

The Potential Combined Effects of Task Complexity and Planning Types on Iranian EFL Learners' Oral Production Performance

Sanaz Mohazabieh¹, Rahman Sahragard^{2*}, Ehsan Rassaei³, Mustafa Zamanian⁴

¹ PhD Candidate, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

² Professor, Department of English Language and Linguistics, Shiraz University, Shiraz, Iran

³ Associate Professor, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

⁴ Assistant Professor, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Received: 2020/01/31

Accepted: 2020/08/05

Abstract: This study examined the combined effects of two task complexity levels (i.e., high- and low-complex) and two planning conditions including pre-task planning and on-line planning on Iranian intermediate language learners' speech production regarding complexity, accuracy, and fluency. To this end, 90 intermediate EFL learners from a language institute in Shiraz were randomly assigned into two control and four experimental groups. At first, the language learners in all groups participated in the speaking pretest. Presented with a series of picture description tasks, the participants were asked to narrate a story. During 10 treatment sessions of picture description task performance, the experimental and control groups attempted different planning time conditions including pre-task planning, online planning, and no-planning along with task complexity levels. Finally, following the last session, the posttest was administered to all participants. The narrations analysis, as well as the results of Mixed between-within groups ANOVAs and a series of one-way ANOVAs, manifested that language learners in the pre-task high complexity group outperformed all other groups in terms of complexity. Moreover, the online low complexity group and online high complexity group outperformed the pre-task planning low complexity, no-planning low complexity, and no-planning high complexity groups regarding accuracy. With regard to fluency, the pre-task planning low complexity group significantly outperformed the no-planning low complexity, no-planning high complexity, and online high complexity groups. It was also concluded that pre-task planning affected language learners' speech fluency. The implications of the results are also addressed.

Keywords: Task Complexity, Planning, Language Learners, Oral Production, CAF.

* Corresponding Author.

Authors' Email Address:

¹ Sanaz Mohazabieh (sanazmohazabieh@gmail.com), ² Rahman Sahragard (rahmansahragard@gmail.com), ³ Ehsan Rassaei (ehsanrassaei@yahoo.com), ⁴ Mustafa Zamanian (mustafazamanian@yahoo.com)

ISSN (Online): 2322-5343, ISSN (Print): 2252-0198 © 2020 University of Isfahan. All rights reserved

Introduction

Given the saliency and currency of task-based instruction in language classroom settings, there has been a substantial increase in the number of research studies probing into different aspects of tasks and their impacts on language learners' oral task performance. Task type and task complexity among other features are instrumental in language learners' oral productions concerning complexity, accuracy, and fluency (CAF). Ellis (2009) holds that the investigation into the impacts of different planning conditions on task performance can inform EFL teachers practicing task-based instruction as to whether or not to provide language learners with time for planning. Based on Ellis (2005), planning time within task-based instruction entails pre-task and within-task planning or online planning.

From a cognitive perspective to task-based instruction, task complexity is conceptualized as any information-processing demands (i.e. memory, reasoning, and attention enforced on the task-performers) by different degrees of definitive task structure (Robinson, 2001b). Lui and Li (2012) define task complexity as the aggregation of any inherent task characteristic affecting task performance. This inclusive definition highlights that task complexity is multifarious, task-dependent, and assessed on the basis of its effect on task performance and the language learner's behavior and perspective (Awwad, 2017). Within task-based language teaching research, however, task complexity is viewed as the amount of attention language learners require performing a task to reach an outcome (Skehan, 2001). It is also advocated that cognitively demanding tasks in terms of their content are likely to deflect attentional resources away from language structures (Skehan & Foster, 2001).

There is a rich portfolio of studies on planning within task-based language teaching. Likewise, task complexity has been extensively touched upon with respect to oral and written productions in the context of language teaching. Notwithstanding innumerate studies on different planning types and task complexity, little research has thus far addressed the combined effects of planning and task complexity on language learners' oral performance in terms of CAF. To fill this gap, the current study seeks to examine the joint effects of task complexity and planning type on EFL learners' oral productions. More specifically, our study is guided by the following research question:

What are the combined effects of task complexity and planning types on Iranian EFL learners' oral production performance concerning a) fluency, b) accuracy and c) complexity?

The current study is motivated by the growing interest in employing tasks as effective

learning tools in task-based language teaching. The findings can redound to EFL teachers and syllabus designers to consider the difficulty level of tasks to appropriately match them with language learners' proficiency level. Moreover, the study may conduce to stakeholders (i.e. speaking examiners) in devising a speaking marking scheme wherein task complexity and planning conditions are collectively taken into account.

Review of the Related Literature

Planning

Based on Ellis (2005), there are essentially two types of task-based planning, i.e., pre-task planning and within-task planning that can be distinguished with respect to the time of planning. Pre-task planning occurs prior to task performance and entails strategic planning and rehearsal. Strategic planning prepares language learners to undertake the task by focusing on the content which is to be encoded and communicated. Rehearsal planning concerns task repetition in which the first task performance makes language learners prepared for upcoming task performance. Within-task planning or online planning refers to the time given to language learners to prepare what to say while performing the task. The amount of time during this planning type is a function of task performance under unpressured or pressured conditions with the former denoting that language learners have the opportunity to prepare what to say under no time limitations, while the latter implies that language learners are given limited time to plan their utterances undertaking the task.

Task Complexity

Task complexity is conceptualized as the cognitive features of a task which can increase or decrease cognitive demands imposed on learners (Robinson, 2001b, 2005). Based on this definition, task complexity is characterized by various dimensions which can be subjected to manipulation in creating materials for language learners (Zarei, 2013).

In Robinson's (2007) Triadic Componential Framework, task complexity factors can be categorized into two groups (i.e. resource-directing and resource-dispersing) with respect to cognitive resources including attention and memory. Resource-directing factors make cognitive demands on attention and memory resources turning attention to linguistic aspects, while resource-dispersing factors make performative and/or procedural demands on attention and memory resources. Examples of the former are +/- here and now and +/- reasoning demand, whereas the latter include +/- planning, +/- single task. Task complexity variables can be

conceptualized as dimensions, plus or minus a feature, and continuums, along which more of a feature lies (Robinson, 2001a). Robinson (2010) is of the opinion that increasing the level of task complexity in resource-directing dimensions can promote learners' focus on speech leading to a complex syntactic language.

Models of Task Complexity

Trade-off Hypothesis

In this model, Skehan posited that cognitively complex tasks lead to trade-off effects among three linguistic elements of production (i.e. CAF given the limited attentional resources). Consequently, accuracy and complexity are viewed as competing dimensions of the task performance in which one dimension captures less attention than the other (Skehan, 1998, 2001, 2003, 2014; Skehan & Foster, 2001). Based on this model when the cognitive complexity of the task is increased, the language learner is more likely to attract more attention to the negotiation of meaning and thereby promote their fluency to successfully reach the task goal (Izadpanah & Shajeri, 2016). The main prediction of this model is that attentional limitations for the second language learner make various performance dimensions outdo the other for the resources available (Skehan & Foster, 2001). Undertaking a complex task leads to trade-off effects between form on the one hand and fluency and meaning on the other. Given the limited attentional capacity for form, a trade-off is made between accuracy and linguistic complexity (Michel, 2011).

The Cognition Hypothesis

Based on Robinson's Cognition Hypothesis, increasing task complexity will result in better linguistic performance and production which is linguistically more accurate, syntactically more complex, and lexically more diverse (Kuiken & Vedder, 2011). The Cognition Hypothesis proposes a multiple-resources approach wherein language learners attend to different dimensions of language while performing a cognitively demanding task (Robinson, 2007, 2011). The main premise of the Cognition Hypothesis is that the increase in task complexity can account for syllabus design as well as task sequencing. As the cognitive demands of tasks are increased, attentional resources are increasingly involved (Lee, 2018).

CAF

Language proficiency in the second language is primarily discussed with respect to CAF (Ellis, 2003, 2008). In fact, these aspects have established a triad framework to examine and assess second language output and language proficiency.

Complexity refers to the learner's capacity to produce complex structures that may not be appropriately controllable (Skehan & Foster, 1999). It is associated with the organization of speech production, namely elaborate language, and a wide range of syntactic structures (Foster & Skehan, 1996).

Accuracy is the ability to deliver error-free language performance (i.e. avoiding challenging forms while speaking a target language) showing higher degrees of control in the use of language (Skehan & Foster, 1999). Moreover, accuracy encompasses the correctness and acceptability of second language learners' speech patterns (Bulte & Housen, 2012).

