Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(2), Spring 2022, pp. 75-111
DOI: 10.22099/TESL.2021.41004.3017

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

The Comparative Effect of Rehearsal, Strategic, and Online Planning on the Fluency, Accuracy, Complexity, Lexical Variety, and Attentional Allocation of EFL Learners' Argumentative Writing

Hoda Divsar *
Payame Noor University

© Shiraz University

Abstract

This study examined the effect of rehearsal, strategic, and online planning on the intermediate undergraduate EFL learners' writing complexity, accuracy, fluency (CAF), lexical variety, and the cognitive psychological performance of attentional allocation to linguistic aspects during planning. To this end, 80 intermediate university students were randomly divided into three experimental groups and one control group. The participants' performance was compared based on measures of Wigglesworth and Storch's (2009) fluency, Storch and Wigglesworth's (2007) accuracy, Foster and Skehan's (1999) complexity, McKee, Malvern, and Richards' (2000) vocd-D model of lexical diversity. The study focused on rehearsal, strategic, and online planning in argumentative writing tasks. Retrospective interviews were conducted right after task performance to examine the participants' attentional allocation under three planning-time conditions. It was revealed that rehearsal task planning resulted in higher CAF and lexical variety in comparison to the other groups. Noticeable variation was also evidenced in the participants' attentional allocation in terms of CAF across different time planning conditions. It was also shown that the learners paid more attention to the conceptualization of ideas during pre-task planning through sentences or clauses while performing a task. Little attention was paid to formulizing complex or accurate

Received: 13/07/2021 Accepted: 22/09/2021

* Assistant Professor, Email: hodadivsar@gmail.com

utterances. The findings have numerous pedagogical implications for EFL teachers, university instructors, and EFL students.

Keywords: Accuracy, Attentional Allocation, Complexity, Fluency, Lexical variety, Online planning, Rehearsal planning, Strategic planning, Writing skill

© Shiraz University

Examining differential influences of task conditions and planning time on EFL learners' complexity, accuracy, and fluency (CAF) is among the sources of inspiration for most researchers (Ahmadian, 2012; Ellis, 2009; Fazilatfar, Kasiri, & Nowbakht, 2020). As Ellis (2009) defined, online planning is the process through which learners pay attention to form while they monitor their production when carrying out a task. Based on the performance planning time, whether a priori to or during a given task, Ellis (2005) distinguished two planning types, namely pre-task planning and within-task (online planning), each of which is divided into two other subtypes (see Figure 1). As Schmidt (2001) puts it, pre-task planning takes in 'preparatory attention' that contributes to greater accuracy and speed. The pre-task is categorized into rehearsal and strategic planning, whereas the within-task planning is of pressured and unpressured types. In rehearsal planning, the learners accomplish a task earlier than the repeated performance to be prepared for the subsequent performance. In other words, the first performance of the task is a preparatory performance for the main one. In strategic planning, the learners contemplate what they are going to write in terms of content as well as the language before engaging in writing the task. It involves practitioners paying attention to what they are going to produce in terms of content and language before undertaking the task. In pressured planning, the learners are expected to accomplish the task quickly under time limitations, while in the unpressured one, the learners enjoy sufficient time to complete the task leisurely. Setting limited time to perform a task typically denotes refuting the opportunity for online planning.

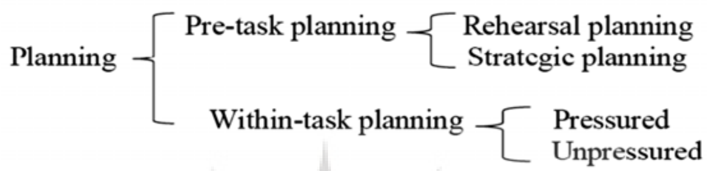

Figure 1
Task-based Planning Types (Ellis, 2005, p. 4)

The rationale of planning lies in Skehan's (1998) limited attentional capacity that planning reduces the load of processes and consequently allows the learners to get prepared for conceptualizing and formulating what they want to produce. This preparation leads learners to be more accurate, complex, and fluent. Based on the limited attentional capacity model, Skehan (1998, 2014) presented the Trade-off Hypothesis, in which learners cannot deal with all aspects of the L2 language, particularly in more cognitively challenging tasks. For Skehan (1998), the most robust competition resides in the trade-off between accuracy and complexity. It is claimed that due to learners' limited memory capacity, performing a task imposes a burden on the learners' attention leading them to prioritize one facet of language over the other aspects to deal with the difficulties of a task that leads to the imbalance allocation of attention to the language features, namely accuracy, fluency, or complexity.

