Journal of English language Teaching and Learning

University of Tabriz

Volume 13, Issue 28, (Fall & Winter 2021)

The Joint Effects of Teacher-led and Collaborative Planning conditions and Task Complexity on L2 Oral Production

Sanaz Mohazabieh

Department of English Language, Islamic Azad University, Shiraz Branch, Shiraz, Iran.
sanazmohazabieh@gmail.com

Rahman Sahragard *(Corresponding Author)*

Department of Foreign Languages and Linguistics, Shiraz University, Shiraz, Iran.
rahman.sahragard@gmail.com

Ehsan Rassaei

Majan University College, Mascat, Oman.
ehsanrassaei@yahoo.com

Mustafa Zamanian

Department of English Language Islamic Azad University, Shiraz Branch, Shiraz, Iran.
mustafazamanian@yahoo.com

ARTICLE INFO:

Document Type: **Research Paper**

Received date:
2021.06.08
Accepted date:
2021.11.16

Print ISSN: **2251-7995**
Online ISSN: **2676-6876**

Keywords:

Task Complexity, Strategic Planning, Oral Production, Accuracy, Fluency, Complexity, EFL learners

Abstract

This quantitative study aimed to investigate the combined effects of two types of strategic planning, namely collaborative and teacher-led planning conditions and task complexity on Iranian intermediate language learners' oral production in terms of complexity, accuracy, and fluency. To achieve this purpose, 90 EFL learners were selected through convenience sampling from a language institute in Shiraz, Iran, and randomly assigned to two control and four experimental groups. The study adopted a quasi-experimental design in the form of pretest, treatment, and posttest. In the first step, all participants took part in a speaking pretest in which they were required to narrate a story based on a series of picture description tasks. While the experimental groups underwent 10 treatment sessions of picture description task performance along with two planning types i.e., teacher-led and collaborative planning conditions, the control groups were not allowed to plan the task performance. In the last session, the language learners took a posttest whose results were compared with those of the pretest. The findings revealed that the language learners in the collaborative planning groups outperformed the other groups in terms of both fluency and complexity. Further, teacher-led groups did better than the other groups in terms of accuracy. This study carries crucial implications for EFL teachers, material developers, syllabus designers, and speaking skill examiners.

DOI: 10.22034/ELT.2021.46449.2401
Citation: Mohazabieh, S., Sahragard, R., Rassaei, E., Zamanian, M. (2021). The Joint Effects of Teacher-led and Collaborative Planning Conditions and Task Complexity on L2 Oral Production. *Journal of English Language Teaching and Learning*, 13(28), 219-244. Doi: 10.22034/ELT.2021.46449.2401

Introduction

In light of the pivotal role of tasks in language teaching and learning contexts, there has been a burgeoning interest in exploring the effects of manipulating task complexity along planning conditions on language learners' oral outputs in terms of complexity, accuracy, and fluency (henceforth, CAF). The literature abounds with studies examining the planning time conditions under which tasks are undertaken. To Ellis (2005), planning time conditions encompassing pre-task and within-task planning otherwise known as online planning are usually differentiated in terms of the time when the act of planning takes place. Pre-task planning occurs before task performance and is of two types i.e., strategic planning and rehearsal; the former allows learners to prepare to undertake the task by attending to the content to be encoded, while the latter entails the repetition of the task before the main task performance (Ellis, 2005, 2009). Strategic planning can be further divided into three sub-categories in terms of participatory structure; that is, whether planning is performed by learners working on their own, with the teacher, or collaboratively with other learners (Ellis, 2005).

Task complexity affects learners' language behavior and performance while engaging in task execution (Robinson, 2001c; Skehan, 1998). Defined as the outcome of the information processing demands including attention, memory, and reasoning imposed by the task structure on the learner (Robinson, 2001c), task complexity may be approximately cognitively complicated for learners due to the task design that influences how learners process information while undertaking a task. In essence, task complexity is conceptualized as the level of attention learners require while carrying out a task to attain an outcome (Skehan, 2001).

A review of the literature on task-based instruction reveals that most of the studies on planning have widely addressed individual planning. However, due to the collaborative nature of interactions in the classroom or real-life settings, language learners in pairs or groups, and even the teacher may be involved in the process of planning. Given these possibilities, it is worthwhile to examine the impacts of strategic planning that is teacher-fronted or group-based. Besides, no research studies to the best of the authors' knowledge have elaborated on the joint effects of strategic planning and task complexity on language learners' oral outputs with reference to CAF. To fill this lacuna in the literature, this line of research intends to investigate the simultaneous effects of task complexity along with two types of strategic planning, namely teacher-led and collaborative strategic planning on EFL learners' oral production. Consequently, this research study seeks to answer the following research question:

What are the combined effects of task complexity and strategic planning types i.e., teacher-led and collaborative planning on Iranian EFL learners' oral productions in terms of a) fluency, b) accuracy, and c) complexity?

The results of this study might contribute to the literature by providing an in-depth theoretical account of the impacts of strategic planning and task complexity on language use and acquisition. Along the same lines, given the salient roles of CAF in attaining greater language proficiency, it is important to examine how the interaction of strategic planning and task complexity can assist language learners to promote these three components of L2 oral output. Further, according to Ellis (2009), exploring the effects of different planning time conditions on task performance can conduce language teachers to decide whether or not to

afford learners with time for planning. Likewise, the results might inform language teachers and materials developers to draw on task difficulty as a tool for selecting tasks in accordance with language learners' proficiency level. Moreover, the findings might redound to researchers in creating a speaking or writing marking scheme for classroom settings or high-stakes examinations. Such a scheme can help teachers and speaking or writing examiners collectively attend to task complexity and strategic planning conditions.

Review of Literature

Pre-task Planning

Pre-task planning allows learners to map form and meaning using their language knowledge that is not operating automatically (Ellis, 2005). In practice, planned discourse helps learners extend what they intend to say (Foster & Skehan, 1999). Pre-task planning as a means of attaining a pedagogical focus-on-form can reduce the limited capacity of working memory by providing learners with the 'cognitive window' required to focus on form while learners are mostly attending to conveying a message (Ellis, 2005). However, according to Skehan (1998), the processing limitations of learners' working memory can be alleviated when presented with time to plan the content and form of a task strategically.

Pre-task planning makes learners equally enhance their attentional focus on meaning and form (Bygate, 2016; Skehan, 2014) and improves speech fluency by encouraging learners to deeply and meaningfully process the propositional content and linguistic forms of their speech; it allows learners to use the content and form of their speech production before generating it orally for real-life interaction (East, 2014; Long, 2015). To enhance various elements of speech production, strategic planning can be manipulated by guided planning which concerns directing learners' attention while preparing for a task (Ellis, 2009). In contrast, unguided planning devotes time to learners to plan on their own without any teacher-fronted guidance on language form or content. Accordingly, learners can draw on their linguistic knowledge to prepare for a task (Thompson, 2014).

Task Complexity

Task complexity concerns the cognitive dimensions of tasks, of which there are two distinct types: resource-directing and resource-dispersing dimensions (Robinson, 2005) integral to Robinson's (2007) Triadic Componential Framework. The former entails +/- here and now and +/- reasoning demand (Robinson, 2007) which turns the learner's attention to specific linguistic characteristics of a task and is associated with linguistic demands or the content occurring on learners manipulated by transforming its respective variables (Thompson, 2014). Robinson (2010) holds that increasing tasks along resource-directing factors can raise learners' attention to speech production and thereby generating syntactically complex language. Likewise, Kormos (2011) argues that complex tasks on resource-directing dimensions' trigger wider lexical variety and more complex syntactic structures. Reasoning demand relates to the resource-directing dimension in which tasks do not require reasoning on the side of learners; it involves a simple transfer of information and necessitates less linguistic effort and resources than tasks demanding reasoning (Robinson, 2005). Resource-dispersing factors, on the other hand, relate to the performance demands that occur on learners manipulated by altering the dimensions including strategic planning (Thompson, 2014) which distracts learner's attention

from various aspects of the task (Robinson, 2005). Giving no planning time planning during task performance leads to the complexity of a task in that it can disperse attention over the various elements of the task (Robinson, 2001a).