Fluency is defined as a language learner's ability to produce language in real-time without inordinate pauses (Skehan & Foster, 1999). According to Bulte and Housen (2012), several scholars claim that fluency embraces three main aspects including speed fluency, breakdown fluency, and repair fluency.

Empirical Studies

There is a handful of research (e.g., Ahmadian, Tavakoli, & Vahid Dastjerdi, 2012; Moattarian, Tahririan, & Alibabae, 2019; Nasiri & Atai, 2017) investigating the CAF triad with respect to planning time conditions and task complexity in oral production in task-based instruction research. Most of these studies (e.g., Baleghizadeh & Nasrollahi Shahri, 2017; Gilabert, 2007; Khoram, 2019; Yuan & Ellis, 2003), however, have addressed either task complexity or planning time in terms of performance dimensions.

Moattarian et al. (2019) investigated the impact of task complexity, collaborative pre-task planning, and proficiency on EFL learners' interactions. The participants of the study (n=128) were from two different language proficiency levels who were required to carry out three different tasks. The researchers in the study carefully analyzed the language learners' interactions quantitatively and qualitatively. The findings suggested that cognitively complex tasks offered more learning opportunities.

Baleghizadeh and Nasrollahi Shahri (2017) investigated the impact of online planning, rehearsal, and strategic planning on the CAF of oral productions of 40 low and intermediate level EFL learners who carried out picture description tasks in three conditions including a first pre-task planning and a second pre-task planning condition as well as an online planning

condition. Their results demonstrated that rehearsal and strategic planning significantly impacted fluency. However, they did not affect accuracy and complexity. Besides the impact of language proficiency, the findings related to task complexity showed a significant pattern of interaction.

Nasiri and Atai (2017) examined the combined impacts of no planning, strategic planning, and online planning on the CAF of 80 advanced language learners' oral production performing simple and complex narrative tasks. Based on their findings, no planning in both tasks was found to be the least effective. Strategic planning helped the participants significantly enhance their complexity as well as fluency in simple tasks. Their fluency significantly improved in the complex task. Moreover, online planning significantly promoted their accuracy in simple and complex tasks. Joint planning led to the development of accuracy and complexity in the complex task, on the one hand, fluency and accuracy in the simple task on the other. Concerning the impact of task complexity, the interaction between task complexity and the CAF turned out to be significant. However, our study differs from this study in a number of ways. For example, their participants were advanced language learners, while intermediate language learners were involved in our study. Likewise, they considered joint planning in other terms (online and strategic planning combined) in addition to other planning conditions mentioned above.

Conducting a study on 40 high-school students with low-intermediate proficiency levels, Ryu (2017) asked them to describe four sets of pictures in various conditions including simple no planning, complex no planning, simple planning, and complex planning. The findings revealed that task complexity positively influenced syntactic complexity and accuracy. Task complexity negatively impacted lexical complexity and fluency with the decrease in lexical complexity. Concerning planning, syntactic complexity, and fluency were significantly higher, whereas lexical complexity and accuracy were found to be negatively impacted. Likewise, task complexity and planning were found to impact different elements of CAF. As evident, pre-task planning and within-task planning were not addressed.

Yuan and Ellis (2003) investigated the impacts of pre-task and online planning on oral productions of 42 students majoring in English in a Chinese university. The three groups (pre-task planning, online planning, and no planning) undertook an oral narrative elicited through a series of pictures in the planning conditions in question. Their results indicated that pre-task planning improved complexity, while online planning impacted accuracy and grammatical complexity. Further, pre-task planners generated more fluent language than online planners.

Gilabert (2007) studied the simultaneous manipulation of task complexity along with planning time on L2 narrative oral productions. The participants of his study were 48 lower intermediate university students at Ramon Llull University in Barcelona. The findings of his study demonstrated that simple and complex narrative tasks undertaken under planned conditions elicited more lexically complex oral discourse as well as focused attention to form, with fluency being impacted negatively.

Method

Design

This study featured a quasi-experimental design in which the participants were non-randomly selected and homogenized based on their proficiency level and then they were randomly assigned into four experimental and two control groups. This design was adopted as it was not feasible to randomly choose the participants and to place them in specific classes for ethical, practical, and time constraints.

Participants

An initial number of 110 Iranian male and female English learners whose ages ranged from 16 to 45 were selected through convenience sampling from all intermediate language classes in a language institute in Shiraz, Iran. Out of the total of 130 intermediate language learners who enrolled in English classes, 110 learners filled out consent forms and were willing to take part in the study. To ensure that all the language learners enjoyed the same proficiency level (i.e., the intermediate level of English in this study), the Oxford Placement Test was administered to 110 learners in all intermediate classes. Next, 90 language learners (53 females and 37 males) obtaining scores within plus or minus one standard deviation of the mean score for intermediate proficiency were selected for the study.

Instrumentation

Oral Presentation Tasks

Oral presentation tasks constituted the first means of data collection for this study. The participants were asked to narrate a story according to a series of pictures presented to them. To investigate the combined impacts of planning type and task complexity on speaking CAF, a series of scrambled pictures was selected for high complex task groups. Low complex task groups, in contrast, received unscrambled pictures. However, several researchers (e.g., Ellis &

Yuan 2004; Ishhikawa, 2006) have employed pictures for narrative tasks as they are more cognitively demanding than other tasks (Skehan & Foster, 1997).

Proficiency Test

To homogenize the language learners regarding their language proficiency levels, the Oxford Placement Test was administered to language learners in all intermediate classes of the institute. All the participants had started their English learning procedure from elementary levels at this institute. However, to make sure that all the participants were homogeneous concerning the proficiency level, the placement test was administered. Language proficiency levels of EFL learners have also been controlled in similar studies (e.g., Farrokhi, & Sattarpour, 2017; Gilabert, 2007; Salimi, 2015). This facilitated the comparison of published results across related research.

Measures of Learners' Oral Production

Fluency

Repair fluency was considered for the purpose of this study. It was measured by counting the number of repeated words or phrases, false starts (incomplete utterances), phrases or clauses repeated with some syntactical, morphological modifications (reformulations), and replacements of some lexical items for others (Elder & Iwashita, 2005; Skehan & Foster, 1999).

Accuracy

Accuracy refers to the ability to generate error-free utterances (Housen & Kuiken, 2009). To measure accuracy, the researchers in the present study estimated the number of error-free clauses and divided them by the total number of clauses. To this end, all lexical, syntactical, and morphological errors were counted. Evidently, high mean scores are indicative of a fewer number of errors and better performance. This measure was also employed in some previous research studies (e.g., Yuan & Ellis, 2003).

Complexity

In this study, complexity was estimated through the number of clauses per C-unit (i.e., Communication-Unit). It was determined by dividing the number of clauses in the participants' oral production by the number of C-units displaying independent utterances indicative of referential or pragmatic meaning (Foster & Skehan, 1996).

Procedure

To answer the research question, the researcher selected intermediate classes in a language institute in Shiraz. The Oxford Placement Test was employed to ensure the groups' homogeneity. The researcher selected the students obtaining scores between one SD above and below the mean which turned out to be 90 learners for the study. Then, two and four groups, 15 participants each, were randomly selected as the control and experimental groups, respectively. A pretest (i.e. the monologic narrative task) was planned to take place to measure their speaking ability at the initial stage. It was followed by 10 sessions of treatment concerning different task complexity levels and planning conditions. The control groups, including a simple task group and a complex task group, received no treatment (planning), while the experimental groups received treatment in terms of task complexity (low and high) and planning type (online planning and pre-task planning).

The participants were informed that the narrative tasks and the tests were for research purposes. Further, they were assured that the data obtained would not be used for the end-of-course grades. However, the purpose of the research was not precisely clarified to avoid participant bias and to reduce the Hawthorne effect.

Picture narration tasks utilized in the study were in accordance with Robinson's task complexity criteria. In cognitively complex tasks, language learners were required to find the right order of the pictures to narrate them. Not only had the narrative comic strips the possibility of being interpreted differently by different participants, but also in terms of complex task groups, unscrambling the pictures added to the complexity of them. Moreover, they needed various degrees of attention on language learners with less known and predictable information leading to an increasingly cognitive load and consequently impacting the task performance (Foster & Skehan, 1996).

The narrative tasks employed for the treatment and the pretest and posttest narratives were a series of picture strips selected from Quino an Argentinian Cartoonist. These tasks were selected for two reasons. First, similar types of tasks were employed in other studies (e.g., Abdoahzadeh & Fard Kashani, 2012; Heidari-Shahreza, Dabaghi, & Kassaian, 2011; Kim, 2009; Nuevo, 2006; Robinson, 2001a) making the comparison of oral performance results easier and more reliable. Second, these tasks were mono-logic, not dialogic, thereby providing a basis for developing measures of learner performance unaffected by interactional variables.