As for lexical variety, a bulk of studies investigated lexis as a sub-dimension of complexity; however, Skehan (2009) contended that a nuance contrast exists between lexical elegance and structural complexity. Based on this duality, the study distinguished lexis from complexity, considering it an

© Shiraz University

independent writing dimension. As such, it is difficult to predict its competition with other aspects of language performance. Thus, after the Trade-off Hypothesis, this study explored lexical variety as a competing aspect of writing performance in addition to CAF.

Although abundant studies have considered the impacts of task features as well as time conditions concerning the CAF triad (Abdi Tabari, 2017), planning time is still a controversial issue that is not settled yet and is open to a variety of interpretations. Whereas some scholars underscored the significance of free-writing in minimizing the cognitive involvement of coherence, detection of notions, and translation (Johnson, 2014), other scholars confirmed the advantage of planning in reducing the processing load of conceptualization, thus, leading to a higher quality of the manuscripts (Ellis & Yuan 2004). The cornerstone of the debate lies in Kellogg's (1990) two contrasting suppositions, namely, Interaction and Overload Hypotheses concerning the role of planning conditions in the text quality. Broadly speaking, the overload hypothesis asserts that planning conditions improve the quality of the text through preparing a macrostructure that can free the writers from the restriction of attentional capacity to handle the problems that arise from the translation processes and consequently reduce the risk of the trade-off effect. In contrast, the Interaction Hypothesis focuses on the free-writing strategy, reviewing processes, and the dynamic nature of planning that leads to the emergence of fresh ideas.

Despite a plethora of research carried out on the impacts of pre-task planning on L2 production (Rostamian, Fazilatfar, & Jabbari, 2018; Fazilatfar et al., 2020), fairly inadequate attention has been paid to the influence of online planning on writing concerning CAF and the allocated attention to each of these features. Likewise, researchers have not resolved the conflict between

© Shiraz University

© Shiraz University

the presumptions that the effect of pre-task planning would not be detected throughout the writing and that the planning time length might not decrease the query of online planning (Ong, 2014, Ong & Zhang, 2010). The dispute merits a controversial perspective of whether learners go through planning through the transcription stage or move from planning to translation of concepts and organizations (Ong, 2014; Ong & Zhang, 2010). Leaning on the aforementioned inconsistent arguments about the impact of planning time and task condition (Johnson, Mercado, & Aceve, 2012), it is still in the researchers' interest to discover the effect of length of time and the strategic planning and text organization concerning fluency, lexical diversity, and text quality. A profound understanding of the various types of planning and their effects stimulate both SLA researchers, who are mainly concerned with examining the L2 acquisition theories and teachers dealing with supporting language learners to learn languages more efficiently and effectively. The present study tried to provide evidence on the differential effects of planning time conditions on the intermediate EFL learners' textual performance as measured in terms of CAF and lexical variety.

Literature Review

Planning refers to allocating the amount of preparation time during which learners resolve which linguistic features and devices to tackle to get the intended meaning across (Ellis, 2005). Inspired by continuing controversies, the issue of planning conditions has been an impetus for a large number of studies (Ahmadian & Tavakoli, 2011; Fazilatfar et al., 2020; Kargozari, Soleimani, Jafarigohar, & Hemmati, 2016) to investigate L2 learners' attentional focus and task performance. Based on Skehan's (2009) limited attention capacity, learners' attentional resources are presumed to be selective

© Shiraz University

and limited that implies they have to decide on the type of cognitive processes to concentrate on through the writing process that may contribute to trade-off impacts among the metacognitive processes, and sequentially different text quality. In other words, when there are two or more latent foci of attention, the task performer cannot dedicate equal attentional resources to all foci simultaneously, and he might prioritize one of these focal points at the cost of others (Schmidt, 2001). Limited attention capacity gave rise to the trade-off hypothesis that stimulated researchers concerning the circumstances under which a given task is best completed (Skehan, 1998, 2009, 2014). It is assumed that when a task performer is performing a task, he cannot instantaneously allot equal attentional resources to all performance features, which might lead to subordinate performance in others unless he is supported through handling task conditions (Skehan, 2009). This is where planning as a multipurpose pedagogical medium makes a move. Encouraging learners to plan their tasks could be regarded as offering task performers enough time to go through mental processes on their productions before or while doing the task (Ahmadian, 2012).