There are two main hypotheses associated with task complexity, namely Skehan's (1998) Limited Capacity Hypothesis and Robinson's (2001a, 2001b, 2003, 2005) Cognition Hypothesis. The former otherwise known as the Trade-off Hypothesis favors a single-resource model of attention based on which, carrying out a complicated task brings about trade-off effects between accuracy on the one hand and complexity on the other. Due to the limited attentional capacity for the form, a trade-off is made between accuracy and complexity (Michel, 2011). The latter on the other hand posits that learners can access multiple and noncompetitive pools of attention and that increasing task complexity leads to more accurate, complex, and lexically diverse language (Kuiken & Vedder, 2011). The Cognition Hypothesis supports a multiple-resources approach in which different aspects of language are attended to by learners while undertaking a cognitively complex task (Robinson, 2007, 2011). That is, there exists no trade-off between accuracy and complexity elements of language production (Robinson, 2011). In essence, increasing the cognitive demands of tasks would result in more attentional resources (Lee, 2018).

CAF Triad.

Language proficiency in L2 research is always addressed in terms of complexity, accuracy, and fluency (Ellis, 2003, 2008). In effect, CAF has been employed as a measure of learners' language output in several studies (Robinson, 2001a; Skehan, 1998; Yuan & Ellis, 2003, 2005).

Complexity relates to the number of clauses the learner links or uses within a sentence representing their interlanguage development and reconstruction (Skehan, 1996). Cognitive complexity refers to the difficulty learners encounter while processing language under different circumstances (Housen, Kuiken, & Vedder, 2012). Complexity is the extent to which learners generate complex language (Ellis & Barkhuizen, 2005) with the possibility that the elaborated language may not be controllable effectively (Skekan & Foster, 1999).

Accuracy is defined as the extent to which the language generated by learners complies with the norms of a target language (Yuan & Ellis, 2003). That is, accuracy points to the extent to which a language learner's productions diverge from target-language norms (Housen et al., 2012). Further, accuracy entails the correctness and appropriateness of language learners' speech utterances (Bulte & Housen, 2012).

Ellis & Barkhuizen (2005) define fluency as producing target language in real-time without excessive pausing and hesitation. Tavakoli & Skehan (2005) maintain that undue pausing is evaluated in terms of breakdown fluency measures that include the length and number of pauses. Lennon (2000) views fluency as the quick, continuous, correct, direct, and exact translation of flows of ideas or intentions into language under the time-related limitations of spontaneous processing

Empirical Studies

A few researchers (e.g., Gilabert, 2007; Kang, 2018; Moattarian, Tahririan, & Alibabaee, 2019; Nasiri & Atai, 2017; Yuan & Ellis, 2003) have examined planning time conditions and task complexity in oral production in terms of the CAF triad.

In a recent study, Mohazabieh, Sahragard, Rassaei, and Zamanian (2020) investigated the simultaneous effects of high and low task complexity levels and individual pre-task planning and on-line planning on Iranian EFL learners' oral performance in terms of CAF. They asked the study participants to narrate stories based on several pictures given to them as tasks. Their results demonstrated that the pre-task high complexity group performed better than the other groups regarding complexity and that pre-task planning impacted EFL learners' oral fluency. They also revealed that the pre-task high complexity group outperformed the pre-task low complexity group in terms of accuracy.

Moattarian et al. (2019) examined the effect of task complexity and collaborative pre-task planning. They selected 128 language learners with two different proficiency levels and asked them to perform three different tasks. Delving into the participants' interactions, they realized that undertaking tasks of a higher level of cognitive complexity led to more learning gains.

Kang (2018) also examined the impacts of individual, collaborative, and no planning on oral task performance of 65 Korean middle school learners along with simple and complex conditions of task complexity with respect to CAF. They revealed that collaborative planning greatly contributed to task completion and accuracy. Their results also showed that individual planning resulted in greater fluency than collaborative planning. Moreover, they did not report an interaction effect between the complexity of tasks and the planning conditions.

Nasiri and Atai (2017) explored the joint impacts of planning types including no planning, strategic planning, and online planning on learners' oral production regarding CAF. Their participants were 80 advanced EFL carrying out simple and complex narrative tasks. The results of their study revealed that strategic planning assisted learners to improve their complexity and fluency in simple tasks and only their fluency in complex tasks. Further, joint planning (strategic and online) resulted in the improvement of accuracy and complexity in the complex task, on the one hand, fluency and accuracy in the simple task on the other.

Exploring the effects of pre-task and online planning on oral productions, Yuan and Ellis (2003) carried out a study on 42 participants studying English in a university in China. They assigned the participants to three groups i.e., pre-task planning, online planning, and no planning to narrate a set of pictures. Their results suggested that pre-task planning enhanced complexity, whereas online planning affected accuracy and complexity. They also demonstrated that pre-task planning led to the production of more fluent language than online planning.

Operationalizing task complexity along planning time simultaneously, Gilabert (2007) utilized oral narrative tasks to examine the effects of these variables on oral productions of 48 intermediate language learners selected from Ramon Llull University in Barcelona. The results manifested that simple and complex narrative tasks carried out under planned conditions

triggered more lexically complex oral output and more focus on form and fluency being negatively affected.

Method

Design

The study sought to explore the joint impacts of two strategic planning types, i.e., collaborative and teacher-led planning conditions and task complexity on language learners' oral production. The study adopted a quasi-experimental design in which the language learners were non-randomly selected, were made homogeneous in terms of their proficiency level, and were randomly assigned to four experimental and two control groups. The study employed a 3*2 factorial design encompassing three levels for planning conditions and two levels for task complexity including high and low complex tasks. The rationale for employing the quasi-experimental design was that the randomization of participants was not feasible.

Participants

Initially, 102 Iranian intermediate language learners were selected based on convenience sampling among six intact classes from a private language institute in Shiraz, Iran. To homogenize the language learners, an Oxford Placement Test was administered to the language learners. Analyzing the results of the placement test, the main researcher(the teacher) selected 90 participants(female = 53 and male= 37) obtaining scores between one standard deviation above and below the mean and randomly assigned them to two control and four experimental groups of 15 language learners each. They were all Persian native speakers with the age range of 16 to 45. Noteworthy to mention is that all the participants completed consent forms and agreed to take part in the study.

Instruments

Oxford Placement Test

The first instrument of the study was an Oxford Placement Test employed to homogenize the participants in terms of their language proficiency. It should be noted that similar studies (e.g., Farrokhi, & Sattarpour, 2017; Gilabert, 2007; Salimi, 2015) have used placement tests to control learners' language proficiency levels to ease the comparison of reported findings across similar research.

Oral Presentation Tasks

Oral presentation tasks in the form of scrambled and unscrambled pictures were the second instruments of the study presented to language learners to high and low complex task groups, respectively to narrate stories. The use of pictures has been a common way for narrative tasks in several research studies (e.g., Ellis & Yuan 2004; Ishhikawa, 2006). These types of tasks are, however, more cognitively demanding than other tasks (Skehan & Foster, 1997).