The whole project lasted for 10 sessions, each session taking one hour and 45 minutes.

The language learners had about 3 hours and 30 minutes of English training each week. However, in each session, about one hour was devoted to the regular instruction and the coursebook, Touchstone Series Book 3 (McCarthy, McCarten, & Sandiford, 2005), and 30 to 45 minutes were allocated to performing the treatments by allocating different planning time conditions along with manipulating the difficulty of narrative tasks to prepare the participants for further data collection. To examine the effectiveness of the treatment, both control and experimental groups took the posttest of speaking. As a posttest, the language learners of each group were required to narrate a story in accordance with the planning type and task complexity level they were given during the treatment sessions. For cognitively complex tasks, the participants were supposed to put the frames of the comic strip in the correct order of occurrence to narrate the story. But the participants in the low complex task groups were given a series of ordered pictures to narrate. Their speaking performances were audiotaped to be coded and scored later by three different raters, with 20 being the maximum grade. The language learners' posttest scores were then compared with their pretest scores to examine the effectiveness of the treatments.

Pilot Study

It must be noted that 12 intermediate learners from the same language institute participated in the pilot study in which they performed all the tasks under the planning conditions of the study. Taking their performances into consideration, the researchers decided to give language learners 1 to 3 minutes to narrate each task in the pre-task planning and no-planning groups. The two control groups (no-planning) were given a short introduction to task performance. This would help them realize that they were not required to do planning for their speaking tasks. The two pre-planning experimental groups received the same introduction, followed by a 10-minute planning time before the speaking task. To boost the chances of pre-planning, the learners were asked to take notes about what they intended to talk about but were reminded that they could not use the notes before speaking. The participants in the two other experimental groups used online planning to tell the story. Online planning is the strategy used by speakers to take notice of the formulation of linguistic structures during speech planning for their language production. Thus, online planners were supposed to produce a narrative story for each picture without being given much time before their oral performance. The online planning groups were required to start the task after 30 seconds, but they had unlimited time to monitor their speech plan as they were narrating the story. To wit, being provided with ample time to perform their narration

task, they were under no time pressure to finish the narration. However, based on the result of the pilot study, narrating a short story by online planners did not take more than 5 minutes.

Data Analysis

IBM SPSS Statistics 24 was used for statistical analyses to answer the research question formulated earlier. Pearson Product Moment Correlation Coefficient was employed to ensure the inter-rater reliability of the pretest and posttest scores assigned by the three raters. Noteworthy to mention is that the scores reported here are the means of scores given by the three raters. Moreover, the normality of the data gathered by the researcher was checked using Kolmogorov-Smirnova and Shapiro-Wilk and the data were normally distributed for all groups' pretest and posttest scores.

Descriptive statistics were run for language learners' fluency, accuracy, and complexity pretest and posttest scores. To investigate the potentially significant differences among the groups prior to the treatments, One-way ANOVA was run on pretest scores. Mixed between-within groups ANOVAs were run on the language learners' fluency, accuracy, and complexity scores separately, with the combination of task complexity and planning type (no-planning low complexity, no-planning high complexity, pre-task low complexity, pre-task high complexity, online low complexity, online high complexity) and time including pretest and posttest as independent variables and language learners' scores as the dependent variable. To explore the specific differences among the groups, One-way ANOVAs and Tukey's pairwise post hoc comparisons were also conducted on learners' posttest scores.

The Results of the Research Question

The research question sought to identify if the combination of task complexity and planning affect Iranian EFL learners' oral production performance concerning a) fluency, b) accuracy, and c) complexity.

Normality

All the statistical techniques and the specific assumptions used in this study retained the normality of the data. Table 1 depicts the normality test results.

Table 1. *Testing the Normality of the Data*

	Groups	Kolmogorov-Smirnov ^b			Shapiro-Wilk			
		Statistic	df	Sig.	Statistic	df	Sig.	
Fluency	Pretest	low-pretask	.129	15	.200*	.966	15	.791
		low, no-planning	.218	15	.053	.911	15	.140
		high, no-planning	.111	15	.200*	.959	15	.670
		low-online	.122	15	.200*	.963	15	.749
		high-online	.198	15	.117	.950	15	.526
		high-pretask	.106	15	.200*	.971	15	.877
	Posttest scores	low-pretask	.147	15	.200*	.945	15	.452
		low, no-planning	.214	15	.063	.917	15	.174
		high, no-planning	.181	15	.200*	.946	15	.466
		low-online	.122	15	.200*	.950	15	.524
		high-online	.142	15	.200*	.962	15	.732
		high-pretask	.185	15	.176	.922	15	.207
		low-pretask	.124	15	.200*	.941	15	.393
		low, no-planning	.177	15	.200*	.951	15	.535
Accuracy	Pretest	high, no-planning	.172	15	.200*	.939	15	.375
		low-online	.143	15	.200*	.961	15	.702
		high-online	.214	15	.062	.933	15	.307
		high-pretask	.127	15	.200*	.965	15	.775
		low-pretask	.120	15	.200*	.979	15	.963
	Posttest scores	low, no-planning	.162	15	.200*	.903	15	.107
		high, no-planning	.210	15	.074	.878	15	.044
		low-online	.153	15	.200*	.916	15	.165
		high-online	.145	15	.200*	.912	15	.147
		high-pretask	.190	15	.149	.957	15	.638
Complexity	Pretest	low-pretask	.092	15	.200*	.968	15	.830
		low, no-planning	.089	15	.200*	.991	15	1.000
		high, no-planning	.178	15	.200*	.926	15	.234
		low-online	.202	15	.101	.935	15	.319
		high-online	.209	15	.076	.919	15	.183
		high-pretask	.159	15	.200*	.930	15	.275
	Posttest scores	low-pretask	.134	15	.200*	.960	15	.698
		low, no-planning	.118	15	.200*	.961	15	.710
		high, no-planning	.139	15	.200*	.946	15	.469
		low-online	.127	15	.200*	.942	15	.414
		high-online	.123	15	.200*	.958	15	.651
		high-pretask	.127	15	.200*	.949	15	.503

Since all the significance values turned out to be above 0.05, it can be concluded that the data met the normality assumption.

Fluency

The descriptive statistics for the groups' pretests and posttests on fluency are depicted in Table 2.

Table 2. Descriptive Statistics of Fluency Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
					Lower Bound	Upper Bound			
Pretest	low-pretask	15	14.8477	.73413	.18955	14.4412	15.2543	13.35	16.23
	low, no-planning	15	14.4068	.82311	.21252	13.9509	14.8626	12.94	15.50
	high, no-planning	15	14.2936	1.41191	.36455	13.5117	15.0755	12.01	16.54
	low-online	15	14.5363	.90281	.23310	14.0364	15.0363	13.20	16.24
	high-online	15	14.0433	.74418	.19215	13.6312	14.4554	12.75	15.33
	high-pretask	15	14.5408	.86903	.22438	14.0596	15.0221	12.61	15.93
	Total	90	14.4448	.94902	.10004	14.2460	14.6435	12.01	16.54
Posttest scores	low-pretask	15	17.4098	1.06721	.27555	16.8188	18.0008	16.02	19.49
	low, no-planning	15	15.6786	.73310	.18929	15.2726	16.0846	14.74	17.15
	high, no-planning	15	15.4215	1.69584	.43786	14.4823	16.3606	13.09	18.44
	low-online	15	16.3420	1.15572	.29841	15.7020	16.9821	14.74	18.96
	high-online	15	15.4721	1.17368	.30304	14.8221	16.1220	12.97	17.16
	high-pretask	15	16.0396	1.06088	.27392	15.4521	16.6271	13.38	17.76
	Total	90	16.0606	1.33864	.14111	15.7802	16.3410	12.97	19.49

a. Proficiency = fluency

To identify the potentially significant differences between the groups in terms of fluency before the treatment, the one-way ANOVA on the groups' pretest scores was performed. The results are displayed in Table 3.

The One-way ANOVA on fluency pretest scores did not show any statistically significant difference among the groups ($F(5, 84) = 1.23, p = .30$), suggesting that the groups were homogenous with regard to fluency before the treatment.

In the next step, to explore the impacts of treatment on the experimental groups and control learners' fluency over time, the language learners' pretest and posttest scores were analyzed using Mixed between-within groups ANOVAs.