Types of Planning and Complexity, Accuracy, and Fluency in Writing

Ellis' (2005) typology of planning is among the widely preferred frame of reference for operationalizing the concept of planning. One of the earliest studies was carried out by Ellis (1987), who worked on the influence of task planning on the accuracy of ESL learners' writing and speaking and uncovered a positive relationship between planning and grammatical accuracy. Haghverdi, Khalaji, and Biria (2020) revealed the supremacy of the strategic group in producing more fluent narrative writings, confirming Skehan's model that planning decreases cognitive loads and yields higher

© Shiraz University

quality. Soleimani and Kargozari (2014) found that the online group displayed higher accuracy in narrative writing than the pre-task group. Rahmanian (2004) inspected the association between pre-task and online planning and CAF and unveiled the superiority of pre-task planners concerning fluency. However, the difference in the accuracy and complexity of the online planning group did not attain a significant level. Concerning the effect of task types, he unveiled that descriptive tasks were more straightforward than narrative ones and liberated more attentional resources toward complexity and accuracy. Concerning complexity, however, the most extended amount of planning time was an influential factor leading to significantly higher complexity. Nevertheless, no beneficial effects were identified by Khomeijani Farahani and Faryabi (2016), who discovered the influences of pre-task planning and no planning on the EFL argumentative writing in terms of CAF.

Taking both writing processes and product into account with respect to the CAF triad, Rostamian et al. (2018) studied the influence of cognitive processes and planning time on the quality of L2 narrative writing. Concerning the involved cognitive processes, the findings unveiled the supremacy of online planning in integrating translation and evaluation; however, the pre-task planning condition reduced the rate of writing processes. Regarding the CAF in the quality of the narrative writings, it was revealed that the establishment of pre-task and online planning did not lead to simultaneous improvement of writing CAF, a finding which is not in line with those mentioned in the aforementioned sections. Consequently, it can be inferred that the Overload Hypothesis and the Limited Attentional Capacity Model were backed by the findings. A plethora of studies on strategic planning has verified the positive influence of strategic planning on fluency, as measured in miscellaneous ways such as temporal phenomena as well as repair. For instance, Ellis and Yuan (2004) and Marzban and Norouzi (2010) found that the amount and type of strategic planning significantly influence learners' performance, particularly fluency and complexity, but to a minor degree accuracy. The findings related to the effect of strategic planning documented generally the positive influence on complexity (Foster & Skehan 1996; Ong, 2014; Ong & Zhang, 2013), and the result would seem to be more noticeable in grammatical complexity than in lexical density (Ellis, 2005; Ong, 2014).

Reviewing the literature, we can see that the studies on planning are not consistent concerning L2 writing in particular, and language performance in general (Ellis & Yuan, 2004; Khomeijani Farahani & Meraji, 2011). Contrary to the studies that have documented a positive effect for planning time on the divergent aspects of language output, especially regarding accuracy, complexity, and fluency (Fazilatfar et al., 2020), Khomeijani Farahani and Faryabi (2016) and Johnson et al. (2012) indicated that pre-task planning did not significantly affect L2 writing performance. Neither Johnson et al. (2012) nor Rahimpour and Safarie (2011) spotted any differences concerning the influence of the various types of planning. More interestingly, Ong and Zhang (2010) acknowledged that pre-task planning negatively affected the fluency and lexical complexity of L2 writing. Therefore, as Johnson (2014) claimed, planning, as one of the task implementation issues, deserves further concern and entails further requirements in the writing domain.

