

The Effect of Portfolio Assessment on Iranian EFL Learners' Expository Writing Accuracy and Complexity

Amir Masoud Shabanloie¹, Alizadeh Khodaverdi²

1, Department of Humanities, Islamic Azad University, Ahar, East Azarbaijan, Iran,
university.ph.d115@gmail.com

2, Department of Humanities, Islamic Azad University, Ahar, East Azarbaijan, Iran,
khodaverdializadeh@gmail.com

1

Abstract

This quasi-experimental study examined the effect of portfolio assessment (PA) on Iranian EFL learners' expository writing accuracy and complexity. It was hypothesized that engaging learners in the process of assessing their performance can focus their attention on formal and textual features and thereby promote their writing. 20 male learners at the intermediate and high level of proficiency in Iran Language Institute (ILI), Khoy Branch who were within the age range of 15-20 were participated in the study and were randomly assigned as the control and experimental groups after their homogeneity was assessed based on a Oxford Placement Test and a writing test. Both groups received one session as pretest and three sessions as posttests. The experimental group was provided with some information about the nature and goal of the portfolio assessment at the beginning of the class and additionally experienced both paper correction by their teacher and portfolio activities such as collecting each participant's papers in personal files and special kind of feedback which is related to their accuracy and complexity. The paired samples t-test analysis of their writing post-test revealed that the experimental group surpassed the control group and produced more accurate but same complex texts. The findings highlight the importance of engaging intermediate and advanced TEFL students not only in the learning process but also in the process of evaluating their own progress over time and offer pedagogical implications.

Key Words: accuracy, complexity, interactive feedback, portfolio Assessment, writing

1. Introduction

Writing is often used as a mode of production for learning, created on the basis of the writer's mastery of grammar and vocabulary. Writing is a difficult skill to master and evaluating or assigning it is even harder to achieve. Unlike the traditional evaluation techniques such as tests which provide a unilateral flow of judgment and mere feedback, portfolios as assessment procedures enable learners to promote their self-evaluation and critical thinking (Venn, 2004). Portfolio Assessment (PA) is not a type of assessment item, but rather a compilation of student work. In the education area specifically in terms of writing assessment, academic portfolios are one type of authentic assessment (Milton & Arend E, 2010). Grab and Kaplan (1996) in their book "theory and practice of writing" agreed that half of the worlds population does not master how to write adequately and effectively. The interactive nature of PA can escalates the complexity and accuracy of the evaluation process which as posed by Karrol (as cited in Reid, 2000), originates from the difficulty and complexity of theoretically defining writing in the first place and the challenge of simultaneously controlling various factors involved in writing and assigning in the second place. PA also contains learners' collection of written text that may be related to their strengths and weaknesses depending on the teachers and learners agreement. In addition the application of correct grammar is an important aspect of any good piece of writing because writing language is grammatically more complex than spoken language. According to these factors, the researcher motivated to study PA in order of examine how its affect learners performance in expository writing accuracy and complexity.

2

2. Review of the Related Literature

As seifoor (2016) noted, PA as a way of facilitating teaching and learning process studied in two major lines of research. The first line is mainly concerned with the effect of portfolio assessment on learners' achievement in writing (Qinghua, 2010; Barootchi&Keshavaraz, 2002; Nezakatgoo, 2011; Song & August, 2002), the second with learners' reflection, comments and attitudes toward portfolio assessment (Starck, 1999; Kear, Coffman, McKenna, & Ambrosio, 2000; Spencer, 1999). The present study is of the first type which is specifically done on L2 learners'. As Hamp-Lyons (2006) asserted, the most of the literature concerning portfolios comes from first language writing and there is rare literature on the use of portfolios for L2 learners in assessment fields. There is also limited number of quantitative research considering the effect of PA on EFL students writing performance descriptors.

Up to now, a number of studies have scrutinized the effect of PA on learning and teaching in general way and few of them were done in a way that they investigate the PA effect on learners writing performance descriptors like accuracy and complexity in more specific way.

