of variables affecting oral feedback effectiveness in the hubbub of foreign language classrooms have been acknowledged. The replications of observational and quasi experimental classroom studies undertaken can be of particular interest with a view to determining the generalizability of previous studies while further illustrating, in more fine-grained ways, how oral feedback effectiveness is differentially affected by instructional setting.

References

Ahmadi, A., & Sadeghi, E. (2016). Assessing English language learners' oral performance: A comparison of monologue, interview, and group oral test. *Language Assessment Quarterly, 13*(4), 341-358.

Ammar, A., & N. Spada (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition, 28*, 543–574.

British Council (n.d.). *Typical Speaking Tasks.* Retrieved October 13, 2017 from *http://learnenglishteens.britishcouncil.org/exams/speaking-exams/typical-speaking-tasks.*

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing, 26*(3), 341–366.

Brown, H. D. (2007). *Principles of language learning and teaching (5th Ed).* White Plains, NY: Pearson Education, Inc.

Cathcart, R., & Olsen, J. (1976). Teachers' and students' preferences for correction of classroom and conversation errors. In J. Fanselow, & R. Crymes (Eds.), *On TESOL 76* (pp. 41-53). Washington, DC: TESOL.

Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning, 27*, 29-46.

Corder, P. (1974). Idiosyncratic dialects and error analysis. In J. Richards (Ed.), *Error Analysis*. London: Longman.

Dekeyeser, R. (1993).   The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal, 77*(1), 501-514.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing, 11*(2), 125–144.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language.* Oxford: Oxford University Press.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*(2), 339-368.

Erdogan, V. (2005). Contribution of error analysis to foreign language teaching. *Mersin University Journal of the Faculty of Education, 1*(2), 261-270.

Fanselow, J. (1977). The treatment of error in oral work. *Foreign Language Annals, 10*, 583-593.

Foster, P., & Skehan, P. (1996). The influence of planning and task-type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299–323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–375.

Gan, Z. (2012). Complexity measures, task type, and analytic evaluations of speaking proficiency in a school-based assessment context. *Language Assessment Quarterly, 9*(2), 133–151.

Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research, 27*, 3-4, 255-270.

Hendrickson, J. (1978) Error correction in foreign language teaching: Recent theory, research and practice. *Modern Language Journal, 62*, 387-398.

Hepford, E. A. (2017). Dynamic second language development: The interaction of complexity, accuracy, and fluency in a naturalistic learning context. Unpublished Ph.D. Dissertation. Temple University.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency* (pp. 1-21). Amsterdam: John Benjamins.

Hunter, J.D. (2011). A multi-method investigation of the effectiveness and utility of delayed corrective feedback in second language oral

production. Unpublished Ph.D. Dissertation (Module 3), University of Birmingham, UK.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*(2), 217-230.

Iravani, H. (2007). *IELTS Speaking Tests: With Answers & Sample Interviews.* Tehran: Zabankadeh.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49.

Katayama, A. (2007). Learners' perceptions toward oral error correction. In K. Bradford-Watts (Ed.), *JALT 2006 Conference Proceedings.* Tokyo: JALT.

Krashen, S. (1994). The pleasure hypothesis. *Georgetown University Roundtable on Languages and Linguistics*.

Krashen, S. (1985). *The input hypothesis.* London: Longman.

Leaper, D. A., & Riazi, M. (2014). The influence of prompt on group oral tests. *Language Testing, 31*(2), 177–204.

Lee, A. H. (2013). Effects of corrective feedback on l2 speech perception: Perceptual decisions, linguistic hypotheses, and negative evidence. Unpublished MA thesis. McGill University.

Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics, 12*, 180-196.

Li, Sh. (2014). Oral corrective feedback. *ELT Journal, 68*(2), 196-198.

Lightbown, P. M., & Spada, N. (2006). *How languages are learned (3rd Ed.).* Oxford: Oxford University Press.

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in Communicative Language Teaching: Effects on second language acquisition. *Studies in Second Language Acquisition, 12*, 429-448.

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 361-378). Oxford: Oxford University Press.

Long, M. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences, 379*, 259-278.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37-61.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1–40.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction*

in second language acquisition: A series of empirical studies (pp. 408-452). Oxford, UK: Oxford University Press.

Meihami, H., & Rashidi, N. (2018). The effect of topical knowledge on ESP learners' writing quality: complexity, accuracy, and fluency measures. *XLinguae, 11*(4), 45-58.

Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing, 31*(2), 147–175.

Ortega, L. (2012). *Interlanguage complexity: A construct in search of theoretical renewal. Linguistic Complexity: Second language acquisition, indigenization, contact.* Berlin: Walter de Gruyter.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492-518.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*, 590–601.

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics 35*(2), 184-207.

Quinn, P. (2014). Delayed versus immediate corrective feedback on orally Produced passive errors in English. Unpublished Ph.D. Dissertation. University of Toronto.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In R. DeKeyser (Ed.), *Practice in second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141-160). New York: Cambridge University Press.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43*(1), 1–32.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. In J. M. Norris, & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 131–164). The Netherlands: John Benjamins.

Saito, K. (2013). Re-examining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*(1), 1-29.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning 62*(2), 595–633.

Selinker, L. (1974). Interlanguage. In J. Richards (Ed.), *Error Analysis.* England: Longman.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skinner, B. F. (1957). *Verbal behavior.* London: Methuen.

Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: the garden path technique. *Studies in Second Language Acquisition, 11*, 385-395.

Truscott, J. (1999). What's wrong with oral grammar correction? *The Canadian Modern Language Review, 55*, 437-456.

Varnosfadrani, A. D., & Basturkmen, H. (2009). The effectiveness of implicit and explicit error correction on learners' performance. *System, 37*, 82-98.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal, 96*(4), 576-598.

Bibliographic information of this paper for citing: