Research paper

The Role of Task Types with Different Task Complexity in EFL Learners' Speaking Performance Complexity, Accuracy, and Fluency across Language Proficiency

Kazem Najmi

Department of Foreign Language, Faculty of Literature, Humanities and Social Sciences, Islamic Azad University, Science and Research Branch, Tehran, Iran

Citation
Najmi, K. (2021). The role of task types with different task complexity in EFL learners' speaking performance complexity, accuracy, and fluency across language proficiency. *Journal of new advances in English Language Teaching and Applied Linguistics*, *3*(2), 629-651.

10.22034/Jeltal.2021.3.2.5

Received
2021-05-20

Revised
2021-08-25

Accepted
2021-10-13

Keywords:
accuracy, assessment, complexity, fluency, task complexity, task types

Abstract

The purpose of the current study was to find out the role of task types with different task complexity in complexity, accuracy, and fluency (CAF) of speaking performance in elementary, intermediate, and advanced EFL learners. In order to achieve the above-mentioned goal, 60 participants were randomly selected from 90 composing 21 elementary, 20 intermediate, and 19 advanced participants through Oxford Placement Test. The selected participants were asked to perform different task types (i.e. personal information exchange, narrative, and decision-making) and learners' speaking CAF was measured and analyzed. These three tasks were chosen because they vary the importance of background knowledge and general familiarity. The sampled participants were asked to perform the prepared tasks in different task complexities (low, mid, and high) and their oral CAF was measured and analyzed. To measure CAF, percentage of error-free C-units for accuracy, clauses per C-units for grammatical complexity, type-token ratio (TTR) for lexical complexity, the number of tokens (words)/total task time (per minutes) for fluency were used. The results of mixed between-within subjects analysis of variance indicated that task types were significant factors on CAF. Furthermore, the results showed that proficiency level was statistically significant on CAF as well. The findings of this study provide pedagogical implications and recommendations for language teachers, material developers, and language assessors.

*__Corresponding Author__: Kazem Najmi
Address: Science and Research Branch, Daneshgah SQ., Hesarak, Tehran, Iran.
Tel: (+98) 21-44865152 **E-mail:** *kazem13najmi@gmail.com*

Introduction

Task Based Language Teaching (TBLT) stems from constructivist learning theories. Accordingly, students are expected to convey meaning and messages rather than a specific

form or pattern. According to constructivist learning theories, a) knowledge is constructed by students based on prior experience and understanding; b) learning is the search for meaning by linking prior knowledge with experience; c) students work actively in groups with self-reflection, and d) teachers are facilitators and co-constructors of knowledge with students through inquiry (Farrell & Jacobs, 2010).

These assumptions indicate that learning is not viewed as a passive process of absorbing information transmitted by a teacher. Rather, learners actively obtain to create sense of experiences and new information through the filter of their functions, interests, prior experiences, and knowledge (Frear & Bitchener 2015). Meanwhile, researchers (Norris & Ortega, 2009; Richards & Rodgers, 2001; Schmidt, 2001) highlighted that TBLT integrates theoretical and empirical foundations for good pedagogy with a focus on tangible learning outcomes in the form of "tasks". Thus, tasks are thought to be the core unit of planning and instruction in teaching. Also, Cook (2013) and Willis (1996) indicate that in TBLT, learning and teaching ought to be organized around communicative tasks are carried out in the target language.

To end with, although different researchers have made different classifications concerning task types, task cognitive complexity, i.e., the degree of a task cognitive involvement, has been considered as the main distinctive factor in determining task types. In this line, Laufer and Hulstijn (2001) proposed their involvement load hypothesis or task-induced involvement stating that word learning and retention depend on the amount of mental effort or involvement a task imposes. Consonant with this hypothesis, it can predict that measures of performance would be the highest in the task type which is the most cognitively demanding one.

What is missing in previous studies of the influence of task type on learners' CAF performance in different proficiency levels. Moreover, a need was felt to double check the learners' levels of CAF in personal information exchange, narrative, and decision-making tasks when there is an elementary, intermediate, and advance level of language proficiency.

Literature Review

Considered as language learning goals, tasks are activities in which a learner engages to gain an objective, and which necessitate the use of language (Kim, Nam, & Lee, 2016) Richards and Renandya (2002) also define a task as an activity that learners carry out using their available language resources and leads to a real outcome. Finally, Samuda and Bygate (2008) consider a task as a holistic pedagogical activity that involves language use and has a pragmatic, non-linguistic outcome. Concerning ways of clustering language tasks, (Long & Porter, 1985) developed the concept of 'task type' that allows for classifying concrete language tasks based on their common characteristics, even though they might be different in detail. Long (1985) believes that while tasks are specified by verb plus noun phrase a task type is specified by a verb alone or by combination of a verb and generic noun phrase. Consequently, his categorization of task types is only based on the feature 'language action'. Considering another perspective, (Johnson & Johnson, 1998) classify tasks according to the extent to which they

are focused on language and communication. Littlewood (cited in Johnson & Johnson, 1998) used the labels 'pre-communicative' and 'communicative' activities to refer to roughly the same distinction between 'skill-getting' and 'skill-using'. Pre-communicative activities include part-skill 'structural' and 'quasi-communicative' activities focusing on the aspects of the target system and their meaning in a way which is clearly language focused. However, communicative activities are referred to those activities that include features of communication. Information-gap 'functional communication' activities or 'social interaction' activities within a simulated social context are considered as examples of communicative activities. To sum up, although some researchers have made different classifications concerning task types, task cognitive complexity, the degree of a task cognitive involvement, has been considered as the main distinctive factor in determining task types. In this line, Laufer and Hulstijn (2001) introduced an involvement load hypothesis or task-induced involvement stating that word learning and retention depend on the amount of mental effort or involvement a task imposes. Consonant with this hypothesis, it can be predicted that measures of performance would be the highest in the task type which is the most cognitively demanding one especially when it deals with the variety of tasks related to the learners' levels of proficiency.

By observing three task types (i.e., personal information exchange, decision making, and narrative tasks) and three levels of proficiency (i.e., elementary, intermediate, and advanced) the present study sought to bridge the gap existing in previous studies concerning the clear role of task types on the participants' oral CAF. So to this aim the following research question was proposed.

RQ: What is the role of task types in complexity, accuracy, and fluency of oral performance in Iranian elementary, intermediate, and advanced EFL learners?