Measures of Learners' Oral Production

Fluency

In this study, fluency was determined in terms of repair fluency measured by counting the number of repeated words or phrases, false starts, phrases or clauses repeated with some syntactical, morphological reformulations, and replacements of some lexical items for others (Elder & Iwashita, 2005; Skehan & Foster, 1999).

Accuracy

Accuracy relates to the ability to generate grammatically correct utterances (Housen & Kuiken, 2009). In the present study, accuracy was gauged by counting the number of target-like clauses and dividing them by the total number of clauses. Some scholars (e.g., Yuan & Ellis, 2003) have also used this accuracy measure.

Complexity

Complexity, in this study, concerned the number of clauses per Communication-Unit (C-unit) gauged by dividing the number of clauses in the subjects' oral production by the number of C-units depicting independent utterances that reflect referential or pragmatic meaning (Foster & Skehan, 1996).

Procedure

Initially, a pretest in the form of the monologic narrative task was administered to all participants to gauge their speaking ability. While the experimental groups received 10 intervention sessions concerning low and high task complexity levels and strategic planning types in the form of collaborative and teacher-led planning, the two control groups entailing low and high task complexity groups received no intervention.

A small-scale pilot study was performed with 12 intermediate language learners from the same language institute with similar educational backgrounds. They undertook low and complex tasks under the two types of strategic planning and no planning conditions. Following the pilot study, we decided to allocate 10 minutes to strategic planners and 30 seconds to no planners in the main study to carry out the task.

The control groups were presented with a short introduction to task performance to inform them that they did not need to do planning prior to performing tasks. In line with Foster and Skehan (1996), and Yuan and Ellis (2003), the strategic planning groups were allocated 10 minutes to reflect on the content and language and to plan their speaking tasks. Accordingly, the experimental groups received some guidelines as to how to do planning within 10 minutes before the speaking task. The participants of the collaborative planning groups were instructed on how to do planning in groups of four. They were then seated in pairs of four and discussed the tasks. Along the same lines, the teacher-led strategic group received teacher-fronted assistance on the content and form and the planning process was led by the teacher. To best take advantage of strategic planning types, the experimental groups' participants were allowed to take notes concerning what they wished to say but were told that they could not use the notes before their oral performance.

The subjects of the study were informed of the true purpose of the narrative tasks and assured that their data in this regard would not affect their final course grades. Nonetheless, the prime purpose of the study was not identified to mitigate participant bias.

Picture narration tasks chosen for this study were in line with Robinson's (2001a) task complexity criteria. The language learners performing cognitively complex tasks were required to identify the correct order of the pictures while narrating them. The language learners could differently interpret the narrative comic strips. As for the complex task groups, unscrambling the pictures could increase the complexity of the tasks in question. Further, such tasks required

varying levels of attention of language learners with unclear and predictable information giving rise to a greater cognitive load and thereby influencing the performance of the task (Foster & Skehan, 1996).

The picture strips utilized as the narrative tasks in this study were taken from Quino, an Argentinian cartoonist. The selection of these tasks could be justified on two grounds. Firstly, other researchers (e.g., Abdoahzadeh & Fard Kashani, 2012; Heidari-Shahreza, Dabaghi, & Kassaian, 2011; Kim, 2009; Nuevo, 2006; Robinson, 2001a) employed similar types of tasks which can facilitate the comparison of oral performance results. Secondly, these tasks are of mono-logic nature that forms a foundation for creating measures of learner performance uninfluenced by dialogic factors.

The study was conducted in 10 sessions of one hour and 45 minutes (two sessions a week). The lead researcher (the teacher) allocated about one hour to cover the main coursebook, namely, Touchstone Series Book 3. Approximately 30 to 45 minutes was devoted to commencing the instructional interventions by performing high and low tasks under different strategic planning conditions. To investigate the effects of the treatment, a speaking post-test was administered to both control and experimental groups. In the posttest, the participants in each group narrated a story based on their relevant planning type and task complexity level. As for the high complex tasks, the learners ordered the frames of the comic strip based on their occurrence and narrated the story. The subjects in the low complex task groups, on the other hand, narrated a set of ordered pictures. Their oral outputs were then recorded to be codified and assessed by three different raters. The maximum score was decided to be 20. The post-test scores were finally compared with those of the pretest scores to investigate the effectiveness of the instructional interventions.

Data Analysis

After collecting the data, a Pearson Product Moment Correlation Coefficient was run to check the inter-rater reliability of the pretest and posttest scores of the three raters. The scores reported in this study are the means of scores assigned by the three raters. Further, the normality of the data was checked using the Kolmogorov-Smirnov[a] and the Shapiro-Wilk and the assumption of normality of the data was retained for all groups' pretest and posttest scores. Likewise, descriptive statistics were performed for the participants' pretest and posttest scores in terms of CAF. One-way ANOVA was performed on pretest scores to examine whether there existed potentially significant differences among the groups. Also, mixed between-within groups ANOVAs were carried out on the participants' CAF scores, with the combination of task complexity and strategic planning type (no-planning low complexity, no-planning high complexity, collaborative planning low complexity, collaborative planning high complexity, teacher-led planning low complexity, teacher-led planning high complexity) and time entailing pretest/ posttest and the participants' scores considered as independent and dependent variables, respectively. To locate the exact differences among the groups, One-way ANOVAs and Tukey's pairwise post hoc comparisons were also run on the participants' posttest scores.

Results

The research question intended to examine whether the combination of task complexity and strategic planning types including teacher-led and collaborative planning conditions affect Iranian EFL learners' oral production concerning a) fluency, b) accuracy, and c) complexity.

Results of the Normality Test

In the first place, the normality of the data was probed by running the Kolmogorov-Smirnov and the Shapiro-Wilk tests (Table 1).

Table 1. *Normality Tests of the Pre and Post-test Scores*

		Groups	Kolmogorov-Smirnov[b]			Shapiro-Wilk		
			Statistic	df	Sig.	Statistic	df	Sig.
Fluency	Pretest	Low-collaborative	.124	15	.200*	.966	15	.789
		Low, no-planning	.176	15	.200*	.942	15	.409
		High, no-planning	.112	15	.200*	.945	15	.451
		Low, teacher-led	.146	15	.200*	.948	15	.492
		High, teacher-led	.144	15	.200*	.928	15	.253
		High-collaborative	.131	15	.200*	.964	15	.754
	Post-test	Low-collaborative	.185	15	.180	.940	15	.385
		Low, no-planning	.184	15	.184	.941	15	.397
		High, no-planning	.129	15	.200*	.957	15	.633
		Low, teacher-led	.136	15	.200*	.941	15	.389
		High, teacher-led	.114	15	.200*	.947	15	.482
		High-collaborative	.158	15	.200*	.947	15	.477
Accuracy	Pretest	Low-collaborative	.143	15	.200*	.919	15	.185
		Low, no-planning	.150	15	.200*	.966	15	.803
		High, no-planning	.128	15	.200*	.955	15	.609
		Low, teacher-led	.171	15	.200*	.962	15	.723
		High, teacher-led	.150	15	.200*	.943	15	.423
		High-collaborative	.134	15	.200*	.960	15	.691
	Post-test	Low-collaborative	.108	15	.200*	.968	15	.821
		Low, no-planning	.151	15	.200*	.939	15	.368
		High, no-planning	.208	15	.081	.875	15	.079
		Low, teacher-led	.172	15	.200*	.919	15	.185
		High, teacher-led	.208	15	.081	.951	15	.544
		High-collaborative	.118	15	.200*	.977	15	.949
Complexity	Pretest	Low-collaborative	.101	15	.200*	.968	15	.828
		Low, no-planning	.091	15	.200*	.988	15	.998
		High, no-planning	.158	15	.200*	.937	15	.344
		Low, teacher-led	.233	15	.077	.883	15	.053
		High, teacher-led	.201	15	.106	.922	15	.203
		High-collaborative	.153	15	.200*	.920	15	.194
	Post-test	Low-collaborative	.141	15	.200*	.952	15	.564
		Low, no-planning	.129	15	.200*	.955	15	.606
		High, no-planning	.106	15	.200*	.972	15	.882
		Low, teacher-led	.119	15	.200*	.953	15	.578
		High, teacher-led	.124	15	.200*	.936	15	.332
		High-collaborative	.138	15	.200*	.936	15	.336

As displayed in Table 1, the non-significant values which are all more than the significance level (p<0.05) demonstrate that the present data i.e., the pre and post-test scores of all groups were normally distributed and parametric tests can be utilized to address the research question.