Table 3. One-way ANOVA regarding the Difference between Groups Concerning Fluency Pretest Scores

	Sum of Squares	df	Mean Square	F	Sig.
--	----------------	----	-------------	---	------

Pretest	Between Groups	5.481	5	1.096	1.233	.301
	Within Groups	74.676	84	.889		
	Total	80.158	89			

a. Proficiency = fluency

The homogeneity of variances of the groups and covariance matrices were checked by Levene's test and Box's test, respectively. Tables 4 and 5 demonstrate the pertaining results.

Table 4. *Levene's Test of Equality of Error Variances on Fluency Scores*

	F	df1	df2	Sig.
Pretest	2.193	5	84	.063
Posttest scores	2.329	5	84	.051

Based on Table 4, there existed no significant differences between the groups' variances on fluency pretest ($F(5, 84) = 2.19, p > .05$) and posttest ($F(5, 84) = 2.32, p > .05$).

Table 5. *Box's Test of Equality of Covariance Matrices on Fluency Scores*

Box's M	25.023
F	1.567
df1	15
df2	38594.288
Sig.	.074

According to Table 5, the non-significant results of the test ($M = 25.02, p > .001$) show that the homogeneity of covariance matrices was met. Table 6 depicts the results of the Multivariate test.

Table 6 suggests statistically significant main effects for time, $F(1, 84) = 260.40, p < .001$. The effect size based on Cohen's (1988) criterion is large (partial eta squared = .75). This implies that the language learners benefited from the combination of task complexity and planning type. Along the same lines, a statistically significant interaction effect was detected between time and combinations of task complexity and planning type, $F(5, 84) = 4.44, p < .001$ suggesting that the language learners differentially benefited from the combination of task complexity and planning type. As to the effect size, the partial eta squared turned out to be .20 which is deemed high (Cohen, 1988). Table 7 depicts the results of the Tests of Between-Subjects Effects.

Table 6. *Multivariate Tests on Fluency Pre and Posttest Scores*

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Time	Pillai's Trace	.756	260.407 ^c	1.000	84.000	.000	.756
	Wilks' Lambda	.244	260.407 ^c	1.000	84.000	.000	.756
	Hotelling's Trace	3.100	260.407 ^c	1.000	84.000	.000	.756
	Roy's Largest Root	3.100	260.407 ^c	1.000	84.000	.000	.756
Time * Groups	Pillai's Trace	.209	4.444 ^c	5.000	84.000	.001	.209
	Wilks' Lambda	.791	4.444 ^c	5.000	84.000	.001	.209
	Hotelling's Trace	.265	4.444 ^c	5.000	84.000	.001	.209
	Roy's Largest Root	.265	4.444 ^c	5.000	84.000	.001	.209

Table 7. *Tests of Between-Subjects Effects on Fluency Scores*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	41875.983	1	41875.983	22804.541	.000	.996
Groups	37.468	5	7.494	4.081	.002	.195
Error	154.249	84	1.836			

a. Proficiency = fluency

As seen in Table 7, the main effect of comparing the six types of interventions was significant, $F(5, 84) = 4.08$, $p < .05$, partial eta squared = .19 showing a large effect size. That is, a significant difference was observed in the effectiveness of the six types of combinations of task complexity and planning type.

Having found main effects for the combination of task complexity and planning type, it is of importance to explore which combinations of task complexity and planning type significantly affected language learners' oral productions in terms of fluency. Therefore, a one-way ANOVA was conducted on learners' fluency posttest scores (Table 8).

Table 8. *One-way ANOVA regarding the Difference between Groups Concerning Fluency Posttest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Posttest scores	Between Groups	42.012	5	8.402	6.008	.000
	Within Groups	117.473	84	1.398		

Total	159.485	89
--------------	---------	----

a. Proficiency = fluency

Table 8 revealed a significant difference among the groups concerning the posttest scores ($F(5, 84) = 6.00, p < .001$). The results of the Tukey's pairwise post hoc are depicted in Table 9.

Table 9. Tests of Between-Subjects Effects on Fluency Scores

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
low-pretask	low, no-planning	1.0861*	.34989	.030	.0656	2.1065
	high, no-planning	1.2712*	.34989	.006	.2508	2.2917
	low-online	.6896	.34989	.368	-.3309	1.7100
	high-online	1.3711*	.34989	.002	.3506	2.3915
	high-pretask	.8385	.34989	.169	-.1819	1.8590
low, no-planning	low-pretask	-1.0861*	.34989	.030	-2.1065	-.0656
	high, no-planning	.1851	.34989	.995	-.8353	1.2056
	low-online	-.3965	.34989	.866	-1.4170	.6240
	high-online	.2850	.34989	.964	-.7355	1.3054
	high-pretask	-.2475	.34989	.981	-1.2680	.7729
high, no-planning	low-pretask	-1.2712*	.34989	.006	-2.2917	-.2508
	low, no-planning	-.1851	.34989	.995	-1.2056	.8353
	low-online	-.5816	.34989	.560	-1.6021	.4388
	high-online	.0998	.34989	1.000	-.9206	1.1203
	high-pretask	-.4327	.34989	.818	-1.4531	.5878
low-online	low-pretask	-.6896	.34989	.368	-1.7100	.3309
	low, no-planning	.3965	.34989	.866	-.6240	1.4170
	high, no-planning	.5816	.34989	.560	-.4388	1.6021
	high-online	.6815	.34989	.381	-.3390	1.7019
	high-pretask	.1490	.34989	.998	-.8715	1.1694
high-online	low-pretask	-1.3711*	.34989	.002	-2.3915	-.3506
	low, no-planning	-.2850	.34989	.964	-1.3054	.7355
	high, no-planning	-.0998	.34989	1.000	-1.1203	.9206
	low-online	-.6815	.34989	.381	-1.7019	.3390
	high-pretask	-.5325	.34989	.651	-1.5530	.4879
high-pretask	low-pretask	-.8385	.34989	.169	-1.8590	.1819
	low, no-planning	.2475	.34989	.981	-.7729	1.2680
	high, no-planning	.4327	.34989	.818	-.5878	1.4531
	low-online	-.1490	.34989	.998	-1.1694	.8715
	high-online	.5325	.34989	.651	-.4879	1.5530

As shown in Table 9, the pre-task planning low complexity group ($M = 17.40, SD = 1.06$) significantly outperformed the no-planning low complexity ($M = 15.67, SD = .73$), no-planning high complexity ($M = 15.42, SD = 1.69$), and online high complexity ($M = 15.47, SD = 1.17$) groups.

Accuracy

Table 10 displays the descriptive statistics for language learners' accuracy scores in the pretest and posttest.

Table 10. *Descriptive Statistics of Accuracy Scores*

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
						Pretest	low-pretask		
	low, no-planning	15	14.6068	1.55534	.40159	13.7455	15.4681	12.00	17.48
	high, no-planning	15	14.6035	.96429	.24898	14.0695	15.1375	13.19	16.88
	low-online	15	14.6496	1.10387	.28502	14.0383	15.2609	12.66	16.30
	high-online	15	14.0338	1.02079	.26357	13.4685	14.5991	12.42	16.25
	high-pretask	15	14.6311	.98878	.25530	14.0835	15.1786	13.19	16.40
	Total	90	14.5006	1.11839	.11789	14.2663	14.7348	12.00	17.48
Posttest scores	low-pretask	15	16.0492	1.29850	.33527	15.3301	16.7683	13.45	18.37
	low, no-planning	15	16.0283	1.09999	.28402	15.4192	16.6375	14.54	17.57
	high, no-planning	15	15.9635	1.01352	.26169	15.4023	16.5248	14.93	18.33
	low-online	15	17.3409	.88097	.22747	16.8530	17.8288	15.91	18.94
	high-online	15	17.9307	.94755	.24466	17.4060	18.4554	16.17	19.03
	high-pretask	15	17.2263	.80494	.20784	16.7806	17.6721	16.12	18.39
	Total	90	16.7565	1.26063	.13288	16.4925	17.0205	13.45	19.03

a. Proficiency = accuracy

To see whether the differences in accuracy pretest mean scores were significant or not, the data were submitted to a one-way ANOVA test (Table 11).

Table 11. *One-way ANOVA regarding the Difference between Groups Concerning Accuracy Pretest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	4.193	5	.839	.658	.657
	Within Groups	107.127	84	1.275		
	Total	111.320	89			

a. Proficiency = accuracy

As evident in Table 11, no significant difference was observed between the groups concerning the accuracy pretest scores ($F(5, 84) = .65, p = .65$). It can be inferred that the groups' homogeneity regarding accuracy was met before that treatment.