In the context of Iran, Seifoori (2016) examined the effect of PA on 40 male and female Iranian post-graduate TEFL students at Islamic Azad University. The researcher employed initially, a modified 40-item version of PET to verify the homogeneity of the participants. For posttest she presented eight genres in the class during the 12-session treatment and the

participants were weekly asked to write compositions based on each of these genres after they were presented in class. The findings revealed superior performance of the experimental group who underwent PA compared to control group in terms of accuracy and complexity of writing. The results conformed the effectiveness of the interactive feedback and learners' engagement in self and peer- assessment activities.

In a longitudinal research, Song and August (2002) examined the effect of portfolio assessment on two groups of advanced ESL learners' composition writing. According to their findings PA group performance was higher than that of the control group not only in both writing and college exit exam. Lately, Qinghua (2010) conducted a quasi-experimental study to investigate the impact of PA on writing development on 34 EFL Chinese learners which age ranged from 18 to 22, in two sophomore English major classes of the same size, gender distribution, and writing proficiency. The results showed that there were significant differences between the two groups in terms of accuracy and coherence with the PA group surpassing their counterparts in the control group.

3. Method

3.1 Participants

3

The participants in the current quasi-experimental study include 20 male learners at the intermediate and high level of proficiency in Iran Language Institute (ILI), Khoy Branch who were within the age range of 15-20. They were selected from two classes (with 10 learners in each class). According to institute's policies, Learners have to take part in placement test in order to be placed in the appropriate level based on their language proficiency. In addition at the end of each term learners take an achievement test in order to be eligible to pass to the next level. Therefore learners in each level are almost at the same level of proficiency. All of the participants were native speakers of Azerbaijani who live in Khoy.

3.2 Instruments

To make sure of participants English level, the researcher used the Oxford Placement Test (Allen, 2001). This test involves 60 multiple-choice items in three sections: vocabulary, grammar, and cloze test. More specifically, it involves 20 vocabulary items, 20 grammar items, and 20 cloze test items. The language learners were required to answer the items of this test during 80 minutes. According to Allen (2001), the results of the statistical analysis have shown that, this test is a reliable test (Cronbach's alpha = .87). Moreover, as he noted, the validity index of the test is .83 which is regarded to be satisfactory.

After the above proficiency test, two instruments were used to collect the research data. First, a writing pretest was administered for both experimental and control groups according to one of the TOEFL topics "Computers and Electronic Tools are Necessary part of Student Learning" to see if the participants were homogenous in terms of the accuracy and complexity of their writing. The practice effect was controlled in two ways. Firstly, researcher had noted the participants that the purpose of the initial writing was to delineate

their entry level in writing and verify the compatibility of the syllabus and that they would not expect any kind of feedback. Further, researcher administered the post-test with a 3-week interval at the end of the treatment during which the participants were working on three different writing genres and assignments relating to TOEFL topics. To relieve the burden on the participants and alleviate their stress, both writing tests were administered during the class time as class work.

3.3. Materials and Procedure

Due to some practical problems, random selection was not possible; thus, two intact classes with 10 learners in each class were selected from ili. Both groups will take one session as pretest to determine their homogeneity. This will be done in format of composition writing according to one topic of the TOEFL test. During the posttests, the teacher provided control group with some explicit instructions on the outline and general format of an essay. At the end of session learners in control group were expected to deliver a composition they wrote according to the TOEFL topics which were selected by their teacher. Learners' composition was given to them after corrected by teacher at the beginning of next session. No portfolio activity like gathering papers or special kind of feedback was done with them. In other words control group received traditional assessment.

4

The PA group (which is our experimental group in this study) on the other hand, was provided with some information about the nature and goal of the portfolio assessment at the beginning of the class. Then they were asked to write one composition according to one of the TOEFL topics selected by their teacher during each session. After that they were experienced both paper correction by their teacher and portfolio activities such as collecting each participant's papers in personal files and special kind of feedback which is related to their accuracy and complexity. This will take three regular session of class time.