Method

Participants

Sixty EFL learners from a Language Center located in Tehran constituted the participants of the study. To be more convenient for the researcher and to increase the practicality and manageability of the research, all the participants were selected from the mentioned institute. Since the institute placement test may not be valid, Oxford Placement Tests (2004) was administered to the participants and the results were utilized in placing three homogeneous groups of participants, elementary, inter-mediate and advance.

Instrumentations

Oxford Placement Tests (2004)

The Oxford Placement Tests (OPTs) provide a reliable and efficient means of placing learners at the start of a course of a research. The tests have been calibrated against the levels system provided by the Common European Framework of Reference for Language: Learning, Teaching, Assessment (commonly known as the CEF), which has been adopted by the Association of Language Testers in Europe (ALTE) as well as by government and major

institutions, including exam boards, throughout Europe. The OPTs can clearly and reliably identify any learner's CEF level (on the A1 to C2 CEF scale) and also provide a score which will show where the learner is within that band. The test includes 200 questions which are divided into listening test and grammar test which there are 100 questions in each. Both parts were administered prior to the course to guarantee the participants' homogeneity in terms of English proficiency level. OPT was administered to 90 participants to place them on the three proficiency levels, i.e. elementary, intermediate, and advanced. Based on the obtained scores, learners were placed in three levels. Twenty one learners in elementary level, 20 learners in inter-mediate level and 19 learners in advanced level.

Speaking Tasks

In order to have suitable speaking tasks related to the learners' current level of proficiency, available books in the markets were used. In such books, there are different speaking tasks. Since these books, namely American English Files, English Results, Top Notches, and York Mission Possible books, have different levels, it was more convenient to choose the speaking tasks of these books based on learners' current level of proficiency, levels A, B and C. The criteria for the level of books and their tasks were the Common European Framework of Reference. Council of Europe has established it and the aim is to show the level of proficiency. The levels are often utilized to explain one's ability at speaking, reading, writing and understanding a language.

Procedure

Although the Institute itself homogenized learners based on a simple interview, once again the proficiency of the participants was determined by OPT test. Based on the result of this test, participants were divided into three homogeneous levels of elementary or A, intermediate or B, and advanced or C. In each level there were around 20 participants.

The present study aimed to find out the role of task types in CAF of oral performance. So to this aim 27 different tasks were collected and designed. All the 27 tasks are different from each other regarding their level of complexity and type.

To address the research question, the role of task types in CAF of oral performance was found out. So to this aim, three different types of speaking tasks were chosen. The three tasks chosen for this study were similar in types to those used in Foster and Skehan (1996): personal information exchange, narrative and decision making and they were chosen from American English File, English Result, Top Notch and York Mission Possible books and the level of these books are distinguished by the CEFR. Based on the learners' performances on each type of speaking task, CAF was measured. Then the relationships between the three components of CAF within the levels based on the different types of speaking tasks were compared with the relationships between the three components of CAF based on the different types of speaking tasks of other levels.

Accuracy Measures; to measure accuracy in this study, percentage of error-free C-units (Robinson & Gilabert, 2007) was used.

Complexity Measures; to measure grammatical complexity, clauses per C-unit and to measure lexical complexity, TTR (Robinson & Gilabert, 2007) were used.

Fluency Measures; to measure fluency in this study, total number of tokens (words)/total task time (in minutes): WPM (Robinson & Gilabert, 2007) was used.

Results

OPT Homogeneity Test Results

As mentioned in chapter three, OPT was administered to 90 participants to place them on the three proficiency levels, i.e. elementary, intermediate, and advanced. The descriptive statistics, as represented in Table 1, indicates that the mean, median and mode of the OPT scores were 123.26, 119, and 89 respectively. Also Table 1 reflects that the normality of the scores is proved as the significance level for Kolmogorov-Smirnov Z test of normality (p = .28) was greater than .05 and therefore not significant. It shows that the OPT scores have normal distribution.

Table 1
Descriptive Statistics for OPT Score (out of 200)

N	Mean	Median	Mode	SD	Kolmogorov-Smirnov Z	Sig. (2-tailed)
90	123.26	119	89	37.04	.989	.282

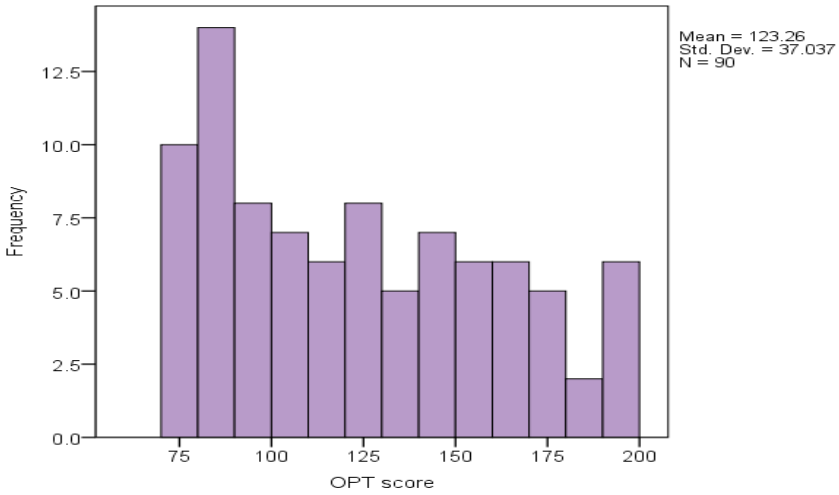
As seen in Table 2, based on the results of OPT, form among 90 students, those students whose OPT scores were between 90 and 119 were chosen as elementary participants ($\bar{x}$ = 103.38, SD = 8.51, n = 21), those students who scored between 120 and 149 were selected as intermediate ones ($\bar{x}$ = 133.30, SD = 9.76, n = 20),, and finally, those students whose scores were between 150 to 189 were considered as advanced participants ($\bar{x}$ = 166.37, SD = 10.74, n = 19)

Table 2
Descriptive Statistics for OPT Score across Three Proficiency Levels

Level	N	Range score	Mean	SD	SEM
Elementary	21	90 – 119	103.38	8.506	1.856
Intermediate	20	120 – 149	133.30	9.761	2.183
Advanced	19	150 – 189	166.37	10.745	2.465

Figure 1 below displays the distribution of the homogeneity test scores on a normal curve.

Figure 1
Histogram of OPT scores



Analysis of the Research Question

The second research question of the present study dealt with the role of task types in CAF of oral performance. In order to answer this research question, mixed between-within subjects analysis of variance (also called split-plot ANOVA design) was applied. According to Pallant (2013, p. 284), there is a situation where we want to combine two approaches of between-subjects design and within-subjects design in our study, with one independent variable being between-subjects (proficiency level in the current study) and the other a within-subjects variable (task type in the current study and research question).