In order to identify if the treatment had any effect on the experimental and control learners' accuracy over time, mixed between-within groups ANOVA was employed. First, the homogeneity of variances was checked (Table 12).

Table 12. *Levene's Test of Equality of Error Variances on Accuracy Scores*

	F	df1	df2	Sig.
Pretest	1.084	5	84	.375
Posttest scores	.907	5	84	.481

Table 12 displays the equality of variances on the accuracy pretest ($F(5, 84) = 1.08, p > .05$) and posttest ($F(5, 84) = .90, p > .05$). The results of the homogeneity test of covariance are displayed in Table 13.

Table 13. *Box's Test of Equality of Covariance Matrices on Accuracy Scores*

Box's M	38.360
F	2,402
df1	15
df2	38594.288
Sig.	.12

According to Table 13, the homogeneity of covariance matrices was maintained ($M = 38.36, p > .001$). Table 14 reports the results of the Multivariate test.

Table 14. *Multivariate Tests on Accuracy Pre and Posttest Scores*

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Time	Pillai's Trace	.802	340.429 ^c	1.000	84.000	.000	.802
	Wilks' Lambda	.198	340.429 ^c	1.000	84.000	.000	.802
	Hotelling's Trace	4.053	340.429 ^c	1.000	84.000	.000	.802
	Roy's Largest Root	4.053	340.429 ^c	1.000	84.000	.000	.802

Time * Groups	Pillai's Trace	.397	11.074 ^c	5.000	84.000	.000	.397
	Wilks' Lambda	.603	11.074 ^c	5.000	84.000	.000	.397
	Hotelling's Trace	.659	11.074 ^c	5.000	84.000	.000	.397
	Roy's Largest Root	.659	11.074 ^c	5.000	84.000	.000	.397

Table 14 displays a substantial main effect for time $F(1, 84) = 340.42, p < .001$, partial eta squared = .80 showing a large effect size with all the groups reflecting an improvement in accuracy scores across the two time periods (pre to posttest). Likewise, a significant interaction was detected between time and combinations of task complexity and planning type, $F(5, 84) = 11.07, p < .001$, partial eta squared = .39 displaying a large effect size. In other words, language learners benefited differentially from the combinations of task complexity and planning type. Table 15 shows the results of the Tests of Between-Subjects Effects.

Table 15. Tests of Between-Subjects Effects on Accuracy Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	43965.201	1	43965.201	26751.246	.000	.997
Groups	20.9511	5	4.190	2.550	.034	.132
Error	138.053	84	1.643			

a. Proficiency = accuracy

According to Table 15, the main effect comparing the six types of interventions was significant, $F(5, 84) = 2.55, p < .05$, partial eta squared = .13 suggesting that there was a significant difference in the effectiveness of the six types of combinations of task complexity and planning type. The partial eta squared also indicates that the effect size was large. Consequently, it can be inferred that the language learners' accuracy enhanced over time and they benefited from the combination of task complexity and planning type from pretest to posttest.

To confirm where the differences occurred between groups, One-way ANOVA and Tukey's pairwise post hoc comparison were performed on language learners' accuracy posttest scores. Tables 16 and 17 demonstrate the one-way ANOVA results and Tukey's pairwise post hoc comparisons, respectively.

Table 16. One-way ANOVA regarding the Difference between Groups Concerning Accuracy

Posttest Scores

		Sum of Squares	df	Mean Square	F	Sig.
Posttest scores	Between Groups	54.006	5	10.801	10.377	.000
	Within Groups	87.433	84	1.041		
	Total	141.439	89			

a. Proficiency = accuracy

The results displayed in Table 16 indicated significant differences among the groups concerning the posttest scores on accuracy ($F(5, 84) = 10.37, p < .001$).

Based on the post hoc analysis in Table 17 and the descriptive statistics depicted in Table 10, the online low complexity group ($M = 17.34, SD = .88$) and the online high complexity group ($M = 17.93, SD = .94$) significantly outperformed the pre-task planning low complexity ($M = 16.04, SD = 1.29$), no-planning low complexity ($M = 16.02, SD = 1.09$), and no-planning high complexity ($M = 15.96, SD = 1.01$), groups. Therefore, it can be concluded that the language learners who employed online planning were more accurate than those using pre-task planning low complexity and no planning groups.

Additionally, the pre-task planning high complexity group significantly outperformed the pre-task planning low complexity ($M = 16.04, SD = 1.29$), no-planning low complexity ($M = 16.02, SD = 1.09$), and no-planning high complexity ($M = 15.96, SD = 1.01$) groups.

Table 17. Post-Hoc Tukey HSD Test of the Groups' Accuracy Posttest Scores

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Tukey HSD	low, no-planning	.02086	.37253	1.000	-1.0657	1.1074
	high, no-planning	.08564	.37253	1.000	-1.0009	1.1722
	low-online	-1.29173*	.37253	.010	-2.3782	-.2052
	high-online	-1.88152*	.37253	.000	-2.9680	-.7950
	high-pretask	-1.17718*	.37253	.026	-2.2637	-.0907

low, no-planning	low-pretask	-.02086	.37253	1.000	-1.1074	1.0657
	high, no-planning	.06478	.37253	1.000	-1.0217	1.1513
	low-online	-1.31259*	.37253	.009	-2.3991	-.2261
	high-online	-1.90238*	.37253	.000	-2.9889	-.8159
	high-pretask	-1.19804*	.37253	.022	-2.2846	-.1115
high, no-planning	low-pretask	-.08564	.37253	1.000	-1.1722	1.0009
	low, no-planning	-.06478	.37253	1.000	-1.1513	1.0217
	low-online	-1.37737*	.37253	.005	-2.4639	-.2909
	high-online	-1.96716*	.37253	.000	-3.0537	-.8806
	high-pretask	-1.26282*	.37253	.013	-2.3493	-.1763
low-online	low-pretask	1.29173*	.37253	.010	.2052	2.3782
	low, no-planning	1.31259*	.37253	.009	.2261	2.3991
	high, no-planning	1.37737*	.37253	.005	.2909	2.4639
	high-online	-.58979	.37253	.612	-1.6763	.4967
	high-pretask	.11455	.37253	1.000	-.9720	1.2011
high-online	low-pretask	1.88152*	.37253	.000	.7950	2.9680
	low, no-planning	1.90238*	.37253	.000	.8159	2.9889
	high, no-planning	1.96716*	.37253	.000	.8806	3.0537
	low-online	.58979	.37253	.612	-.4967	1.6763
	high-pretask	.70434	.37253	.415	-.3822	1.7909
high-pretask	low-pretask	1.17718*	.37253	.026	.0907	2.2637
	low, no-planning	1.19804*	.37253	.022	.1115	2.2846
	high, no-planning	1.26282*	.37253	.013	.1763	2.3493
	low-online	-.11455	.37253	1.000	-1.2011	.9720
	high-online	-.70434	.37253	.415	-1.7909	.3822

*. The mean difference is significant at the 0.05 level.

Complexity

The descriptive statistics for pre and posttest scores on complexity are provided in Table 18.

Table 18. Descriptive Statistics of Complexity Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					Pretest			
low-pretask	15	14.5934	1.20942	.31227	13.9236	15.2631	12.14	16.42
low, no-planning	15	14.2172	1.33450	.34457	13.4781	14.9562	11.48	16.83

	high, no-planning	15	14.1499	1.16621	.30111	13.5041	14.7957	12.45	15.95
	low-online	15	14.1930	.79312	.20478	13.7537	14.6322	13.02	15.84
	high-online	15	13.8284	.78647	.20307	13.3928	14.2639	12.30	15.25
	high-pretask	15	14.1851	1.33257	.34407	13.4471	14.9230	10.83	16.31
	Total	90	14.1945	1.11785	.11783	13.9603	14.4286	10.83	16.83
	low-pretask	15	16.2566	.94753	.24465	15.7318	16.7813	14.31	17.66
	low, no-planning	15	15.5540	1.26755	.32728	14.8521	16.2560	13.43	17.65
	high, no-planning	15	16.0094	1.02530	.26473	15.4416	16.5772	14.24	17.51
Posttest scores	low-online	15	15.7825	.98276	.25375	15.2383	16.3268	14.45	17.49
	high-online	15	15.9145	1.16429	.30062	15.2697	16.5592	14.14	17.95
	high-pretask	15	17.5158	1.37915	.35610	16.7521	18.2795	15.46	19.94
	Total	90	16.1721	1.27860	.13478	15.9043	16.4399	13.43	19.94

a. Proficiency = complexity

In the next step, a One-way ANOVA was performed to explore potential significant differences between the groups with respect to the pretest complexity scores. The results are shown in Table 19.