In both pretest and posttests the participants were told that their compositions have to be about 100 words in length and they should finish their compositions within 20 minutes of class time. Finally the researcher will used t-test to analyze the data of the study. To ensure the comparability of the participants in the two groups, the Levene's test of equality of variance was run due to the participants in both groups ($F = .120, p > .05$). Although in order to check for the normality of the data, the researcher used Kolmogorov-Smirnov and Shaphiro-Wilk test of normality.

3.4 measures

Grammatical accuracy might be measured as the ratio of error-free terminal units (t-units) or in terms of inaccuracy as the ratio of errors per t-unit (Larsen-Freeman, 2006). A t-unit is defined as each independent utterance providing referential or pragmatic meaning (Foster & Skehan, 1999) and may be made up of one simple independent finite clause or an independent finite clause plus one or more dependent finite or non-finite clauses. In this study, accuracy was quantified as the percentage of error-free clauses in overall writing (Ellis & Yuan, 2004; Foster & Skehan, 1999; Tavakoli & Skehan, 2005; Yuan & Ellis, 2003).



Following Foster and Skehan (1999), complexity was also measured as the ratio of subordinate clauses to the overall t-units produced. These two methods supplanted the levels of the EFL composition profile to address the reliability concern posed by Song and August (2002) who underscored the increased subjectivity as a serious problem in PA.

4. Results

4.1 Proficiency test (Oxford Placement Test)

In order to reassurance of participants English level homogeneity, moreover ili placement and proficiency test, the researcher used the Oxford Placement Test (Allen, 2001). This test involves 60 multiple-choice items in three sections: vocabulary, grammar, and cloze test. More specifically, it involves 20 vocabulary items, 20 grammar items, and 20 cloze test items. To analyze the result of the proficiency test (Oxford Placement test) the researcher conducted an Independent-Samples t-test. The results of descriptive statistics are presented in Table 4.1.

5

Table 4.1 Descriptive statistics of proficiency test (Oxford Placement Test)

Group Statistics					
	group	N	Mean	Std. Deviation	Std. Error Mean
Proficiency	Control group	10	.9450	.01080	.00342
	PA group	10	.9470	.01252	.00396

Table 4.2 Results of the Independent-Samples t-test for proficiency test (Oxford Placement Test)

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	95% Confidence Interval of the Difference	
							Lower	Upper
Proficiency	Equal variances assumed	.213	.650	-.383	18	.707	-.01298	.00898
	Equal variances not assumed			-.383	17.623	.707	-.01300	.00900

According to the Table 4.1, the mean score of individual group is .9450, with a standard deviation of .01252. Based on descriptive statistics, the mean score of both groups performance in Oxford placement Test are very close. In order to determine there was a significance difference between the groups mean score, the results of Independent-Samples t-test are demonstrated in the table 4.2.

As Table 4.2 indicates, there was no significant difference, $t(18) = 0.383$, $P = 0.707 > 0.5$, between the PA and control groups mean scores when the variances are assumed equal ($P = 0.650$). This reflects the homogeneity of both groups at the beginning of the study.

4.2 Descriptive Statistics for pretest and posttests in both groups

The total number of research participants was two classes with 10 learners in each one. In order to conduct a study, one class is selected as PA group and another one as control group. The studied groups were examined during one pretest session and three posttests sessions which is represented in the Table 4.3 below.

6

Table 4.3 Descriptive Statistics for pretest and posttests in both groups

Variable	Group	N	Mean	Std. Deviation	Std. Error of Mean
pretest. Accuracy	Control Group	10	.2940	.01430	.00452
	PA	10	.3010	.01524	.00482
pretest. Complexity	Control Group	10	1.6730	.00949	.00300
	PA	10	1.6770	.01252	.00396
posttest. Accuracy 1	Control Group	10	.3850	.01434	.00453
	PA	10	.2270	.01418	.00448
posttest. Accuracy 2	Control Group	10	.3840	.02011	.00636
	PA	10	.2410	.01729	.00547
posttest. Accuracy 3	Control Group	10	.3830	.01636	.00517
	PA	10	.2390	.01287	.00407
posttest. Complexity 1	Control Group	10	1.6670	.02406	.00761
	PA	10	1.6570	.00949	.00300
posttest. Complexity 2	Control Group	10	1.6780	.01814	.00573