Although a growing number of researches exist on pre-task planning, few studies have focused on online planning in the light of CAF, lexical variety, and allocation of attention (Tavakoli & Rezazadeh, 2014; Khomeijani Farahani & Faryabi, 2016). Hence, conducting further empirical research can illuminate the issue at hand. Given the preceding discussion, the present study

© Shiraz University

explored the role of both pre-task and within-task planning in EFL undergraduate students' argumentative writing for CAF and lexical variety.

The present study addressed these research questions:

- Is there any significant difference among the fluency, accuracy, complexity, and lexical variety of undergraduate EFL learners' argumentative writing task performance under three types of task planning (rehearsal, strategic, and online) conditions?
- How does EFL learners' allocation of attention change in terms of CAF and lexical variety across different time planning conditions?

Method

Participants

Eighty undergraduate EFL learners (males = 31 males and females = 49) majoring in English at BA level with the age range between 21 to 23 years were selected through non-random convenience sampling from universities in Guilan and were randomly assigned into three experimental and one control group. They had already passed prewriting courses, namely, Grammar and Writing 1 and 2, Advanced Writing, and Essay Writing. To homogenize the participants with respect to their proficiency, Oxford Quick Placement Test (Version 1) was given. Based on its manual, those test-takers whose score fell between 40-47 were considered upper-intermediate. The participants' writing proficiency was also measured through two argumentative writing tasks at the pre-test stage to determine their writing ability before the treatment sessions.

Instruments and Materials

The following instruments were employed to gather the necessary data. They are explained in detail as follows.

© Shiraz University

Oxford Quick Placement Test (OQPT). It was given to guarantee the homogeneity of the participating learners concerning their English language ability. This placement test contains 60 multiple-choice questions that assess the participants' English knowledge in terms of usage, prepositions, and vocabulary through cloze passages and fill-in-the-blank items. The test was administered in 30 minutes.

Writing Pre-test. To test the participants' writing ability before the treatment, a topic from IELTS Writing Task 2 (https://www.english-exam.org/IELTS/ielts_writing) was selected to specify the level of the participants' writing ability as well as their abilities in terms of the CAF triad. They were required to write a 120 to 180-word text on the given topic within 30 minutes. Two raters scored the papers to ensure the reliability of scoring. To measure the participants' overall writing quality, Jacobs, Zinkgraf, Wormouth, Hartfiel, and Hughey's (1981) rating scale, including the quality of writing, organization, appropriate use of grammar, vocabulary use, and accuracy of the content, was used. Inter-rater reliability results showed high agreement (***r*** =.87), suggesting that the differences among the four groups were not significant. The rating scales explained below were utilized to assess the writing competencies for the CAF. The final mean scores on each sub-component of each of them were considered to confirm the comparability of the groups concerning CAF before the experiment. To ensure the inter-rater reliability of the obtained scores, the tasks were scored independently by two raters.

Writing Post-test. The purpose of the post-test was to check the post-treatment differences. The topic of the writing post-test was the same for all the participants: "Nowadays, many people choose ready-made food and refuse to cook at home. What are the advantages and disadvantages of such as

© Shiraz University

choice?” The Pearson correlation values obtained for the post-test ($r_{\text{post-test}}$ = .898) showed significant correlations between the two ratings of the writing test ($p < .01$).

Argumentative Writing Tasks. The argumentative writing tasks were chosen to be of the same level as the participants. For each writing task, they were carefully instructed on what to write and how to organize their writing. They were asked to consider a decision that young individuals are asked to make between two potential choices and discuss the opposing sides of that topic.

The topics were selected from the IELTS task 2 website (www.ielts-exam.net). There were eight sessions, and in each session, one topic was given to the participants. The coursebook used for the present study was "15 Days Practice for IELTS Writing," which contains several tasks, practices, sample writings, and useful instructions on writing academic texts.

Rating Scales. The following measures were considered in assessing the four dimensions of English writing output:

a) ***Fluency**:* In this research, following Wigglesworth and Storch (2009), fluency is seen as the average word count, clauses, and T-units in the script.
b) ***Complexity:*** According to Norris and Ortega (2009), subordination is indicative of syntactic complexity at intermediate and upper-intermediate levels. Subordination measures are considered the most potent index of syntactic complexity. In the current study, therefore, following Foster and Skehan (1996), complexity indicated the rate of clauses to T-units and the proportion of dependent clauses to whole clauses (DC/C).