Table 19. One-way ANOVA regarding the Difference between Groups Regarding Complexity Pretest Scores

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	4.436	5	.887	.698	.626
	Within Groups	106.777	84	1.271		
	Total	111.214	89			

a. Proficiency = complexity

The results of One-way ANOVA on complexity pretest scores did not reveal any significant difference among the groups ($F(5, 84) = .69, p = .62$). It shows that the groups were homogenous before the treatment.

To investigate the effects of the treatment on the experimental and control groups learners' complexity over time, the Mixed between-within groups ANOVA was run on the learners' complexity pretest and posttest scores. First, the results of the Levene's test and Box's test checking the groups' homogeneity of variances and homogeneity of covariance matrices,

respectively, are provided in Tables 20 and 21, respectively.

Table 20. *Levene's Test of Equality of Error Variances on Complexity Scores*

	F	df1	df2	Sig.
Pretest	1.191	5	84	.320
Posttest	.870	5	84	.505

According to Table 20, no significant difference was observed between the groups' variances on complexity pretest ($F(5, 84) = .32, p > .05$) and posttest ($F(5, 84) = .505, p > .05$).

Table 21. *Box's Test of Equality of Covariance Matrices on Complexity Scores*

Box's M	30.724
F	1.924
df1	15
df2	38594.288
Sig.	.000

As depicted in Table 21, the homogeneity of covariance matrices was met ($M = 30.724, p > .001$). The Multivariate test results are presented in Table 22.

The Mixed between-within groups ANOVA revealed significant main effects for time, $F(1, 84) = 245.95, p < .001$, partial eta squared = .74 displaying a large effect size, and the interaction between time and the combination of task complexity and planning type $F(2, 84) = 5.27, p < .001$, partial eta squared = .23 representing a large effect size. Table 23 displays the results of Tests of Between-Subjects Effects.

Table 22. *Multivariate Tests on Complexity Pretest and Posttest Scores*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.745	245.958 ^c	1.000	84.000	.000	.745
	Wilks' Lambda	.255	245.958 ^c	1.000	84.000	.000	.745
	Hotelling's Trace	2.928	245.958 ^c	1.000	84.000	.000	.745
	Roy's Largest Root	2.928	245.958 ^c	1.000	84.000	.000	.745
	Pillai's Trace	.239	5.275 ^c	5.000	84.000	.000	.239

Time * Groups	Wilks' Lambda	.761	5.275 ^c	5.000	84.000	.000	.239
	Hotelling's Trace	.314	5.275 ^c	5.000	84.000	.000	.239
	Roy's Largest Root	.314	5.275 ^c	5.000	84.000	.000	.239

Table 23. Tests of Between-Subjects Effects on Fluency Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	41495.874	1	41495.874	22404.586	.000	.996
Groups	22.151	5	4.430	2.392	.044	.125
Error	155.578	84	1.852			

a. Proficiency = complexity

Regarding the complexity, the results ($F(5, 84) = 2.39, p < .05$, partial eta squared = .125 (showing a large effect size) were indicative of significant differences in the effectiveness of the six types of combinations of task complexity and planning type. To specify the differences between the six groups, One-way ANOVA (Table 24) and Tukey's Post hoc comparisons (Table 25) were conducted.

Table 24. One-way ANOVA regarding the Difference between Groups Regarding Complexity Posttest Scores

		Sum of Squares	df	Mean Square	F	Sig.
Posttest	Between Groups	36.590	5	7.318	5.644	.000
	Within Groups	108.909	84	1.297		
	Total	145.498	89			

a. Proficiency = complexity

Table 24 exhibits that there were significant differences among the groups in the complexity posttest scores, $F(5, 84) = 5.64, p < .001$.

Table 25. Hoc Comparisons of the Groups' Complexity Posttest Scores

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Tukey HSD	low, no-planning	.70256	.41578	.542	-.5101	1.9152
	high, no-planning	.24714	.41578	.991	-.9655	1.4598
	low-online	.47404	.41578	.863	-.7386	1.6867
	high-online	.34210	.41578	.963	-.8705	1.5547
	high-pretask	-1.25923*	.41578	.037	-2.4719	-.0466

low, no-planning	low-pretask	-.70256	.41578	.542	-1.9152	.5101
	high, no-planning	-.45542	.41578	.882	-1.6681	.7572
	low-online	-.22852	.41578	.994	-1.4412	.9841
	high-online	-.36046	.41578	.953	-1.5731	.8522
	high-pretask	-1.96179*	.41578	.000	-3.1744	-.7492
high, no-planning	low-pretask	-.24714	.41578	.991	-1.4598	.9655
	low, no-planning	.45542	.41578	.882	-.7572	1.6681
	low-online	.22690	.41578	.994	-.9857	1.4395
	high-online	.09496	.41578	1.000	-1.1177	1.3076
	high-pretask	-1.50637*	.41578	.006	-2.7190	-.2937
low-online	low-pretask	-.47404	.41578	.863	-1.6867	.7386
	low, no-planning	.22852	.41578	.994	-.9841	1.4412
	high, no-planning	-.22690	.41578	.994	-1.4395	.9857
	high-online	-.13194	.41578	1.000	-1.3446	1.0807
	high-pretask	-1.73327*	.41578	.001	-2.9459	-.5206
high-online	low-pretask	-.34210	.41578	.963	-1.5547	.8705
	low, no-planning	.36046	.41578	.953	-.8522	1.5731
	high, no-planning	-.09496	.41578	1.000	-1.3076	1.1177
	low-online	.13194	.41578	1.000	-1.0807	1.3446
	high-pretask	-1.60133*	.41578	.003	-2.8140	-.3887
high-pretask	low-pretask	1.25923*	.41578	.037	.0466	2.4719
	low, no-planning	1.96179*	.41578	.000	.7492	3.1744
	high, no-planning	1.50637*	.41578	.006	.2937	2.7190
	low-online	1.73327*	.41578	.001	.5206	2.9459
	high-online	1.60133*	.41578	.003	.3887	2.8140

As shown in Table 25, the post hoc comparisons test revealed that the pre-task high complexity group ($M= 17.51$, $SD= 1.37$) significantly outperformed all other groups.

Discussion of the Results for the Research Question

The current study intended to examine the joint impacts of planning time conditions and task complexity on language learners' oral productions with regard to CAF.

Complexity

Concerning complexity, the language learners in the pre-task planning high complexity group outperformed all other groups (i.e. pre-task planning low complexity group, no-planning low complexity group, no-planning high complexity group, online planning low complexity group, online planning high complexity group).

The pre-task planning low complexity group's mean score was higher than those of the online groups (high and low complexity) and no planning groups (high and low complexity) in terms of

complexity though not reaching statistical significance. It can be implied that pre-task planning impacted language learners' complexity. That is, the language learners provided with more resources, produced more complex constructions. This result is in harmony with previous studies (e.g., Ahangari & Abdi, 2011; Crooks, 1989; Foster & Skehan, 1996; Gilabert, 2007; Mehnert, 1998; Ortega, 1999), all of which have demonstrated that affording language learners with the opportunity to plan can increase the complexity level of their production. In a similar vein, Yuan and Ellis (2003) found that pre-task planning promotes grammatical complexity. One might reason that the language learners already prioritized complexity or they might already focus on complexifying their productions when provided with planning time.

Considering the mean score difference between the pre-task high complexity group ($M= 17.51$) and the pre-task planning low complexity group ($M= 16.25$), it is concluded that the language learners undertaking more complex tasks did better than those performing low complex tasks in terms of complexity. This finding is in parallel with the Involvement Load Hypothesis developed by Laufer and Hulstijn (2001). Moreover, this result fits neatly with Robinson's (2001c) findings which suggest that complex tasks trigger more complex language than simple tasks. Inconsistent with our findings, Rahimpour (2007) revealed that complex tasks gave rise to the production of less complex language.

A possible explanation for this might be that the attentional resources of the language learners performing complex tasks went beyond the reasonable demand of competently undertaking the tasks in terms of complexity. Likewise, when the language learners were given planning time, they focused on the content of tasks and the preparation for the task making them produce more complex language.

Accuracy

Regarding accuracy, the online low complexity group outperformed the pre-task planning low complexity, no-planning low complexity, and no-planning high complexity groups. The results also revealed that the online high complexity group significantly differed from the pre-task planning low complexity, no-planning low complexity, and no-planning high complexity groups.