Fluency

Table 2 displays the descriptive statistics for all groups' pre-tests and post-tests on fluency.

Table 2. *Descriptive Statistics of Fluency Pre and Post-test Scores*

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Pretest	Low-collaborative	15	14.9144	.76987	.19878	14.4880	15.3407	13.35	16.23
	Low, no-planning	15	14.4734	.89766	.23177	13.9763	14.9705	12.94	15.84
	High-no-planning	15	14.3803	1.46859	.37919	13.5670	15.1936	12.01	16.54
	Low, teacher-led	15	14.6297	.97494	.25173	14.0898	15.1696	13.20	16.24
	High, teacher-led	15	14.1767	1.00148	.25858	13.6221	14.7313	12.75	16.68
	High-collaborative	15	14.6141	.92339	.23842	14.1028	15.1255	12.61	15.93
	Total	90	14.5314	1.02654	.10821	14.3164	14.7464	12.01	16.68
Post-test	Low-collaborative	15	18.8565	.60379	.15590	18.5221	19.1909	17.46	19.85
	Low, no-planning	15	16.0786	.94736	.24461	15.5540	16.6032	14.74	17.82
	High, no-planning	15	15.4881	1.69538	.43775	14.5493	16.4270	13.09	18.44
	Low, teacher-led	15	16.1554	1.09514	.28276	15.5489	16.7619	14.21	17.65
	High-teacher-led	15	15.5654	1.23467	.31879	14.8817	16.2491	12.97	17.16
	High-collaborative	15	17.4263	.95213	.24584	16.8990	17.9536	15.99	18.91
	Total	90	16.5951	1.63181	.17201	16.2533	16.9368	12.97	19.85

To examine if there were any significant differences between the performance of the groups before the treatment, a one-way ANOVA was carried out. The results are demonstrated in Table 3.

Table 3. O*ne-way ANOVA concerning the Difference between Groups with Regard to Fluency Pretest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	4.728	5	.946	.892	.490
	Within Groups	89.059	84	1.060		
	Total	93.787	89			

The results of the one-way ANOVA indicated that there was not any significant difference between the groups with respect to the fluency pretest scores (F (5, 84) =.89, p=.49). This indicates that the groups enjoyed homogeneity in terms of fluency prior to the treatment.

To further investigate whether the treatment impinged upon the experimental groups' fluency scores over time, a Mixed between-within groups ANOVA was conducted. To examine the homogeneity of variances of the groups and covariance matrices, Levene's test and Box's test were run, respectively.

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

	F	df1	df2	Sig.
Pretest	2.061	5	84	.078

Post-test scores	4.478	5	84	.061

As can be seen in Table 4, the differences between the groups' variances on fluency pretest (F (5, 84) = 2.06, p > .05) and post-test (F (5, 84) = 4.47, p > .05) were not significant.

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

Box's M	21.526
F	1.348
df1	15
df2	38594.288
Sig.	.164

The results as depicted in Table 5 (Box's M = 21.52, p> .001) revealed that the assumption of homogeneity of covariance matrices was retained.

In addition, the Multivariate test was run to assess the effect of the interaction of task complexity and strategic planning on the fluency of the participants' performance over time.

Table 6. *Multivariate Tests for Fluency Pre and Post-test Scores*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.848	467.348[c]	1.000	84.000	.000	.848
	Wilks' Lambda	.152	467.348[c]	1.000	84.000	.000	.848
	Hotelling's Trace	5.564	467.348[c]	1.000	84.000	.000	.848
	Roy's Largest Root	5.564	467.348[c]	1.000	84.000	.000	.848
Time * Groups	Pillai's Trace	.565	21.793[c]	5.000	84.000	.000	.565
	Wilks' Lambda	.435	21.793[c]	5.000	84.000	.000	.565
	Hotelling's Trace	1.297	21.793[c]	5.000	84.000	.000	.565
	Roy's Largest Root	1.297	21.793[c]	5.000	84.000	.000	.565

As revealed in Table 6, the main effects for time, Wilk's Lambda= .15, (F (1, 84) = 467.34, p < .001) were statistically significant. Following Cohen's (1988) criterion, the effect size (partial eta squared= .84) suggested a large actual difference in mean scores over time from the pre to the post-test. There was also a significant interaction between time and combinations of task complexity and planning type, Wilk's Lambda= .43, (F (5, 84) = 21.79, p < .001), partial eta squared=.56 (large effect size). This implies that language learners benefited differentially from the combinations of task complexity and planning type.

The Tests of the between-subjects effects were also performed to explore if there were significant differences between the groups on fluency, regardless of time.

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	43598.593	1	43598.593	22420.051	.000	.996
Groups	88.304	5	17.661	9.082	.000	.351
Error	163.349	84	1.945			

As Table 7 illustrated, the main effect of comparing the six types of combinations was significant, F (5, 84) = 9.08, p<.05, partial eta squared=.35 (large effect size) reflecting a statistically significant difference between the effectiveness of the six types of combinations of task complexity and the planning type.

In the next step, to specify the differences between the six groups, a One-way ANOVA and Tukey's Post hoc comparisons were conducted. Tables 8 and 9 demonstrate the results of the One-way ANOVA and Post hoc tests, respectively.

Table 8. O*ne-way ANOVA concerning the Difference between Groups in terms of Fluency Posttest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Posttest scores	Between Groups	128.257	5	25.651	19.816	.000
	Within Groups	108.733	84	1.294		
	Total	236.990	89			

As presented in Table 8, the One-way ANOVA generated significant results (F (5, 84) = 19.81, p<.001), signaling that there existed a significant difference among the groups concerning the post-test scores.

Table 9. *Post-Hoc Tukey HSD Test of the Groups' Fluency Post-test Scores*

Dependent Variable	(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Post-test scores	Low-collaborative	Low, no-planning	2.77786*	.41544	.000	1.5662	3.9895
		High, no-planning	3.36834*	.41544	.000	2.1567	4.5800
		Low, teacher-led	2.70110*	.41544	.000	1.4894	3.9128
		High, teacher-led	3.29107*	.41544	.000	2.0794	4.5027
		High-collaborative	1.43020*	.41544	.011	.2185	2.6419
	Low, no-planning	Low-collaborative	-2.77786*	.41544	.000	-3.9895	-1.5662
		High, no-planning	.59048	.41544	.714	-.6212	1.8021
		Low, teacher-led	-.07676	.41544	1.000	-1.2884	1.1349
		High, teacher-led	.51321	.41544	.818	-.6984	1.7249
		High-collaborative	-1.34766*	.41544	.020	-2.5593	-.1360
	High, no-planning	Low-collaborative	-3.36834*	.41544	.000	-4.5800	-2.1567
		Low, no-planning	-.59048	.41544	.714	-1.8021	.6212
		Low, teacher-led	-.66724	.41544	.597	-1.8789	.5444
		High, teacher-led	-.07727	.41544	1.000	-1.2889	1.1344
		High-collaborative	-1.93814*	.41544	.000	-3.1498	-.7265
	Low, teacher-led	Low-collaborative	-2.70110*	.41544	.000	-3.9128	-1.4894
		Low, no-planning	.07676	.41544	1.000	-1.1349	1.2884
		High, no-planning	.66724	.41544	.597	-.5444	1.8789
		High, teacher-led	.58997	.41544	.715	-.6217	1.8016
		High-collaborative	-1.27090*	.41544	.034	-2.4826	-.0592
	High, teacher-led	Low-collaborative	-3.29107*	.41544	.000	-4.5027	-2.0794
		Low, no-planning	-.51321	.41544	.818	-1.7249	.6984
		High, no-planning	.07727	.41544	1.000	-1.1344	1.2889
		Low, teacher-led	-.58997	.41544	.715	-1.8016	.6217
		High-collaborative	-1.86087*	.41544	.000	-3.0725	-.6492