Four mixed between-within subjects ANOVA were performed since there were three factors (i.e. Complexity, accuracy, and fluency) and oral complexity was two-fold grammatical and lexical). All of these factors are explained below:

Grammatical Complexity

In terms of grammatical complexity, task type was the within-subjects variable and proficiency level was considered as the between-subjects variable. Table 3 represents the results of the descriptive statistics for grammatical complexity. According to Table 3, the mean score for oral grammatical complexity gained on decision-making task ($\bar{x}$ = 2.64, SD = .56) is the highest, followed by narrative task ($\bar{x}$ = 2.54, SD = .58), and then personal information exchange ($\bar{x}$ = 2.46, SD = .57). Also, Table 3 shows that the mean score for oral grammatical complexity obtained by the advanced learners is the greatest, followed by the intermediate, and then the elementary learners in the all three task types, i.e. personal information, narrative, and decision-making.

Table 3
Descriptive Statistics for Oral Grammatical Complexity Scores by Task Type and Proficiency Level

Task type	Proficiency level	Mean	Std. Deviation	N
Personal information	Elementary	1.810	.197	21
	Intermediate	2.556	.267	20
	Advanced	3.075	.255	19
	Total	2.459	.575	60
Narrative	Elementary	1.903	.215	21
	Intermediate	2.659	.294	20
	Advanced	3.145	.322	19
	Total	2.548	.585	60
Decision-making	Elementary	2.017	.204	21
	Intermediate	2.722	.266	20
	Advanced	3.242	.303	19
	Total	2.640	.564	60

As Table 4 below displays, the assumption of homogeneity of covariance for conducting ANOVA is violated (Box's M = 44.91, $p < .05$). So the significance level was lowered from .05 to .01 in order to compensate for this shortcoming.

Table 4
Box's Test of Equality of Covariance Matrices for Oral Grammatical Complexity by Task Type and Proficiency Level

Box's M	F	df1	df2	Sig.
44.909	3.454	12	15493.572	.000

The results of Levene's test (Table 5) indicate that the assumption of homogeneity of variance is met as well as the significance value was greater than .05 for all three task types, i.e. personal information exchange, narrative, and decision-making.

Table 5
Levene's Test of Equality of Error Variances for Oral Grammatical Complexity by Task Type and Proficiency Level

Task type	Levene Statistic	df1	df2	Sig.
Personal Information	.880	2	57	.420
Narrative	1.769	2	57	.180
Decision-making	1.088	2	57	.344

As it can be seen in Table 6 below (multivariate tests), the partial eta square index is .41, which shows that 41 percent of the variance in the oral grammatical complexity scores is due to task type; this is quite a large effect size (.41 > .138). The obtained results for Wilks' Lambda ($F (2, 56) = 19.13$, $p < .01$) indicated that there is a significant difference in oral grammatical complexity scores gained on the three task types.

Furthermore, multivariate tests (Table 6) revealed that the interaction effect of Factor * Level is not significant (F (4, 112) = .99, p > .01).

Table 6
Multivariate Tests ANOVA for Oral Grammatical Complexity Scores by Task Type and Proficiency Level

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Factor	Pillai's Trace	.406	19.127	2.000	56.000	.000	.406
	Wilks' Lambda	.594	19.127	2.000	56.000	.000	.406
	Hotelling's Trace	.683	19.127	2.000	56.000	.000	.406
	Roy's Largest Root	.683	19.127b	2.000	56.000	.000	.406
Factor * Level	Pillai's Trace	.014	.196	4.000	114.000	.940	.007
	Wilks' Lambda	.986	.192	4.000	112.000	.942	.007
	Hotelling's Trace	.014	.189	4.000	110.000	.944	.007
	Roy's Largest Root	.011	.301	2.000	57.000	.742	.010

Pairwise comparisons were provided to compare the effect of different task types on oral grammatical complexity. As evident from Table 7, the difference between decision-making and personal information exchange (p = .000, p < .01) is the most significant one, followed by the difference between decision-making and narrative (p = .003, p < .01), and then the between narrative and personal information exchange (p = .006, p < .01).

Table 7
Pairwise Comparisons for the Effect of Task Type on Oral Grammatical Complexity

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.
Personal Information	Narrative	-.088	.031	.006
	Decision-making	-.180	.029	.000
Narrative	Decision-making	-.092	.030	.003

Tests of between-subjects effects (Table 8) detected a statistically significant effect for proficiency level (F (2, 57) = 156.75, p < .01, Eta square= .85) on oral grammatical complexity.

Table 8
Tests of Between-Subjects Effects for Oral Grammatical Complexity Regarding Proficiency Level

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1187.164	1	1187.164	7895.164	.000	.993
Level	47.140	2	23.570	156.752	.000	.846
Error	8.571	57	.150			

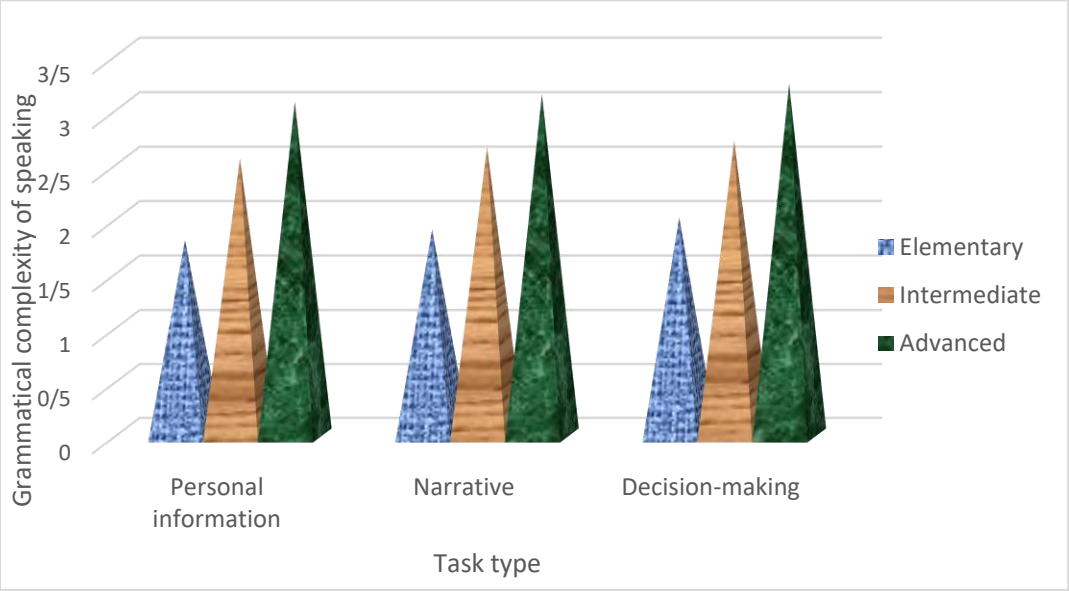
Pairwise comparisons were prepared to compare the effect of different proficiency level on oral grammatical complexity. As seen in Table 9, the difference between the three possible pairs of proficiency levels (p = .000, p < .01) is significant.