	PA	10	1.6760	.01838	.00581
posttest. Complexity 3	Control Group	10	1.6790	.00994	.00314
	PA	10	1.6840	.01350	.00427

As Table 4.3 indicates, at the pretest stage, the mean score of PA and control group are almost the same in terms of accuracy and complexity, which can demonstrate two groups' accuracy and complexity homogeneity before the treatment. At posttests stages, in term of accuracy, mean score of collaborative group has decreased significantly from 0.3 to 0.22, 0.24, and 0.23 respectively. But mean score of individual group show slight difference. Regarding to complexity there is no significant change between pretest with posttests stages.

4.3 Normality Test

In order to check for the normality of the data, the researcher used the Kolmogorov-Smirnov and Shaphiro-Wilk test of normality which is represented in the Table 4.4 below

7

Table 4.4 results of the normality of the data obtained from the performance of the two groups composition writing on accuracy and complexity pretest and posttests.

group		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Control group	pretest.Accuracy	.210	10	.200*	.955	10	.732
	pretest.Complexity	.224	10	.168	.911	10	.287
	posttest.accuracy1	.164	10	.200*	.968	10	.876
	posttest.accuracy2	.183	10	.200*	.951	10	.680
	posttest.accuracy3	.166	10	.200*	.946	10	.627
	posttest.complexity1	.250	10	.078	.862	10	.081
	posttest.complexity2	.230	10	.144	.922	10	.378
	posttest.complexity3	.240	10	.107	.886	10	.152
PA group	pretest.Accuracy	.226	10	.158	.929	10	.441
	pretest.Complexity	.205	10	.200*	.929	10	.436
	posttest.accuracy1	.211	10	.200*	.937	10	.520
	posttest.accuracy2	.238	10	.115	.907	10	.260
	posttest.accuracy3	.231	10	.139	.924	10	.392
	posttest.complexity1	.224	10	.168	.911	10	.287
	posttest.complexity2	.214	10	.200*	.941	10	.569

posttest.complexity3	.217	10	.200*	.896	10	.198
----------------------	------	----	-------	------	----	------

The results of the Table 4.4 show that, the Kolmogorov-Smirnov and Shaphiro-Wilk test values related to the accuracy and complexity variables in the pretest stage and the three posttest stages are insignificant at the 0.05 level. This shows the normal distribution of the data in all pretest and posttests stages.

4.4 The Impact of Portfolio Assessment

Having verified the initial homogeneity of the PA and control groups and normality of the data, it was safe to analyze the post-test stages scores to answer the research question. Hence, the descriptive statistics of the groups' post-test accuracy and complexity measures were estimated to explore the impact of the PA on the accuracy and complexity of the participants' writing. Table 4.5 and Table 4.6 present the results.

Table 4.5 Results of the Independent-Samplest-testforcomparing means of two groups writing complexity

8

	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
							Lower	Upper
pretest. Complexity	.499	.489	-.805	18	.431	-.00400	-.01443	.00643
posttest. Complexity1	2.362	.142	1.223	18	.237	.01000	-.00718	.02718
posttest. Complexity2	.092	.765	.245	18	.809	.00200	-.01515	.01915
posttest. Complexity3	1.275	.274	-.943	18	.358	-.00500	-.01614	.00614

Table 4.6 Results of Independent-Samples t-test for comparing means of two groups writing accuracy

	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
							Lower	Upper
pretest. Accuracy	.022	.883	-1.059	18	.303	-.00700	-.02088	.00688



posttest.accuracy1	.000	1.000	24.776	18	.000	.15800	.14460	.17140
posttest.accuracy2	.002	.966	17.051	18	.000	.14300	.12538	.16062
posttest.accuracy3	.879	.361	21.875	18	.000	.14400	.13017	.15783

According to Table 4.5, results of levene's test show that the assumption of homogeneity of variances is respected ($P > 0.05$). Although Independent Samples t-test results for equality of variances show that there is no significant difference between the means of complexity at pre-test stages ($t = 0.805$, $sig = 0.431$). Similarly, in posttest stage one ($t = 1.22$, $sig = 0.237$), two ($t = 0.245$, $sig = 0.809$), and three ($t = 0.943$, $sig = 0.358$) there is no significant difference between means of two groups writing complexity.