High-collaborative	Low-collaborative	-1.43020*	.41544	.011	-2.6419	-.2185
	Low, no-planning	1.34766*	.41544	.020	.1360	2.5593
	High, no-planning	1.93814*	.41544	.000	.7265	3.1498
	Low, teacher-led	1.27090*	.41544	.034	.0592	2.4826
	High, teacher-led	1.86087*	.41544	.000	.6492	3.0725

The Post-hoc comparisons using the Tukey HSD test indicated that language learners in the collaborative planning low or high complexity groups were significantly more fluent than the other groups. Additionally, the collaborative planning low complexity group (M= 18.85, SD= .60) significantly outperformed the collaborative planning high complexity group (M= 17.42, SD= .95) in terms of fluency.

Accuracy

Table 10 gives an overview of descriptive statistics for the accuracy scores in the pre and post-tests.

Table 10. *Descriptive Statistics of Accuracy Pre and Post-test Scores*

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Pretest	Low-collaborative	15	14.5587	1.09518	.28277	13.9522	15.1652	12.82	15.95
	Low, no-planning	15	14.7068	1.41802	.36613	13.9215	15.4921	12.32	17.48
	High, no-planning	15	14.6568	.99079	.25582	14.1082	15.2055	13.19	16.88
	Low, teacher-led	15	14.0938	1.01072	.26097	13.5340	14.6535	12.42	16.25
	High, teacher-led	15	14.7030	1.13852	.29396	14.0725	15.3334	12.66	16.30
	High-collaborative	15	14.6511	.99666	.25734	14.0991	15.2030	13.19	16.40
	Total	90	14.5617	1.10767	.11676	14.3297	14.7937	12.32	17.48
Post-test scores	Low-collaborative	15	15.6625	1.04719	.27038	15.0826	16.2424	13.45	17.27
	Low, no-planning	15	15.9216	1.26478	.32656	15.2212	16.6221	13.29	17.57
	High, no-planning	15	15.8902	.86537	.22344	15.4110	16.3694	14.93	17.34
	Low, teacher-led	15	18.4374	.82926	.21411	17.9781	18.8966	16.27	19.94
	High, teacher-led	15	17.3676	.83329	.21516	16.9061	17.8290	16.04	18.91
	High-collaborative	15	16.0930	.73081	.18869	15.6883	16.4977	14.61	17.31
	Total	90	16.5620	1.36531	.14392	16.2761	16.8480	13.29	19.94

To investigate the potentially significant differences between the groups regarding the pretest accuracy scores, a one-way ANOVA was conducted.

Table 11. O*ne-way ANOVA concerning the Difference between Groups regarding Accuracy Pretest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	4.155	5	.831	.665	.651
	Within Groups	105.042	84	1.250		
	Total	109.197	89			

According to Table 11, the results did not reveal any significant difference among the groups (F (5, 84) =.66, p=.65), confirming the homogeneity of the groups with regard to the fluency before the treatment.

To investigate the effects of the treatment on the learners' complexity over time, a Mixed between-within groups ANOVA was run on the experimental and control groups' complexity pretest and post-test scores. The assumptions of the Mixed between-within groups ANOVA i.e., the homogeneity of variances and homogeneity of covariance matrices were first checked utilizing Levene's test and Box's test, respectively.

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

	F	df1	df2	Sig.
Pretest	.576	5	84	.718
Posttest	1.499	5	84	.199

As revealed in Table 12, no significant difference was observed between the groups' variances on accuracy pretest (F (5, 84) = .57, p > .05) and post-test scores (F (5, 84) = 1.49, p > .05).

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

Box's M	25.121
F	1.573
df1	15
df2	38594.288
Sig.	.081

The results of Box's test confirmed the homogeneity of covariance matrices (M = 25.12, p > .001). Next, the Multivariate test was performed to assess the effect of the interaction of task complexity and strategic planning on the accuracy of the participants' performance over time. Table 14 depicts the results of the Multivariate test.

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

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.822	388.866[c]	1.000	84.000	.000	.822
	Wilks' Lambda	.178	388.866[c]	1.000	84.000	.000	.822
	Hotelling's Trace	4.629	388.866[c]	1.000	84.000	.000	.822
	Roy's Largest Root	4.629	388.866[c]	1.000	84.000	.000	.822
Time * Groups	Pillai's Trace	.614	26.735[c]	5.000	84.000	.000	.614
	Wilks' Lambda	.386	26.735[c]	5.000	84.000	.000	.614
	Hotelling's Trace	1.591	26.735[c]	5.000	84.000	.000	.614
	Roy's Largest Root	1.591	26.735[c]	5.000	84.000	.000	.614

Table 14 confirmed that there existed a significant main effects for time, Wilk's Lambda= .17, F (1, 84) =388.86, p < .001, partial eta squared= .82 displaying a large effect size, and the interaction between time and the combination of task complexity and planning type, Wilk's Lambda= .38, F (5, 84) = 26.73, p < .001, partial eta squared= .61 representing a large effect size.

A Test of the Between-Subjects main effect was also conducted to identify the differences between the groups on accuracy regardless of time.

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	43590.916	1	43590.916	25931.653	.000	.997
Groups	33.103	5	6.621	3.938	.003	.190
Error	141.203	84	1.681			

The results of the Between-Subjects Effects Test ($F (5, 84) = 3.93$, $p < .05$, partial eta squared = .19 (indicating a large effect size) as shown in Table 15 revealed that there existed significant differences in the effectiveness of the six types of combinations of task complexity and planning type. To discover whether there were any meaningful differences between the six groups with regard to the accuracy post-test scores, a One-way ANOVA and Tukey's Post hoc comparisons were run.

Table 16. O*ne-way ANOVA concerning the Difference between Groups in terms of Accuracy Post-test Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Post-test scores	Between Groups	90.845	5	18.169	20.334	.000
	Within Groups	75.058	84	.894		
	Total	165.903	89			

According to Table16, a statistically significant difference was detected among the groups' post-test scores on accuracy ($F (5, 84) = 20.33$, $p < .001$).