Table 9
Pairwise Comparisons for the Effect of Proficiency on Oral Grammatical Complexity

(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.
Elementary	Intermediate	-.736	.070	.000
	Advanced	-1.244	.071	.000
Intermediate	Advanced	-.508	.072	.000

A Bar Graph (Figure 2) was drawn to demonstrate the results. As the Bar Graph shows, oral grammatical complexity mean score has been recorded in the following ascending hierarchical order: personal information exchange tasks, narrative tasks, and decision-making tasks. Meanwhile, regarding proficiency level, following ascending hierarchical order is obvious for oral grammatical complexity mean score: elementary, intermediate, and advanced.

Figure 2
Bar Graph for oral grammatical complexity gained on different task types and proficiency levels



Lexical Complexity

Table 10 represents the results of the descriptive statistics for lexical complexity. According to Table 10, the mean score for oral lexical complexity gained on decision-making task ($\bar{x}$ = 68.28, SD = 8.96) is the highest, followed by narrative task ($\bar{x}$ = 66.78, SD = 9.32), and then personal information exchange ($\bar{x}$ = 65.27, SD = 9.39).

Also, Table 10 shows that the mean score for oral lexical complexity obtained by the advanced learners is the greatest, followed by the intermediate, and then the elementary learners in the all three task types, i.e. personal information, narrative, and Decision-making.

Table 10
Descriptive Statistics for Oral Lexical Complexity Scores by Task Type and Proficiency Level

Task type	Proficiency level	Mean	Std. Deviation	N
Personal information	Elementary	55.001	3.456	21
	Intermediate	65.602	2.472	20
	Advanced	76.283	4.280	19
	Total	65.274	9.393	60
Narrative	Elementary	57.495	4.724	21
	Intermediate	66.801	4.730	20
	Advanced	77.043	5.011	19
	Total	66.787	9.323	60
Decision-making	Elementary	58.405	3.214	21
	Intermediate	68.648	2.986	20
	Advanced	78.823	3.339	19
	Total	68.285	8.962	60

As Table 11 below displays, the assumption of homogeneity of covariance for conducting ANOVA is not violated (Box's M = 16.98, $p > .05$).

Table 11
Box's Test of Equality of Covariance Matrices for Oral Lexical Complexity by Task Type and Proficiency Level

Box's M	F	df1	df2	Sig.
16.980	1.306	12	15493	.207

The results of Levene's test (Table 12) indicate that the assumption of homogeneity of variance is met as well as the significance value was greater than .05 for all three task types, i.e. personal information exchange, narrative, and decision-making.

Table 12
Levene's Test of Equality of Error Variances for Oral Lexical Complexity by Task Type and Proficiency Level

Task type	Levene Statistic	df1	df2	Sig.
Personal information	2.868	2	57	.065
Narrative	.317	2	57	.730
Decision-making	.496	2	57	.612

As it can be seen in Table 13 below (multivariate tests), the partial eta square index is .81, which shows that 81 percent of the variance in the oral lexical complexity scores is due to task type; this is quite a large effect size (.81 > .138). The obtained results for Wilks' Lambda ($F (2, 56) = 119.55$, $p < .05$) indicated that there is a significant difference in oral lexical complexity scores gained on the three task types. Furthermore, multivariate tests (Table 13) revealed that the interaction effect of Factor * Level is not significant ($F (4, 112) = .94$, $p > .05$).

Table 13
Multivariate Tests ANOVA for Oral Lexical Complexity Scores by Task Type and Proficiency Level

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Factor	Pillai's Trace	.810	119.55	2.000	56.000	.000	.810
	Wilks' Lambda	.190	119.55	2.000	56.000	.000	.810
	Hotelling's Trace	4.270	119.55	2.000	56.000	.000	.810
	Roy's Largest Root	4.270	119.55	2.000	56.000	.000	.810
Factor Level *	Pillai's Trace	.064	.946	4.000	114.00	.440	.032
	Wilks' Lambda	.936	.942	4.000	112.00	.443	.033
	Hotelling's Trace	.068	.936	4.000	110.00	.446	.033
	Roy's Largest Root	.064	1.815	2.000	57.00	.172	.060

Pairwise comparisons were provided to compare the effect of different task types on oral lexical complexity. As evident from Table 14, the difference between decision-making and personal information exchange ($p = .000$, $p < .05$) is the most significant one, followed by the difference between decision-making and narrative ($p = .004$, $p < .05$), and then the between narrative and personal information exchange ($p = .009$, $p < .05$).

Table 14
Pairwise Comparisons for the Effect of Task Type on Oral Lexical Complexity

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.
Personal Information	Narrative	-1.484	.551	.009
	Decision-making	-2.997	.197	.000
Narrative	Decision-making	-1.512	.509	.004

Tests of between-subjects effects (Table 15) detected a statistically significant effect for proficiency level ($F (2, 57) = 187.10$, $p < .05$, Eta square= .87) on oral lexical complexity.