Moreover, it can be concluded from the results that the language learners who employed online planning were more accurate than the other groups. Little research has explored the effect of within-task planning on CAF (Ellis 2009). The studies that have investigated this effect have found an increase in both accuracy and complexity (Ahmadian & Tavakoli 2011; Yuan & Ellis 2003). However, under online planning, language learners pay more attention to the formulation phase and are involved in pre- and post-monitoring of their productions (Yuan & Ellis, 2003). This seems to be in line with Dekeyser's (2003) argument that formulating language under online planning forces language users to draw on their implicit knowledge.

The findings of the current study regarding accuracy are in agreement with the one obtained by Nasiri and Atai (2017) who found that online planners performing simple and complex tasks significantly improved their accuracy. This result lends support to Yuan and Ellis's (2003) study in which online planning was found to positively impact accuracy. Along the same lines, this study replicates the findings of Khoram (2019) who reported that online planning assisted language learners to substantially improve their accuracy both in simple and complex tasks.

Additionally, the pre-task planning complex group significantly outperformed the pre-task planning low complexity, no-planning low complexity, and no-planning high complexity groups. Considering the significant difference between the pre-task planning complex group and the pre-task planning low complexity, it is inferred that the language learners performing complex tasks produced more accurate speech acts than those doing low complex tasks. This confirms Kuiken and Vedder's (2007) argument stating that task complexity influences linguistic performance. That is the increase in cognitive task complexity results in more accurate language output.

The higher mean score of the online high complexity group compared with other groups in terms of accuracy can be justified with regard to Skehan's (1998) dual-mode system proposal suggesting that under pressured online planning, language learners rely on their exemplar-based system entailing a large number of prefabricated chunks which imposes lower cognitive demands on the language learner leading to more accurate sentences (Ahmadian et al, 2012).

Fluency

Concerning fluency, the pre-task planning low complexity group significantly outperformed the no-planning low complexity, no-planning high complexity, and online high complexity groups.

Given the higher mean scores in the pre-task planning low complexity group and the pre-task planning high complexity group compared with the other groups, it is inferred that pre-task planning impacted language learners' speech fluency.

This finding is in accord with that of Yuan and Ellis's study (2003) in which language learners in the pre-task planning groups generated more fluent language than did the online planning groups. Although no significant differences were observed between the pre-task

planning high complexity group and the no planning and online planning groups, the mean score was higher than these groups. One tentative explanation for the positive effect of pre-task planning on fluency is that the language learners did not rely on their grammatical rules which typically loads working memory. Consequently, their attentional resources process meaning in an effective manner, thereby increasing the rate of speech fluency. Likewise, based on Nasiri and Atai (2017), under the pre-task planning condition, the language learners did not plan while performing the task. Thus, they undertook it more fluently. This line of explanation is in accord with the common belief in the language teaching literature that online planning decreases language learners' fluency.

Moreover, the mean score of the pre-task planning low complexity group was higher than the pre-task planning high complexity group in terms of their fluency. This suggests that the pre-task planners carrying out low complex tasks produced more fluent language. One justification might be that those language learners doing low complex tasks were less cognitively involved. This finding is in accord with the results of Foster and Skehan (1996), Wendel (1997), Mehnert (1998), and Ortega (1999) who found that pre-task planning significantly affected L2 fluency. However, this disagrees with the findings of Gilabert (2007) and Yuan and Ellis (2003) which suggested that pre-task planning did not enhance fluency. Concerning the impact of task complexity on fluency, our finding runs counter with that of Salimi and Dadashpour (2012) who revealed that task complexity led to an increase in fluency. However, consistent with our result in this regard, Brown, Anderson, Shilcock, and Yule (1984) found that fluency decreased as a result of the complex task.

Conclusion

The current study investigated the combined impacts of task complexity and planning on language learners' oral productions with regard to CAF.

The findings exhibited that the language learners in the pre-task planning low complex task group were more fluent than the other groups. Likewise, pre-task planning impacted complexity and fluency while online planning affected accuracy more.

Regarding the planning conditions, pre-task planning produced positive impacts on complexity and fluency. Likewise, online planning influenced accuracy more than did no-planning and pre-task planning conditions.

As for task complexity, our findings confirm Robinson's Cognition Hypothesis in which

the development of the language learners' speaking skill is resultant of employing more challenging tasks. To wit, increasing the difficulty of the task to a reasonable level can effectively enhance the learners' speaking ability. At this point, EFL teachers should develop language learners' ability to accomplish real-world tasks. By involving language learners in increasingly complex cognitive and interactive activities, teachers help them develop their language learning.

The results obtained from this study with respect to complexity groups under the online planning condition confirmed Skehan's (1998) Limited Capacity Hypothesis meaning that increasing task complexity did not lead to higher accuracy and complexity simultaneously which is suggestive of a trade-off effect between accuracy and complexity. By contrast, the results concerning pre-task planning condition coupled with high complex tasks resulted in better gains in complexity and accuracy which lends support to Robinson's Cognition Hypothesis.

This study yields insights into the design and implementation of tasks in language teaching classroom settings. Drawing on the competing goals of CAF, language learners attempt to strike up a balance between these measures of speaking. Thus, the findings of the current study can redound to EFL teachers and materials designers to create tasks that place emphasis on each of these measures. Language teachers are required to embed the competing demands of CAF. At this point, EFL teachers need to teach language learners to be heedful of various elements of language including grammar for more accurate linguistic output and fewer false starts and reformulations, either lexical or morphological, and observing the appropriate rate of speech for improving disfluency of oral performance. Moreover, EFL teachers should adopt a wide variety of tasks that rely upon various skills to improve complexity, accuracy, and complexity. In other words, EFL teachers need to keep a good balance of tasks to ensure that CAF measures are not overlooked.

Given the limited time available for planning conditions in real-life situations, EFL teachers need to attain situational authenticity wherein language learners should be involved in performing real-life tasks. However, as this position is not always possible or practical in classroom settings, EFL teachers need to ensure interactional authenticity in which teachers encourage language learners to take on communication strategies (i.e. the ones practiced in real-life situations).

To improve language learners' CAF measures in oral production, EFL teachers can create a well-balanced task development wherein language learners' competence to use the target

language is aligned with respect to CAF.

By considering the findings of the current study, EFL teachers can manipulate planning time, encouraging pre-task planning and online planning in a way by which language learners can produce the target language in an actual testing situation.

In light of the findings of this study, EFL teachers can provide language learners with instruction on how to plan rather than simply allocate them sufficient time for planning. This would help language learners take advantage of planning time and make them prepared for speaking.

The present study has some pedagogical implications for task designers and language assessment specialists. The findings of the study can contribute to the establishment of a sound and a fine-grained assessment rubric for grading and task sequence. Moreover, the results suggest that language teachers should attend to the cognitive abilities of language learners and cognitive load tasks. Further, the cognitive complexity of tasks should be taken into account by language testers when designing tasks.

The current study suffered from some limitations that should be addressed. First, this research study was performed in an EFL setting among Iranian language learners. Consequently, the findings will be generalizable only in an EFL context. Second, the time allocated to treatment was 10 sessions. More sessions of treatment, if allocated, more implications would emerge.

The third limitation concerns the small sample size of the study ($n = 90$). Thus, generalizations should be made with caution.

Future studies should employ a mixed-methods approach to study task complexity or planning (i.e. performing post-task interviews) and think-aloud protocols to delve into the cognitive processes involved.

References

- Abdollahzadeh, S., & Fard Kashani, A. (2012). The Effect of Task Complexity on EFL Learners' Narrative Writing Task Performance. *Journal of English Language Teaching and Learning*, 3(8), 1-28.
- Ahangari, S., & Abdi, M. (2011). The Effect Of Pre-Task Planning on the Accuracy and Complexity of Iranian EFL Learners' Oral Performance. *Procedia-Social and Behavioral Sciences*, 29, 1950-1959.
- 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.