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

Dependent Variable	(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Post-test scores	Low-collaborative	Low, no-planning	-.25914	.34517	.975	-1.2658	.7476
		High, no-planning	-.22769	.34517	.986	-1.2344	.7790
		Low, teacher-led	-2.77485*	.34517	.000	-3.7815	-1.7682
		High, teacher-led	-1.70506*	.34517	.000	-2.7118	-.6984
		High-collaborative	-.43051	.34517	.812	-1.4372	.5762
	Low, no-planning	Low-collaborative	.25914	.34517	.975	-.7476	1.2658
		High, no-planning	.03145	.34517	1.000	-.9752	1.0381
		Low, teacher-led	-2.51572*	.34517	.000	-3.5224	-1.5090
		High, teacher-led	-1.44593*	.34517	.001	-2.4526	-.4392
		High-collaborative	-.17138	.34517	.996	-1.1781	.8353
	High, no-planning	Low-collaborative	.22769	.34517	.986	-.7790	1.2344
		Low, no-planning	-.03145	.34517	1.000	-1.0381	.9752
		Low, teacher-led	-2.54716*	.34517	.000	-3.5539	-1.5405
		High, teacher-led	-1.47737*	.34517	.001	-2.4841	-.4707
		High-collaborative	-.20282	.34517	.992	-1.2095	.8039
	Low, teacher-led	Low-collaborative	2.77485*	.34517	.000	1.7682	3.7815
		Low, no-planning	2.51572*	.34517	.000	1.5090	3.5224
		High, no-planning	2.54716*	.34517	.000	1.5405	3.5539
		High, teacher-led	1.06979*	.34517	.031	.0631	2.0765
		High-collaborative	2.34434*	.34517	.000	1.3376	3.3510
		Low-collaborative	1.70506*	.34517	.000	.6984	2.7118

	High, teacher-led	Low, no-planning	1.44593*	.34517	.001	.4392	2.4526
		High, no-planning	1.47737*	.34517	.001	.4707	2.4841
		Low, teacher-led	-1.06979*	.34517	.031	-2.0765	-.0631
		High-collaborative	1.27455*	.34517	.005	.2679	2.2812
	High-collaborative	Low-collaborative	.43051	.34517	.812	-.5762	1.4372
		Low, no-planning	.17138	.34517	.996	-.8353	1.1781
		High, no-planning	.20282	.34517	.992	-.8039	1.2095
		Low, teacher-led	-2.34434*	.34517	.000	-3.3510	-1.3376
		High, teacher-led	-1.27455*	.34517	.005	-2.2812	-.2679

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

The results of the Post-hoc test in Table 17 revealed that teacher-led planning low and high complexity groups significantly outperformed the other groups. Furthermore, the results indicated that the learners in the teacher-led planning low complexity group (M= 18.43, SD= .82) were significantly more accurate than those in the teacher-led planning high complexity group (M= 17.36, SD= .83).

Complexity

Table 18 presents the descriptive statistics for the pre and post-test scores on complexity.

Table 18. *Descriptive Statistics of Complexity Pre and Post-test Scores*

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Pretest	Low-collaborative	15	14.6667	1.16260	.30018	14.0229	15.3105	12.14	16.42
	Low, no-planning	15	14.2038	1.31862	.34047	13.4736	14.9340	11.48	16.83
	High, no-planning	15	14.1232	1.14176	.29480	13.4909	14.7555	12.45	15.95
	Low, teacher-led	15	14.2556	.95347	.24618	13.7276	14.7836	13.02	16.78
	High, teacher-led	15	13.8884	.80278	.20728	13.4438	14.3329	12.30	15.25
	High-collaborative	15	14.2917	1.31877	.34050	13.5614	15.0221	10.83	16.31
	Total	90	14.2382	1.12414	.11849	14.0028	14.4737	10.83	16.83
Posttest scores	Low-collaborative	15	17.1532	.99622	.25722	16.6015	17.7049	15.35	19.18
	Low, no-planning	15	15.5487	1.30572	.33714	14.8256	16.2718	13.43	17.65
	High, no-planning	15	15.7609	.81630	.21077	15.3089	16.2130	14.24	17.09
	Low, teacher-led	15	15.7045	.92013	.23758	15.1950	16.2141	14.45	17.49
	High, teacher-led	15	15.8811	1.21500	.31371	15.2083	16.5540	14.14	18.85
	High-collaborative	15	18.3271	1.12303	.28996	17.7052	18.9490	16.28	19.94
	Total	90	16.3959	1.46025	.15392	16.0901	16.7018	13.43	19.94

To investigate whether there were any differences between the groups in terms of complexity before the treatment, a one-way ANOVA was conducted.

Table 19. O*ne-way ANOVA concerning the Difference between Groups regarding Complexity Pretest Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	4.854	5	.971	.758	.583
	Within Groups	107.614	84	1.281		
	Total	112.468	89			

As shown in Table 19, the One-way ANOVA on pretest complexity scores did not show any statistically significant difference among the groups (F (5, 84) =.75, p=.58), reflecting that the language learners in all groups were homogenous with respect to complexity before the treatment.

Next, a Mixed between-within groups ANOVA was performed to examine the effects of the treatment on the experimental and control groups' scores on complexity over time. Before conducting the Mixed between-within groups ANOVA, the Levene's test and Box's test were performed to check the groups' homogeneity of variances and homogeneity of covariance matrices, respectively.

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

	F	df1	df2	Sig.
Pretest	.672	5	84	.646
Posttest	.794	5	84	.557

According to Table 20, there existed no significant differences between the groups' variances on fluency pretest (F (5, 84) = .67, p > .05) and posttest scores (F (5, 84) = .79, p > .05).

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

Box's M	15.201
F	.952
df1	15
df2	38594.288
Sig.	.504

As revealed in Table 21, the homogeneity of covariance matrices was retained (M = 15.20, p > .001).

To investigate the effect of the interaction of task complexity and strategic planning on the complexity of the participants' performance over time, a Multivariate test was run. The results of the Multivariate test are displayed in Table 22.

Table 22. *Multivariate Tests for Complexity Pretest and Post-test Scores*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.863	527.744[c]	1.000	84.000	.000	.863
	Wilks' Lambda	.137	527.744[c]	1.000	84.000	.000	.863
	Hotelling's Trace	6.283	527.744[c]	1.000	84.000	.000	.863
	Roy's Largest Root	6.283	527.744[c]	1.000	84.000	.000	.863
Time * Groups	Pillai's Trace	.534	19.250[c]	5.000	84.000	.000	.534
	Wilks' Lambda	.466	19.250[c]	5.000	84.000	.000	.534
	Hotelling's Trace	1.146	19.250[c]	5.000	84.000	.000	.534
	Roy's Largest Root	1.146	19.250[c]	5.000	84.000	.000	.534

Building upon the results of the Mixed between-within groups ANOVA, it can be conceivably contended that there were significant main effects for time, Wilk's Lambda= .13, F (1, 84) = 527.74, p < .001, partial eta squared= .86 suggesting a large effect size, and the interaction between time and the combination of task complexity and planning type, Wilk's Lambda= .46, F (5, 84) = 19.25, p < .001, partial eta squared= .53 indicating a large effect size.

In the next step, the test of the between-subjects main effect was conducted to determine whether there were significant differences between the groups on the complexity scores, regardless of time. Table 23 displays the results of the Tests of Between-Subjects Effects.

Table 23. *Tests of Between-Subjects Effects on Complexity Scores*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	42230.396	1	42230.396	20679.869	.000	.996
Groups	59.152	5	11.830	5.793	.000	.256
Error	171.537	84	2.042			

The results of the Tests of Between-Subjects Effects indicated a statistically significant difference in the effectiveness of the six types of combinations of task complexity and planning type (F (5, 84) = 5.79, p < .05, partial eta squared = .25 (showing a large effect size)). A One-way ANOVA and Tukey's Post hoc comparisons were performed to specify the differences between the six groups.

Table 24. O*ne-way ANOVA concerning the Difference between Groups regarding Complexity Post-test Scores*

		Sum of Squares	df	Mean Square	F	Sig.
Post-test scores	Between Groups	92.508	5	18.502	15.978	.000
	Within Groups	97.269	84	1.158		
	Total	189.776	89			

Table 24 demonstrated that there were significant differences among the groups in the post-test complexity scores, *F* (5, 84) = 15.97, *p* < .001. Tukey's Post Hoc test was employed to locate the differences among the groups (Table 25).