Table 15
Tests of Between-Subjects Effects for Oral Lexical Complexity Regarding Proficiency Level

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	809623.106	1	809623.106	24279.029	.000	.998
Level	12478.313	2	6239.157	187.100	.000	.868
Error	1900.756	57	33.347			

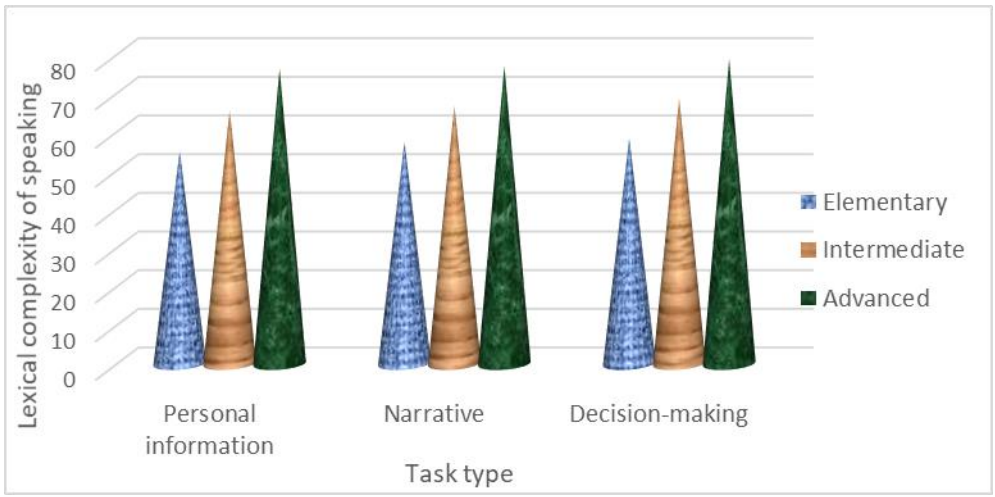
Pairwise comparisons were prepared to compare the effect of different proficiency level on oral lexical complexity. As seen in Table 16, the difference between the three possible pairs of proficiency levels ($p = .000$, $p < .05$) is significant.

Table 16
Pairwise Comparisons for the Effect of Proficiency on Oral Lexical Complexity

(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.
Elementary	Intermediate	-10.050	1.042	.000
	Advanced	-20.416	1.056	.000
Intermediate	Advanced	-10.366	1.068	.000

A Bar Graph (Figure 3) was drawn to demonstrate the results. As the Bar Graph shows, oral lexical complexity mean score has been recorded in the following ascending hierarchical order: personal information exchange tasks, narrative tasks, and decision-making tasks. Meanwhile, regarding proficiency level, following ascending hierarchical order is obvious for oral lexical complexity mean score: elementary, intermediate, and advanced.

Figure 3
Bar Graph for oral lexical complexity gained on different task types and proficiency levels



Accuracy

Related to oral accuracy, task type was the within-subjects variable and proficiency level was considered as the between-subjects variable. Table 17 represents the results of the descriptive statistics oral accuracy. According to Table 17, the mean score for oral accuracy gained on decision-making task ($\bar{x}$ = 60.56, SD = 5.98) is the highest, followed by narrative task ($\bar{x}$ = 59.21, SD = 6.40), and then personal information exchange ($\bar{x}$ = 58.18, SD = 6.50).

Also, Table 17 shows that the mean score for oral accuracy obtained by the advanced learners is the greatest, followed by the intermediate, and then the elementary learners in the all three task types, i.e. personal information, narrative, and decision-making.

Table 17
Descriptive Statistics for Oral Accuracy Scores by Task Type and Proficiency Level

Task type	Proficiency level	Mean	Std. Deviation	N
Personal information	Elementary	51.822	4.074	21
	Intermediate	57.967	2.425	20
	Advanced	65.448	3.333	19
	Total	58.185	6.505	60
Narrative	Elementary	53.341	3.993	21
	Intermediate	59.933	4.904	20
	Advanced	64.924	4.006	19
	Total	59.206	6.403	60
Decision-making	Elementary	54.747	3.216	21
	Intermediate	60.405	2.854	20
	Advanced	67.140	3.460	19
	Total	60.557	5.980	60

As Table 18 below displays, the assumption of homogeneity of covariance for conducting ANOVA is not violated (Box's M = 23.02, $p > .05$).

Table 18
Box's Test of Equality of Covariance Matrices for Oral Accuracy by Task Type and Proficiency Level

Box's M	F	df1	df2	Sig.
23.025	1.771	12	15493.572	.052

The results of Levene's test (Table 19) indicated that the assumption of homogeneity of variance is met as well as the significance value was greater than .05 for all three task types, i.e. personal information exchange, narrative, and decision-making.

Table 19
Levene's Test of Equality of Error Variances for Oral Accuracy by Task Type and Proficiency Level

Task type	Levene Statistic	df1	df2	Sig.
Personal Information	1.554	2	57	.220
Narrative	.934	2	57	.399
Decision-making	1.055	2	57	.355

As it can be seen in Table 20 below (multivariate tests), the partial eta square index is .49, which shows that 49 percent of the variance in the oral accuracy scores is due to task type; this is quite a large effect size (.49 > .138). The obtained results for Wilks' Lambda ($F (2, 56) = 26.86$, $p < .05$) indicated that there is a significant difference in oral accuracy scores gained on the three task types. Furthermore, multivariate tests (Table 20) revealed that the interaction effect of Factor * Level is not significant ($F (4, 112) = 1.42$, $p > .05$).

Table 20
Multivariate Tests ANOVA for Oral Accuracy Scores by Task Type and Proficiency Level

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Factor	Pillai's Trace	.490	26.865	2.000	56.000	.000	.490
	Wilks' Lambda	.510	26.865	2.000	56.000	.000	.490
	Hotelling's Trace	.959	26.865	2.000	56.000	.000	.490
	Roy's Largest Root	.959	26.865	2.000	56.000	.000	.490
Factor * Level	Pillai's Trace	.095	1.424	4.000	114.000	.231	.048
	Wilks' Lambda	.906	1.419	4.000	112.000	.232	.048
	Hotelling's Trace	.103	1.413	4.000	110.000	.234	.049
	Roy's Largest Root	.090	2.573c	2.000	57.000	.085	.083

Pairwise comparisons were provided to compare the effect of different task types on oral accuracy. As evident from Table 21, the difference between narrative and personal information exchange ($p = .06$, $p > .05$) is not significant. However, Table 21 shows that the difference between decision-making and personal information exchange ($p = .000$, $p < .05$) and the difference between decision-making and narrative ($p = .01$, $p < .05$) is significant.

Table 21
Pairwise Comparisons for the Effect of Task Type on Oral Accuracy

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.
Personal Information	Narrative	-.987	.508	.061
	Decision-making	-2.352	.318	.000
Narrative	Decision-making	-1.365	.520	.011

Tests of between-subjects effects (Table 22) detected a statistically significant effect for proficiency level ($F (2, 57) = 85.41$, $p < .05$, Eta square= .75) on oral accuracy.

Table 22
Tests of Between-Subjects Effects for Oral Accuracy Regarding Proficiency Level

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	636727.711	1	636727.711	23128.617	.000	.998
Level	4702.563	2	2351.282	85.408	.000	.750
Error	1569.202	57	27.530			

Pairwise comparisons were prepared to compare the effect of different proficiency level on oral accuracy. As seen in Table 23, the difference between the three possible pairs of proficiency levels ($p = .000$, $p < .05$) is significant.