According to Table 4.6, results of levene's test show that the assumption of homogeneity of variances is respected ($P > 0.05$). Although Independent Sample t-test results for equality of variances show that there is no significant difference between the means of accuracy at pretest stage ($t = 1.05$, $sig = 0.303$). But, in posttest stage one ($t = 24.77$, $sig = 0.001$), two ($t = 17.05$, $sig = 0.001$), and three ($t = 21.87$, $sig = 0.001$) there is significance difference between two groups mean in writing accuracy. Although collaborative pre-writing mean in all three posttest stages is less than individual pre-writing mean in writing accuracy.

In summary, the results of the Table 4.5 using an independent samples t-test show that there is no statistically significant difference between PA expository writing with control group expository writing in terms of writing complexity. Although the results of the Table 4.6 using an Independent Samples t-test show that there is significant difference between PA expository writing and control groups expository writing mean with posttest one, two, and three in terms of writing accuracy. It is showed that PA expository writing mean is significantly lower than control group expository writing mean in terms of accuracy. Therefore the research question is answered in a way that engaging Iranian TEFL learners in PA of their own writing enhance the accuracy of their writing but their complexity stayed the same.

5. Discussion

The present study findings confirmed the greater performance of experimental group who underwent PA compared to the control group in terms of accuracy of writing but showed no significant difference according to complexity of writing. According to Tai (2015), the lack of improvement of complexity might due to the relatively short time of the study. In a meta-analysis conducted by Ortega (2003), syntactic complexity might require up to 12 months of college-level to develop which was less than one month in this study and 18 week in Tai research. Another factor that could cause lack of improvement in syntactic complexity is relatively high proficiency of the participants that in this study learners were already fairly advanced and thus, improvement for them might be harder or take longer to achieve. Furthermore, topic familiarity might also play a part in the lack of complexity improvement (Tedick, 1990). As Skehan (2009) proposed, tasks which are familiar to the learners and

whose structures are clear, such as presenting personal information, lead to higher accuracy and fluency rather than complexity.

In terms of accuracy the findings are compatible with those who reported positive effects from portfolio-based writing assessment on the accuracy and coherence (Qinghua, 2010), and who detected the impact of self-assessment on EFL learners' goal-orientation like (Baleghizadeh & Masoun, 2014). Nezakatgoo (2011) suggested that, evaluating learners' work through portfolio-based assessment could improve their writing on final examinations. Tabatabaei and Assefi (2012), also reported significant impact of PA on the growth of focus, vocabulary, organization, conventions and vocabulary, at upper-intermediate level of proficiency. The results of the study appears to be with the opinions of Seifoori (2016), in a part that PA group had superior performance in terms of accuracy of writing postgraduate and in the process of evaluating their own progress over time and offer pedagogical implications.

The findings of this experimental study show that learners' valued accuracy more than complexity in their expository writings. Moreover, the pattern obtained from researcher results provides further evidence in support of trade-off effects between complexity and accuracy. What seems to be happening here is that subjects were operating undersame information-processing pressure after planning that they had to allocate attention to accuracy at the expense of other goals such as complexity and probably fluency (for Portfolio- based assessment planners).