Table 25. *Tukey's Post Hoc Comparisons of the Groups' Complexity Post-test Scores*

Dependent Variable	(I) Groups	(J) Groups	Std. Error	Sig.	95% Confidence Interval

			Mean Difference (I-J)			Lower Bound	Upper Bound
Post-test scores	Low-collaborative	Low, no-planning	1.60456*	.39293	.001	.4586	2.7506
		High, no-planning	1.39231*	.39293	.008	.2463	2.5383
		Low, teacher-led	1.44870*	.39293	.005	.3027	2.5947
		High, teacher-led	1.27210*	.39293	.021	.1261	2.4181
		High-collaborative	-1.17390*	.39293	.041	-2.3199	-.0279
	Low, no-planning	Low-collaborative	-1.60456*	.39293	.001	-2.7506	-.4586
		High, no-planning	-.21225	.39293	.994	-1.3583	.9337
		Low, teacher-led	-.15586	.39293	.999	-1.3019	.9901
		High, teacher-led	-.33246	.39293	.958	-1.4785	.8135
		High-collaborative	-2.77846*	.39293	.000	-3.9245	-1.6325
	High, no-planning	Low-collaborative	-1.39231*	.39293	.008	-2.5383	-.2463
		Low, no-planning	.21225	.39293	.994	-.9337	1.3583
		Low, teacher-led	.05639	.39293	1.000	-1.0896	1.2024
		High, teacher-led	-.12021	.39293	1.000	-1.2662	1.0258
		High-collaborative	-2.56621*	.39293	.000	-3.7122	-1.4202
	Low, teacher-led	Low-collaborative	-1.44870*	.39293	.005	-2.5947	-.3027
		Low, no-planning	.15586	.39293	.999	-.9901	1.3019
		High, no-planning	-.05639	.39293	1.000	-1.2024	1.0896
		High, teacher-led	-.17661	.39293	.998	-1.3226	.9694
		High-collaborative	-2.62260*	.39293	.000	-3.7686	-1.4766
	High, teacher-led	Low-collaborative	-1.27210*	.39293	.021	-2.4181	-.1261
		Low, no-planning	.33246	.39293	.958	-.8135	1.4785
		High, no-planning	.12021	.39293	1.000	-1.0258	1.2662
		Low, teacher-led	.17661	.39293	.998	-.9694	1.3226
		High-collaborative	-2.44600*	.39293	.000	-3.5920	-1.3000
	High-collaborative	Low-collaborative	1.17390*	.39293	.041	.0279	2.3199
		Low, no-planning	2.77846*	.39293	.000	1.6325	3.9245
		High, no-planning	2.56621*	.39293	.000	1.4202	3.7122
		Low, teacher-led	2.62260*	.39293	.000	1.4766	3.7686
		High, teacher-led	2.44600*	.39293	.000	1.3000	3.5920

The results of the post-hoc test revealed that both collaborative planning low and high complexity groups significantly outperformed all other groups with regard to complexity after the treatment. Additionally, the participants of the collaborative planning high complexity group (M= 18.32, SD= 1.12) were significantly more accurate than those of the collaborative planning low complexity group (M= 17.15, SD= .99).

Discussion

The current study aimed to investigate the combined effects of strategic planning types including collaborative and teacher-led planning conditions and task complexity on language learners' oral productions in terms of CAF. Overall, one of the more important similarities of our study with previous research (e.g. Foster & Skehan, 1996, 1999; Yuan & Ellis, 2003) is that strategic planning was found to play an instrumental role in task performance. In this regard, in our study, strategic planning groups including collaborative and teacher-led planning groups outperformed the no-planning groups in terms of CAF. We found that trade-off effects exist between complexity and accuracy, a conclusion drawn by Foster and Skehan (1996), and

Skehan and Foster (1997), and Mehnert (1998) who all reported that the improvement in accuracy is accomplished at the expense of complexity and vice versa. In what follows, we will discuss the results in terms of CAF one by one.

Complexity

The participants in the collaborative planning low and high complexity groups outperformed the other groups in the study in terms of complexity. That is, collaborative planning improved language learners' oral productions in terms of complexity. The language learners in the collaborative groups might have exploited more resources and thus generated more complex utterances. This finding is in contrast with those of Skehan and Foster (1999) and Geng and Ferguson (2013) who found that group-based planning did not yield significant results concerning complexity. One line of explanation for the inconsistency of our result with those of theirs might be that they utilized a decision-making task, which is a more complicated task than the narrative ones used in our study. However, Skehan and Foster (1999) argue that group-based planning moves towards an emphasis on content that might result in greater language complexity. Nonetheless, this finding agrees with those of some researchers (e.g., Ahangari & Abdi, 2011; Foster & Skehan, 1996; Gilabert, 2007; Mehnert, 1998; Mohazabieh et al., 2020) who have revealed that pre-task planning in the form of individual planning can increase the complexity level of learners' oral production. This is also in harmony with Yuan and Ellis's (2003) results suggesting that pre-task planning increased grammatical complexity.

Given the significant difference between the collaborative planning high complexity group (M= 17.51) and the collaborative planning low complexity group (M= 16.25), it is inferred that the participants performing more complex tasks outperformed those undertaking low complex tasks concerning complexity. Our result in this regard is in harmony with the findings of Mohazabieh et al. (2020) and Robinson's (2001c) suggesting that performing cognitively complex tasks triggers more complex language than simple ones. When given sufficient time to plan the task performance, the language learners might have gained a better knowledge of content to undertake the task and thereby generated more complex language.

Accuracy

The teacher-led planning groups (high and low) were better than the other planning groups in terms of accuracy. This finding fits finely with that of Skehan and Foster (1999) who reported teacher-fronted planning significantly impacted accuracy and resulted in more control over the language. Likewise, Geng and Ferguson (2013) found that teacher-organized planning groups generated more accurate utterances. However, this effect did not reach significance probably because of their small sample size and a different accuracy measure. One possible justification for this finding is that the teacher might have adopted a form-focused approach rather than a meaning-focused approach. Likewise, it is common practice for teachers in Iran to primarily rely on the explicit teaching of morphosyntax making language learners naturally attend to well-formed language. To wit, the teacher might have scaffolded the language learners to prepare grammatically correct forms for a task. Based on Skehan and Foster (1999), it is plausible to speculate that teacher-oriented planning might emphasize form and accuracy. Another explanation is that the teacher might have asked learners to monitor their language

productions while performing low and high complex tasks. However, as put forth by Wang (2014), speech monitoring can contribute to accuracy.

Taking task complexity into account, the teacher-led planning low complex group significantly outperformed the teacher-led planning high complex group in terms of accuracy. That is, the participants performing less complex tasks produced more accurate language. It implies that accuracy and complexity are in competition for attention in teacher-led planning. On this point, Geng and Ferguson (2013) maintain that learners undertaking low complex tasks did not reflect upon the connection between pictures leading to less complex language.