Table 23

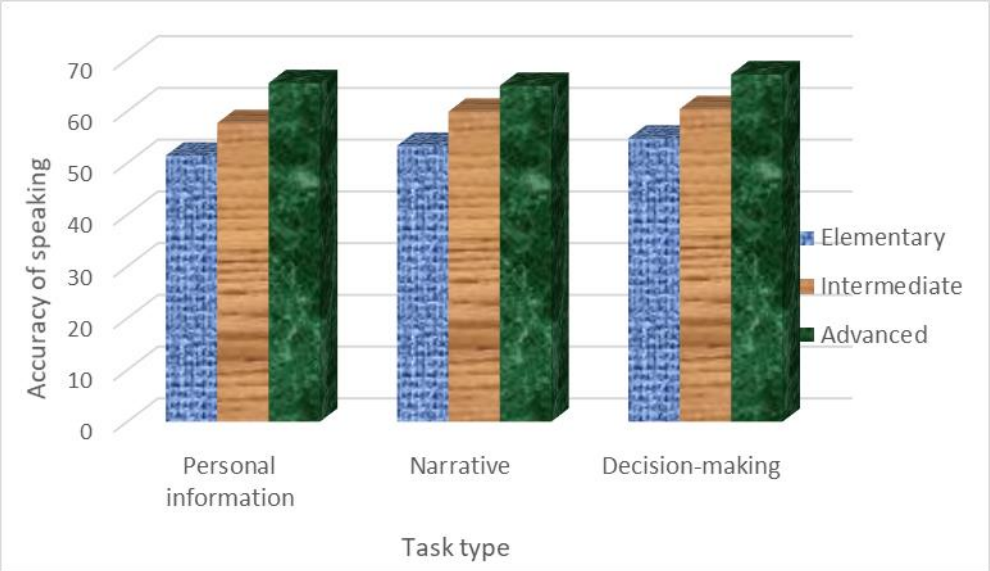
Pairwise Comparisons for the Effect of Proficiency on Oral Accuracy

(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.
Elementary	Intermediate	-6.132	.946	.000
	Advanced	-12.534	.959	.000
Intermediate	Advanced	-6.402	.970	.000

A Bar Graph (Figure 4) was drawn to demonstrate the results. As the Bar Graph shows, oral fluency mean score has been recorded in the following ascending hierarchical order: personal information exchange tasks, narrative tasks, and decision-making tasks. Meanwhile, regarding proficiency level, following ascending hierarchical order is obvious for oral fluency mean score: elementary, intermediate, and advanced.

Figure 4

Bar Graph for oral accuracy gained on different task types and proficiency levels



Fluency

Regarding oral accuracy, task type was the within-subjects variable and proficiency level was considered as the between-subjects variable. Table 24 represents the results of the descriptive statistics for oral fluency. Table 24 represents the results of the descriptive statistics. According to Table 24, the mean score for oral fluency gained on decision-making task ($\bar{x}$ = 78.92, SD = 8.91) is the highest, followed by narrative task ($\bar{x}$ = 77.50, SD = 9.37), and then personal information exchange ($\bar{x}$ = 76.14, SD = 9.48).

Also, Table 24 shows that the mean score for oral fluency obtained by the advanced learners is the greatest, followed by the intermediate, and then the elementary learners in the all three task types, i.e. personal information, narrative, and decision-making.

Table 24
Descriptive Statistics for Oral Fluency Scores by Task Type and Proficiency Level

Task type	Proficiency level	Mean	Std. Deviation	N
Personal information	Elementary	65.791	3.538	21
	Intermediate	76.447	2.470	20
	Advanced	87.244	4.396	19
	Total	76.137	9.484	60
Narrative	Elementary	68.134	4.625	21
	Intermediate	77.518	4.685	20
	Advanced	87.832	5.098	19
	Total	77.499	9.372	60
Decision-making	Elementary	69.126	3.111	21
	Intermediate	79.198	2.913	20
	Advanced	89.449	3.402	19
	Total	78.919	8.911	60

As Table 25 below displays, the assumption of homogeneity of covariance for conducting ANOVA is not violated (Box's M = 21.64, p > .05).

Table 25
Box's Test of Equality of Covariance Matrices for Oral Fluency by Task Type and Proficiency Level

Box's M	F	df1	df2	Sig.
21.639	1.664	12	15493.572	.068

The results of Levene's test (Table 26) indicate that the assumption of homogeneity of variance is met as well as the significance value was greater than .05 for all three task types, i.e. personal information exchange, narrative, and decision-making.

Table 26
Levene's Test of Equality of Error Variances for Oral Fluency by Task Type and Proficiency Level

Task type	Levene Statistic	*df1*	*df2*	*Sig.*
Personal Information	2.881	2	57	.064
Narrative	.375	2	57	.689
Decision-making	.800	2	57	.454

As it can be seen in Table 27 below (multivariate tests), the partial eta square index is .82, which shows that 82 percent of the variance in the oral fluency scores is due to task type; this is quite a large effect size (.82 > .138). The obtained results for Wilks' Lambda (F (2, 56) = 127.38, p < .05) indicated that there is a significant difference in oral fluency scores gained on the three task types. Furthermore, multivariate tests (Table 27) revealed that the interaction effect of Factor * Level is not significant (F (4, 112) = 1.71, p > .05).

Table 27
Multivariate Tests ANOVA for Oral Fluency Scores by Task Type and Proficiency Level

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Factor	Pillai's Trace	.820	127.38	2.000	56.000	.000	.820
	Wilks' Lambda	.180	127.38	2.000	56.000	.000	.820
	Hotelling's Trace	4.549	127.38	2.000	56.000	.000	.820
	Roy's Largest Root	4.549	127.38	2.000	56.000	.000	.820
Factor Level *	Pillai's Trace	.112	1.688	4.000	114.000	.158	.056
	Wilks' Lambda	.888	1.707	4.000	112.000	.154	.057
	Hotelling's Trace	.125	1.724	4.000	110.000	.150	.059
	Roy's Largest Root	.123	3.515c	2.000	57.000	.036	.110

Pairwise comparisons were provided to compare the effect of different task types on oral fluency. As evident from Table 28, the difference between decision-making and personal information exchange ($p = .000$, $p < .05$) is the most significant one, followed by the difference between decision-making and narrative ($p = .006$, $p < .05$), and then the between narrative and personal information exchange ($p = .02$, $p < .05$).