© Shiraz University

c) ***Accuracy:*** Storch and Wigglesworth (2007) see accuracy as the ratio of error-free T-units to all T-units (EFT/T) and the number of error-free clauses to all clauses (EFC/C).
Consequently, in this study, all errors including syntactic errors such as errors in word order and missing items, morphological errors, namely subject-verb agreement, verb tense, incorrect use of prepositions and articles, and improper use of word form, and lexical errors were carefully examined. Errors in the present study included syntactic errors (e.g., errors in word order and missing elements) and morphology (e.g.,). Errors in lexis (word choice) were counted only when the lexical error impeded or obscured the meaning; therefore, the punctuation and spelling errors were ignored.

d) ***Lexical Variety:*** The most frequently used measure of lexical variety is the type-token ratio (TTR) which counts the number of different tokens for each type (Malvern & Richards, 2000). Tokens describe the overall sum of words in a given text, while types represent the number of different items (Nation & Webb, 2011). However, the inherited problem in this method is the dependence of lexical richness measure on the text length (Nation & Webb, 2011). To recompense the flaw of the TTR, various mathematical transformations were tried, among which Malvern and Richard's (2000) D model, based on a curve-fitting, gained more recognition. This study measured lexical diversity using McKee et al.'s (2000) vocd-D model of lexical diversity specified by adjusting D to the equation converges on the TTR value. To calculate the pertinent data, special software called VOCD was utilized (available at http://www. www.textinspector.com). According to McCarthy and Jarvis (2010), vocd-D is based on the likelihood of

© Shiraz University

drawing a particular number of tokens of a certain type from a particular sample without replacement. Each text had more than 150 words, thus meeting the minimum sample size required to compute the valid D score. Each writing was exposed to 15 times of Voc-D analyses; the whole procedure was repeated four times and finally, their average was used as the final D.

Procedure

At the outset, the Oxford Placement Test and the writing pre-test were administered to confirm the homogeneity of the participants' L2 writing as well as their proficiency level. Their writing tasks were then analyzed regarding CAF and lexical variety. The purpose of the pre-test of writing was to ensure the homogeneity of the participants with respect to their writing ability prior to the treatment. Inter-rater reliability was confirmed with a randomly selected sample of 20 writings (about 20% of the whole dataset).

The experimental groups were instructed through three different types of task planning (rehearsal, strategic, and online), and conventional instruction was used for the control group. The first experimental group received the writing instruction and practices in the format of rehearsal planning. In this group, the learners had extra time and sessions before the main writing exercise of argumentative writing. This time is meant to be for practice. In other words, they prepared the argumentative writing task within *rehearsal planning*. The students received 15 minutes for the preparation of the writing task in two sessions within a week, and in the next three sessions, they were required to prepare the same task within 30 minutes. The participants prepared the writing task based on strategic planning in the second experimental group. That is to say, they were required to prepare their argumentative writing task

© Shiraz University

© Shiraz University

with the specific strategic plan that was meant for them. They had 10 minutes to reflect on the writing subject, and then they had 15 minutes to plan for their writing on a piece of paper, and finally, they had 30 minutes to prepare their writing completely. In the third experimental group, the participants had to apply *online planning* for their writing preparation. The online planning group had thirty minutes to compose their ideas, but they were told how to carefully plan on the CAF and lexical variety of their language.

Immediately after the participants completed task performance, retrospective interviews were conducted and recorded for later transcription. Each interview lasted for 25 to 30 minutes in which the participants were asked about what they focused on during three time-planning conditions and task performance by providing them with an analytic table on the sub-components of CAF and lexical variety based on the rating scales mentioned above. Quantitative content analysis was employed to analyze the transcribed interviews. The retrospective interviews were coded in terms of whether they made any remarks on CAF and lexical variety in each type of planning time, and consequently, categories emerged from data through quantitative content analysis. Two experienced EFL instructors coded 25% of the interview data independently to check the reliability of the analysis. The analysis focused on the existence as well as the frequency of the remarks in each interview. There was 89% of agreement in recognizing the pertinent remarks on CAF and lexical variety. The results were statistically evaluated to estimate the learners' prioritized attention by the frequency of use of decontextualized features of CAF and lexical variety. L2 argumentative writings at two data collection times were quantitatively assessed for the writing features. To determine the writing accuracy, the proportion of the error-free T-units to the entire T-units was specified. To determine the writing complexity, the total number of clauses divided by the whole number of T-units was calculated. To ensure the reliability of post-test scoring, two raters scored 20% of the whole samples according to all the aforementioned scales of measurement.