Fluency

The collaborative planning groups significantly performed better than the other groups concerning fluency. This disagrees with the findings of Skehan and Foster (1999) and Kang (2018) who revealed that group-based planning conditions did not show significant results in terms of fluency. Here again, the disparity in the results could be explained in terms of the task type i.e., decision-making tasks employed in their studies. Moreover, negotiating in groups might result in better interaction, but it prevents effective task planning. This result accords with that of Geng and Ferguson (2013) who explored a significant effect of the pair-based planning on fluency. In a similar vein, Nasiri and Atai (2017) revealed that under the strategic planning condition (individual planning) the language learners did not plan the task performance. Consequently, they commenced the task more fluently. In a collaborative planning group, as the name implies the interaction among language learners is collaborative and reasonably and primarily meaning-based with little focus on form. This type of interaction typical of group-based planning might account for greater fluency gains in the collaborative groups compared with the other groups. Another explanation might be that in group-based planning, the language learners might have had the opportunity for task rehearsal that contributed to greater fluency. This justification (if true) is in line with Bygate's (2001) study in which task repetition led to more fluent language. Another line of explanation for the significant effect of collaborative planning on fluency might be that the participants in the collaborative planning groups did not draw on their knowledge of forms. This would reduce short-term memory load and allowed their attentional resources to process meaning more easily and thereby enhanced their fluency.

Given the significantly higher mean score of the collaborative planning low complexity group than the collaborative planning high complexity group in terms of fluency, we can conclude that the collaborative planners undertaking low complex tasks delivered a more fluent performance. This can be explained by the fact that the collaborative planners undertaking low complex tasks were less cognitively involved than the collaborative planners performing high complex tasks. Similarly, Mohazabieh et al (2020) found that individual pre-task planners performing low complex tasks produced more fluent utterances than individual pre-task planners undertaking high complex tasks.

Conclusion

This research study examined the combined effects of strategic planning i.e., collaborative and teacher-led planning conditions and task complexity on language learners' oral productions in terms of CAF.

The results demonstrated that the language learners in the collaborative planning low and high complexity groups outperformed the other groups in terms of complexity. However, the collaborative planning low complexity group was better than the collaborative planning high complexity group in terms of fluency. Likewise, teacher-led planning high and low groups did better than the other groups concerning accuracy. Moreover, the teacher-led planning high complexity group outdid the teacher-led planning low complexity group regarding accuracy. Overall, in respect of CAF, the results confirm Skehan's (1998) Limited Capacity Hypothesis suggesting that high complex tasks did not enhance accuracy and complexity concurrently, which highlights the trade-off between accuracy and complexity. However, these results appear to run counter to Wendel's (1997) claim that the trade-off concerns fluency and accuracy.

In light of the results concerning collaborative planning, language teachers are recommended to practice collaborative planning through interactive tasks and to guide learners in the process of planning prior to task performance to offer ample learning opportunities. In this regard, language teachers can assess learners' speaking skills through dialogical tasks which might better depict their language abilities.

Given the superiority of teacher-led planning high and low complexity groups over the other groups in terms of accuracy, language teachers could encourage language learners in collaborative planning groups to provide their peers with scaffolded corrective feedback on forms to enhance their accuracy while performing a task.

Nevertheless, regarding task complexity, the results partially lend supports to Robinson's (2001a) Cognition Hypothesis in that increasing task complexity led to syntactically more complex language (Kuiken & Vedder, 2011) but not linguistically accurate language. The results in general appear to be more congruent with Skehan's (1998) Trade-Off Hypothesis than Robinson's (2001a) Cognition Hypothesis. Nonetheless, this study might unlock the potentials of the two hypotheses.

Our results depicted that performing more complex tasks would result in the improvement of the language learners' oral performance in terms of complexity. In this regard, EFL teachers are suggested to encourage language learners to commence authentic tasks to help enhance their language complexity. Language teachers thus need to take into consideration language learners' cognitive abilities as well as the cognitive load of tasks. Also, the findings regarding strategic planning conditions combined with task complexity revealed significant gains in either complexity or accuracy underpinning Skehan's (1998) Limited Capacity Hypothesis.

This study sheds more light on task design and performance in classroom settings. In light of the competing goals of accuracy and complexity as shown in the present study, language teachers should maintain a balance between CAF measures. In this regard, materials developers need to design tasks with different planning conditions and task complexity levels.

Owing to time constraints for planning conditions in real-life situations, EFL teachers need to draw on situational authenticity, a situation in which learners engage in real-life tasks by employing communication strategies typical of authentic situations.

The results of this study carry clear pedagogical implications for language teachers, task designers, materials developers, and speaking and writing examiners. The findings can assist

researchers in creating a speaking evaluation scheme for high-stakes examinations. Besides, the findings may help task designers attend to the cognitive complexity of tasks when designing them. Drawing on the results, syllabus designers might incorporate a diverse array of tasks to trigger more complex, accurate, and fluent performance. They can also differently manipulate speaking tasks through planning conditions and task complexity to achieve different pedagogical goals. A possible implication for language teachers is that they can employ an eclectic method of teaching in which CAF elements are conflated and potentialities of task-based language instruction are captured.

Notwithstanding the possible contribution of the present study to researchers in the field of language teaching, some caveats in the study may limit the results. The first caveat concerns the sample size of the study. Conducting the study with more participants could make the findings more reliable and generalizable. A second caveat pertains to the number of treatment sessions. With more treatment sessions, the study might yield different results. Further, the delayed effects of strategic planning types and task complexity task planning were not explored in this study. As a result, further studies are called for to administer a delayed posttest to determine the long-term impact of the variables in question. Finally, although the learners' language proficiency level was controlled, their L1 speaking ability and working memory capacity were not taken into consideration. Future researchers might perform studies in which these variables can be controlled.

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.

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.

Bygate, M. (2016). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*. Philadelphia: John Benjamins.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Hillsdale, NJ: Lawrence Erlbaum Associates.

East, M. (2014). Encouraging innovation in a modern foreign language initial teacher education program: What do beginning teachers make of task-based language teaching? *The Language Learning Journal, 42*(3), 261–274.

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 lan*guage (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 fluency, complexity, and accuracy in L2 oral production. *Applied linguistics*, *30*(4), 474-509.

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

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.

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3, 215-247.*

Geng, X., & Ferguson, G. (2013). Strategic planning in task-based language teaching: The effects of participatory structure and task type. *System*, *41*(4), 982-993.

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.

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

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.

Kang, S. (2018). The effects of individual vs. collaborative pre-task planning on Korean middle school learners' English oral task performance under different task complexity. *Korean Journal of English Language and Linguistics, 18*(3), 306-327.

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.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P.Robinson (Eds.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 39-61). Amsterdam: John Benjamins.

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.

Lennon, P. (2000). The Lexical Element in Spoken Second Language Fluency. In H. Riggenbach (Eds.), *Perspectives on Fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Lee, J. (2018). Task complexity, cognitive load, and l1 speech. *Applied Linguistics*, *40*(3), 506-539.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.

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., Alibabaee, 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.

Mohazabieh, S., Sahragard, R., Rassaei, E., Zamanian., M (2020). The potential combined effects of task complexity and planning types on Iranian EFL learners' oral production performance. *Applied Research on English Language*, *9*(4), 503-538.

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.

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. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, *21*(2), 45- 105.

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.

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.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics 17*(1), 38–62.

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. (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. (2014). The context for researching a processing perspective on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 1–26). 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

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 239-277). Amsterdam: John Benjamins.

Thompson, C. (2014). *Guided planning, task complexity and second language oral development* (Doctoral dissertation)), University of Central Lancashire).

Wang, Z., (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In Skehan, P. (Ed.), *Processing perspectives on task performance* (pp. 27-62). John Benjamins Publishing Company.

Wendel, J. (1997). *Planning and second language narrative production*. Unpublished Ph.D. Thesis, Temple University, Japan

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.

Yuan, F., & Ellis, R. (2005). The effect of careful within-task planning on oral production and written task performance. In Ellis, R. (Ed.), *Language learning & languag*e *teaching: planning and task performance in second language* (pp. 167-192). Amsterdam: John Benjamins Publishing Company.