Table 28
Pairwise Comparisons for the Effect of Task Type on Oral Fluency

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.
Personal Information	Narrative	-1.33	.540	.017
	Decision-making	-2.763	.176	.000
Narrative	Decision-making	-1.430	.502	.006

Tests of between-subjects effects (Table 29) detected a statistically significant effect for proficiency level ($F (2, 57) = 185.02$, $p < .05$, Eta square= .87) on oral fluency.

Table 29
Tests of Between-Subjects Effects for oral Fluency Regarding Proficiency Level

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1089371.043	1	1089371.043	32072.638	.000	.998
Level	12568.565	2	6284.283	185.018	.000	.867
Error	1936.047	57	33.966			

Pairwise comparisons were prepared to compare the effect of different proficiency levels on oral fluency. As seen in Table 30, the difference between the three possible pairs of proficiency levels ($p = .000$, $p < .05$) is significant.

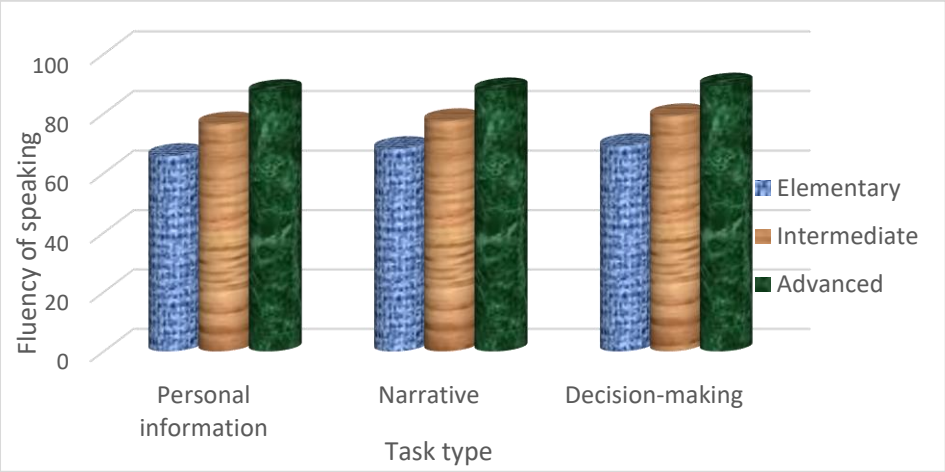
Table 30
Pairwise Comparisons for the Effect of Proficiency on Oral Fluency

(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.
Elementary	Intermediate	-10.037	1.051	.000
	Advanced	-20.491	1.065	.000
Intermediate	Advanced	-10.454	1.078	.000

A Bar Graph (Figure 5) was drawn to demonstrate the results. As the Bar Graph shows, oral fluency mean score has been recorded in the following ascending hierarchical order: personal information exchange tasks, narrative tasks, and decision-making tasks. Meanwhile, regarding proficiency level, following ascending hierarchical order is obvious for oral fluency mean score: elementary, intermediate, and advanced.

Figure 5

Bar Graph for oral fluency gained on different task types and proficiency levels



Discussion

On the basis of the results of statistical analyses, it can be realized how task type affects learners' oral accuracy in a way that they have greater accuracy in the cognitively more complex decision-making tasks. This finding is in accordance with the involvement load hypothesis proposed by Laufer and Hulstijn (2001).They stated that incidental tasks with a higher degree of involvement load are more conducive to the kind of processing that is deemed crucial for learning.

The reason may also lie in Kuiken and Vedder's (2007) argument stating that "task complexity does have an effect on linguistic performance, in the sense that an increase in cognitive task complexity leads to a more accurate text, suggesting that students pay more attention to language form" (p.130). Further, the findings emerging from this study are is in line with Guerrero's (2005) study and Skehan and Foster's (1997) work which showed greater accuracy for a task with a clear inherent structure like the decision-making task in the present study.

In the final place, considering the effect of participatory structures during the personal and decision-making tasks separately this is somehow inconsistent with Foster and Skehan (1999) reporting more accurate performance for teacher-fronted planners than group and solitary planners. However, these findings do not seem to support the studies conducted by Rahmanian (2004), Jafari (2006), and Rezazadeh, Tavakoli, & Eslami Rasekh (2011) since they reported low cognitively demanding tasks are more effective in promoting accuracy. Consequently, it appears that this issue demands further exploration in other EFL contexts.

On the basis of the results of statistical analyses, it can be realized how task type affects learners' oral complexity, accuracy, and fluency. This finding is in accordance with the involvement load hypothesis proposed by Laufer and Hulstijn (2001).They stated that incidental tasks with a higher degree of involvement load are more conducive to the kind of processing that is deemed crucial for learning. Further, the findings emerging from this study are is in line with some researchers' work like Gilabert (2005) and Skehan and Foster (1997) which showed greater accuracy for a task with a clear inherent structure like the decision-making task in the present study. However, these findings do not seem to support the studies conducted by Jafari, (2006); Rahmanian, (2004); Rezazadeh, Tavakoli, and Rasekh, (2011) since they reported low cognitively demanding tasks are more effective in promoting accuracy. Consequently, it appears that this issue demands further exploration in other EFL contexts. However, the findings of the present study ran against the results of other studies like Ishikawa (2006). They argue that the results support Skehan and Fosters' (2001, p: 193) preposition that "prioritization or predisposition (or both) seem to orient performance toward one (or two) of the three areas of accuracy, fluency, and complexity".

Task type had a positive effect on CAF of learners' speaking performance either. The mean score for CAF gained on decision-making task was the highest, followed by narrative task and then personal information exchange. There was also a significant positive relationship between oral grammatical complexity and oral accuracy and a significant positive relationship between oral lexical complexity and oral accuracy. Furthermore, a significant positive relationship between oral grammatical complexity and oral fluency and a significant positive relationship between oral lexical complexity and oral fluency were found out in this study. Additionally, the findings revealed a significant positive relationship between oral fluency and accuracy.

Conclusion

The work presented in this study on CAF presents new perspectives on the empirical study of CAF in SLA, as well as raising important theoretical and methodological questions. Essential to these questions is the need to further clarify testing instruments and better define the constructs to be measured as well as learner internal and external factors surrounding, affecting and perhaps impeding the development or manifestation of CAF in L2 performance. These are all issues for further exploration. It is hoped that this study will contribute to further debate on CAF, shedding light on existing theoretical and methodological issues in the field as well as opening up new areas of inquiry.