6. Conclusion

The way portfolio assessment has improved the learners' performance in writing exposition text during this research is through learning the nature and goal of the PA in writing process and from their previous portfolios that evaluated by the researcher as their teacher and self-assessment. The Learners were better in using accurate words into sentences and paragraph in order to convey their ideas, opinions, and feeling because they had already known the steps in writing process and they learned from their previous portfolios that evaluated by the researcher as their teacher and self-assessment. They were motivated to get the progress of their portfolio assessment. They were better in revising and editing their draft to become the best work in every meeting. Using portfolio assessment improved students' performance in content, organization, vocabulary, language use, and mechanics of their text, so we can see that using portfolio assessment affect students' performance in writing exposition text to a great extent. Besides that, the students give positive responses in the using of portfolio assessment to improve their expository-writing performance. This research suggests that the English teacher should give motivation, responsibility, and chance the EFL learners increasing and assessing the amount of choice in learners' writing. The teacher should be a facilitator and have the responsibility to assess all projects given to the students and give an explanation to the students about the way the teacher assesses the students' portfolio. Finally, whatever the method used by the teacher in the writing process, it was important to make clear the explanation, the direction, and the way in assessing the students' draft. The teacher

has to show their appreciation on the students' portfolio by assessing it and show them the criteria or the components that are important in assessing the students' portfolio.

References

- [1] Allen, D. (2001). Oxford Placement Test. New York: EKT Examination Practice Tests.
- [2] Baleghizadeh, S., & Masoun, A. (2014). The effect of self-assessment on EFL learners' goal orientation. *Iranian Journal of Applied Linguistics*, 17(1), 25-48.
- [3] Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolio assessment and teacher-made tests. *Educational Research*, 44(3), 279-288.
- [4] Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59-84.
- [5] Foster, P., & Skehan, P. (1999). The influence of planning and focus of planning on task-based performance. *Language Teaching Research*, 3, 215-247.
- [6] Grabe, W. and R.B. Kaplan . 1996. Theory and practice of writing: Contrastive rhetoric. *Applied Linguistics and Language Study*. C. N. Candlin: Routledge. 263-283.
- [7] Hamp-Lyons, L. (2006). Feedback in portfolio-based writing courses. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 140-161). Cambridge: Cambridge University Press
- [8] Kear, D. J., Coffman, G. A., McKenna, M. C., & Ambrosio, A. L. (2000). Measuring attitude toward writing: A new tool for teachers. *The Reading Teacher*, 54 (1), 10-23.
- [9] Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- [10] Nezakatgoo, B. (2011). Portfolio as a viable alternative in writing assessment. *Journal of Language Teaching and Research*. 2(4), 747-756.
- [11] Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college level L2 writing. *Applied Linguistics* 24 (1), 492-518.
- [12] Qinghua, L. (2010). The impact of portfolio-based writing assessment on EFL writing development of Chinese learners. *Chinese Journal of Applied Linguistics*, 33(2), 103-116.
- [13] Reid, J. (2000). Advanced EAP writing and curriculum design: What do we need to know? In R. Carter, & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp.28-34). Cambridge: Cambridge University Press
- [14] Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics Journal*, 17(1), 38-62.
- [15] Song, B., & August, B. (2002). Using portfolios to assess the writing of ESL students: A powerful alternative? *Journal of Second Language Writing*, 11(1), 49-72.
- [16] Spencer, D. M. (1999). *An exploration of portfolio assessment and its influence on children's writing* (Unpublished master's thesis). Regina University, Canada.



- [17] Starck, T. L. (1999). *Student portfolios: Impact on writing skills and attitudes* (Unpublished doctoral dissertation). Memphis University, Tennessee
- [18] Tabatabaei, O., & Assefi, F. (2012). The effect of portfolio assessment technique on writing performance of EFL learners. *English Language Teaching*, 5(5), 138-147.
- [19] Tai, H. (2015). Writing development in syntactic complexity, accuracy and fluency in a content and language integrated learning class. *International Journal of Language and Linguistic*, 2(3), 149-156.
- [20] Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.239-277). Amsterdam: Benjamins
- [21] Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes*, 9(1), 123-143.
- [22] Venn, J. (2004). *Assessing students with special needs*. Pearson/Merrill/Prentice Hall.
- [23] Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 oral production. *Applied Linguistics*, 24(1), 1-27.