- Ahmadian, M. J., Tavakoli, M., & Vahid Dastjerdi, H. (2012). The Combined Effects of Online Planning and Task Structure on Complexity, Accuracy, and Fluency of L2 Speech. *The Language Learning Journal*, 43(1), 41–56.
- Awwad, A. A. (2017). *The Effects of Task Complexity Manipulated by Intentional Reasoning Demands on Second Language Learners' Speech Performance: Interaction with Language Proficiency and Working Memory*. Doctoral dissertation, University of Reading, UK.
- Baleghizadeh, S., & Shahri, M. N. N. (2017). The Effect of Online Planning, Strategic Planning, and Rehearsal Across Two Proficiency Levels. *The Language Learning Journal*, 45(2), 171-184.
- Brown, G. Anderson, A. Shilcock, R. & Yule, G. (1984). *Teaching Talk: Strategies for Production and Assessment*. Cambridge: Cambridge University Press.
- Bulté, B. & Housen, A. (2012). Defining and Operationalizing L2 Complexity. In Housen A, Kuiken F, and Vedder I (eds) *Dimensions of L2 Performance and Proficiency: Investigating Complexity, Accuracy, and Fluency in SLA*. Amsterdam/Philadelphia: Benjamins, 21–46.
- Cohen, J. W. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crooks, G. (1989). Planning and Interlanguage Variation. *Studies in Second Language Acquisition*, 11, 367-383.
- DeKeyser, R. (2003). Implicit and Explicit Learning. In *Handbook of Second Language Acquisition*, ed. C. Doughty and M. Long, 313–49. Malden, MA: Blackwell.
- Elder, C. A., & Iwashita, N. (2005). Planning for Test Performance: Does it Make a Difference? In Rod Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 219–238). Amsterdam: John Benjamins.
- Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.
- Ellis, R. (2005). Planning and Task-based Research: Theory and Research. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language*. John Benjamins.
- Ellis, R. (2008). *The Study of Second Language Acquisition (2nd edition)*. Oxford: Oxford University Press.
- Ellis, R. (2009). The Differential Effects of Three Types of Task Planning on the Fluency, Complexity, and Accuracy in L2 Oral Production. *Applied linguistics*, 30(4), 474-509.

- 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(1), 59-84.
- Farrokhi, F., & Sattarpour, S.(2017). Different Task Complexity Factors and Cognitive Individual Differences: The Effects on EFL Writers' Performance. *Applied Research on English Language*, 6(3), 387-410.
- 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.
- Gilabert, R. (2007). The Simultaneous Manipulation of Task Complexity along Planning Time and [+/-Here-and-Now]: Effects on L2 Oral Production. *Investigating Tasks in Formal Language Learning*, 20, 44-68.
- Heidari-Shahreza, M. A., Dabaghi, A., & Kassaian, Z. (2011). The Effects of Manipulating Task Complexity on the Occurrence of Language-related Episodes during Learner-Learner Interaction. *Porta Linguarum*, 17, 173-188.
- Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied linguistics*, 30(4), 461-473.
- Ishikawa, T. (2006). The Effect of Manipulating Task Complexity along the [+/- Here-and-Now] Dimension on L2 Written Narrative Discourse. In M.P. García Mayo (ed.), *Investigating Tasks in Formal Language Learning*. Clevedon: Multilingual Matters, 136-156.
- Izadpanah, S., & Shajeri, E. (2016). The Impact of Task Complexity along Single Task Dimension on EFL Iranian Learners' Written Production: Lexical Complexity. *Journal of Teaching Language Skills*, 35(1), 57-84.
- Kim, S. Y. (2009). Questioning the Stability of Foreign Language Classroom Anxiety and Motivation across Different Classroom Contexts. *Foreign Language Annals*, 42(1), 138-157.
- Khoram, A. (2019). The Impact of Task Type and Pre-task Planning Condition on the Accuracy of Intermediate EFL Learners' Oral Performance. *Cogent Education*, 6(1), 1675466.
- Kuiken, F., & Vedder, I. (2007). Cognitive Task Complexity and Linguistic Performance in French L2 Writing. In M. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 117–135). Clevedon: Multilingual matters Ltd.
- Kuiken, F., & Vedder, I. (2011). Task Complexity and Linguistic Performance in L2 Writing and

- Speaking. In Robinson, P. (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 91-104. 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.
- Lee, J. (2018). Task Complexity, Cognitive Load, and L1 Speech. *Applied Linguistics*, 40(3), 506-539.
- Liu, P., & Li, Z. (2012). Task Complexity: A Review and Conceptualization Framework. *International Journal of Industrial Ergonomics*, 42(6), 553-568.
- McCarthy, M., McCarten, J., & Sandiford, H. (2005). *Touchstone Series*. Cambridge: Cambridge University Press.
- Mehnert, U. (1998). The Effects of Different Lengths of Time for Planning on Second Language Performance. *Studies in Second Language Acquisition*, 20, 83–108.
- Michel, M. C. (2011). Effects of Task Complexity and Interaction on L2 Performance. In Robinson, P. (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 2, 141-173. Amsterdam: John Benjamins.
- Moattarian, A., Tahririan, M., Alibabae, A. (2019). Task Complexity Manipulation and EFL Learners' Interactions in the Process of Collaborative Pre-planning. *Applied Research on English Language*, 8(1), 51-78.
- Nasiri, M., & Atai, M. R. (2017). An Investigation into the Effects of Joint Planning on Complexity, Accuracy, and Fluency across Task Complexity. *Journal of English Language Teaching and Learning*, 9(20), 49-74.
- Nuevo, A. M. (2006). *Task Complexity and Interaction: L2 Learning Opportunities and Development*. Doctoral Dissertation, Georgetown University, USA.
- Ortega, L. (1999). Planning and Focus on Form in L2 Oral Performance. *Studies in Second Language Acquisition*, 21, 109–148.
- Rahimpour, M. (2007). Task Complexity and Variation in L2 Learners' Oral Discourse. *The University of Queensland Working Papers in Linguistics*, 1, 1-9. Retrieved from <https://www.library.uq.edu.au/>
- Robinson, P. (2001a). Task Complexity, Cognitive Resources, and Syllabus Design: A Triadic Framework for Examining Task Influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287- 318). Cambridge: CUP.

- Robinson, P. (2001b). Individual Differences, Cognitive Abilities, Aptitude Complexes, and Learning Conditions in Second Language Acquisition. *Second Language Research*, 17(4), 368-392.
- Robinson, P. (2001c). Task Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework. *Applied Linguistics*, 22(1), 27-57.
- Robinson, P. (2005). Cognitive Complexity and Task Sequencing: Studies in a Componential Framework for Second Language Task Design. *International Review of Applied Linguistics*, 43, 1–32
- Robinson, P. (2007). Task Complexity, Theory of Mind, and Intentional Reasoning: Effects on L2 Speech Production, Interaction, Uptake, and Perceptions of Task Difficulty. *International Review of Applied Linguistics*, 45, 237-257.
- Robinson, P. (2010). Situating and Distributing Cognition Across Task Demands: The SSARC Model of Pedagogic Task Sequencing. In Pütz, M. & Sicola, L. (Eds.), *Cognitive Processing in Second Language Acquisition* (pp. 243–268). Amsterdam, The Netherlands: Benjamins.
- Robinson, P. (2011). Second Language Task Complexity, the Cognition Hypothesis, Language Learning, and Performance. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 3–38). Philadelphia/Amsterdam: John Benjamins.
- Ryu, J. (2017). *Effects of Task Complexity and Planning on L2 Spoken Performance and Attentional Allocation to Meaning and Form of Korean High School Students*. Doctoral Dissertation, Seoul National University, Korea.
- Salimi, A. (2015). The Effect of Focus on Form and Task Complexity on L2 Learners' Oral Task Performance. *Advances in Language and Literary Studies*, 6(6), 54-62.
- Salimi, A., & Dadashpour, S. (2012). Task Complexity and SL Development: Does Task Complexity Matter? *Procedia-Social and Behavioral Sciences*, 46, 726-735.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan, P. (2001). Tasks and Language Performance Assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 167–185). New York/London: Longman.
- Skehan, P. (2003). Task-based Instruction. *Language Teaching*, 36, 1–14.

- Skehan, P. (2014). Limited Attentional Capacity, Second Language Performance, and Task-based Pedagogy. In P. Skehan (Ed.), *Processing Perspectives on Task Performance (Task-based Language Teaching)* (pp. 211–260). Philadelphia/Amsterdam: John Benjamins.
- Skehan, P., & Foster, P. (1997). Task Type and Task Processing Conditions as Influences on Foreign Language Performance. *Language Teaching Research*, 1(3), 185–211.
- Skehan, P., & Foster, P. (1999). The Influence of Task Structure and Processing Conditions on Narrative Retellings. *Language Learning*, 49(1), 93-120
- Skehan, P., & Foster, P. (2001). Cognition and Tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183–205). New York: Cambridge University Press.
- Wendel, J. (1997). *Planning and Second Language Narrative Production*. Doctoral Dissertation, Temple University, Philadelphia, USA.
- 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), 1-27.
- Zarei, A. (2013). The Effects of Task Complexity on English Language Learners' Listening Comprehension. *Issues in Language Teaching*, 1(2), 225-250.