This study largely attempted a quantitative exploration of the data, and hence lacked a qualitative account of the phenomena identified in the data set. This study itself was of an exploratory nature rather than explanatory and its aim was related more to unveiling the nature of the relationship rather than finding explanations to account for such relationships. However, a future expansion of the study may be attempted by supplementing it with a qualitative account of the phenomena found in the data. Qualitative account will enrich the data and open grounds for further analysis and discussions. It is also important to conduct research on other examples

of these three task types, as well as on other task types, and to explore the predictability of the language characteristics associated with such tasks.

Clearly, there have been some limitations in need of recognition by researchers and those involved in speaking testing. Making learners speak English may be influenced by affective filters (e.g. stress, shyness, etc.) or, if not well justified, may bore them. As a result, data may express the learners' unwillingness to keep the conversation. So there is a need for further studies in related fields prior to any generalization in the results.

Foreign and second language teaching has a long history and different methods and approaches have been devised throughout to teach language depending on how language has been viewed. In 1950's for example, language was regarded as a system consisting of discrete parts and rules that were to be taught explicitly in a decontextualised way. Later, it was argued that learning a second language like acquiring mother tongue goes through several stages; and comprehension is needed for meaningful and real use of language in target situations. More recently, Task Based Language Teaching (TBLT) has come to be recognized as a communicative approach to language teaching attracting numerous language teachers and researchers. However, within the framework of TBLT, teaching methods teachers utilize in different classes different classroom activities in terms of types and amounts of interaction, goals of language learning and teaching, etc. Some tend to use language in the form of isolated sentences while others prefer to use language as a whole, such as using tasks as activities with special predetermined goals for meaningful use of language (Ellis, 2003).

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35-59.

Bachman, L. F. (2004). *Statistical analyses for language assessment book*: Cambridge University Press.

Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing, 16*(2), 97-110.

Bradlow, A. R., Kim, M., & Blasingame, M. (2017). Language-independent talker-specificity in first language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America, 141*, 886–899.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses speed and repairs. *Language Testing, 30*, 159–175.

Cohen, L. M., & Manion, L. (1998). L.(1989) Research Methods in Education. *New York: Routledge*.

Cook, V. (2013). *Second language learning and language teaching*. New York: Routledge.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*: Oxford University Press Oxford.

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*, 45-57.

Huensch, A., & Tracy–Ventura, N. (2016). Understanding second language fluency behaviour: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, *38*, 755– 785.

Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, *101*, 275–293.

Hunter, A.–M. (2017). *Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency*. (Unpublished doctoral dissertation). University of Surrey, Guildford, UK.

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *Journal of Asia TEFL, 3*(4).

Jafari, S. (2006). *The effects of planning condition, task structure, and gender on different aspects of Iranian intermediate EFL learners' performance (fluency, accuracy, complexity, and lexical density) in written tasks*. Unpublished doctoral dissertation, Tarbiat Modares University, Tehran, Iran.

Kim, Y., Nam, J., & Lee, S.-Y. (2016). Correlation of proficiency with complexity, accuracy, and fluency in spoken and written production: Evidence from L2 Korean. *Journal of the National Council of Less Commonly Taught Languages, 19*, 147-181.

Kormos, J., & Dénes, M. (2006). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*, 145–164.

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143–170). Philadelphia/Amsterdam: John Benjamins.

Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

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

Long, M. H., & Porter, P. A. (1985). Group work, inter-language talk, and second language acquisition. *TESOL Quarterly, 19*(2), 207-228.

Mizera, G. J. (2006). *Working memory and L2 oral fluency*. University of Pittsburgh.

Najmi, K., Naserbakht, M. (2021). The effect of "Drops" application as gaming on vocabulary learning of Iranian EFL young learners. *Journal of new advances in English Language Teaching and Applied Linguistics*, 3(1), 518-533.

Najmi, K., Siyyari, M., Abassian, G. (2020). The Relationship between Three Components of CAF in Learners' Oral Performance at Different Proficiency Levels. *Journal of English Language Pedagogy and Practice*, 13(26), 254-284.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

Pallant, J. (2013). *SPSS survival manual*: McGraw-Hill Education (UK).

Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *Modern Language Journal*, *102*, 676– 692.

Rezazadeh, M., Tavakoli, M., & Eslami Rasekh, A. (2011). The role of task type in foreign language written production: focusing on fluency, complexity, and accuracy. *International Education Studies*, *4*(2), 169-176.

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching*. Cambridge: Cambridge University Press.

Robinson, P. (2003). The cognitive hypothesis, task design, and adult task-based language learning. *University of Hawai'I Second Langauge Studies Paper 21 (2)*.

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (Vol. 2): John Benjamins Publishing.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics*, *30*(4), 533-554.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*: Springer.

Schmidt, R. (2001). Attention'in P. Robinson (ed.): Cognition and Second Language Instruction. *Cambridge, UK: Cambridge University Press*.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Segalowitz, N., French, L., & Guay, J.–D. (2017). What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion? *Canadian Journal of Applied Linguistics*, *20*, 90–116.

Seyyedi, K., Ismail, S., & Mohamed, A. (2014). The Effect of Task Structure on Second Language Learner's Narrative Writing Performance. *Journal of Foreign Languages, Cultures and Civilizations*, *2*(1), 41-53.

Shiriyan, Z., & Nejadansari, D. (2014). The effect of literature-response activities on the complexity, accuracy, and fluency of Iranian EFL learners' L2 oral productions. *Journal of Applied Linguistics and Language Research*, *1*(2), 12-26.

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

Skehan, P. (1998b). Task-based instruction. *Annual review of applied linguistics*, *18*, 268-286.

Skehan, P. (2009a). Lexical performance by native and non-native speakers on language-learning tasks. In *Vocabulary studies in first and second language acquisition* (pp. 107-124): Springer.

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

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. *Processing perspectives on task performance*, *211*, 211-260.

Skehan, P. (2015). Limited attention capacity and cognition. *Domains and Directions in the Development of TBLT*, *8*, 123-155.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language teaching research*, *1*(3), 185-211.

Tavakoli, P., & Foster, P. (2008). Task design and L2 performance. *Language Learning*, *58*(2), 429-473.

Tavakoli, P., & Hunter, A.–M. (2018). Is fluency being "neglected" in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22,* 330–349.

Willis, J. (1996). A flexible framework for task-based learning. In J. Willis and D. Willis (Eds.), *Challenge and change in language teaching* (pp. 52-62). Oxford: Macmillan

Heinemann.Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied linguistics, 24*(1), 1-18.