Results

The correlations between the scores assigned by the two raters were examined through computing inter-rater reliability coefficients. One-way MANOVA was employed to compare the writing score means. Before running the parametric tests, the main assumptions were examined. To check the assumptions, Box's *M*, homogeneity of variances tests, skewness analyses, and the Shapiro-Wilk test were used.

Inter-Rater Reliability Coefficients. Average Measures were computed individually for the argumentative writing pre-test and post-test. The estimated inter-rater reliability values for the pre-test scores were ($r_{\text{T-units, number of words, and causes}}$= .89), ($r_{\text{error-free T-units}}$ = .82), with 95% CI (.72, .88), ($r_{\text{total T-units}}$ = .95), with 95% CI (.92, .96), (r_{clauses} = .97), with 95% CI (.96, .98), (*D* = 1.00), with 95% CI (1.00, 1.00), and ($P \leq .05$). The estimated inter-rater reliability values between the two raters for the post-test scores were ($r_{\text{T-units, number of words, and causes}}$ = .89), ($r_{\text{error-free T-units}}$ = .99), with 95% CI (.99, .99), ($r_{\text{total T-units}}$ = .99), with 95% CI (.99, .99), (r_{clauses} = .99), with 95% CI (.99, .99), (*D* = 1.00), with 95% CI (1.00, 1.00) ($P \leq .05$). Therefore, inter-rater reliability for the argumentative pre-test and post-test were confirmed.

Descriptive Statistics

The initial analysis included descriptive statistics counting the frequency of writing features such as accuracy, complexity, and D for diversity features in the four groups. Furthermore, given the nature of the research questions and

© Shiraz University

the 1-by-3 research design (types of tasks and three features of writing), a one-way between-groups MANOVA was run to compare the overall effect of types of task planning on the average writing performance.

In the first phase, the three measures, namely, accuracy, complexity, and lexical variety were used to investigate the quality of the learners' argumentative writing pre-test. The mean scores of the *fluency* for the experimental and the control groups were ($M_{Ex.1}$ = .2099), ($M_{Ex.2}$ = 2084), ($M_{Ex.3}$ = .2077), ($M_{Co.}$ = .2069), respectively and the standard deviation for the groups were ($SD_{Ex.1}$ = .0172), ($SD_{Ex.2}$ = .0158), ($SD_{Ex.3}$ = .0166), ($SD_{Co.}$ = .014), respectively. Similarly, for *accuracy*, the mean scores were ($M_{Ex.1}$ = .2094), ($M_{Ex.2}$ = 2083), ($M_{Ex.3}$ = .2055), ($M_{Co.}$ = .2072), respectively. Furthermore, the standard deviations for the groups were ($SD_{Ex.1}$ = .0172), ($SD_{Ex.2}$ = .0158), ($SD_{Ex.3}$ = .0166), ($SD_{Co.}$ = .014), respectively. Likewise, the *complexity* mean scores were ($M_{Ex.1}$ = .616) , ($M_{Ex.2}$ = .618), ($M_{Ex.3}$ = .608), ($M_{Co.}$ = .606), respectively and the degree of the variation of scores were ($SD_{Ex.1}$ = .055), ($SD_{Ex.2}$ = .054), ($SD_{Ex.3}$ = .041), ($SD_{Co.}$ = .042), respectively. Besides, the mean scores of the *lexical variety* for the L2 argumentative writing pre-test were ($M_{Ex.1}$ = .774), ($M_{Ex.2}$ = .762), ($M_{Ex.3}$ = .752), ($M_{Co.}$ = .746), respectively. Moreover, the degree of the variation of scores were ($SD_{Ex.1}$ = .040), ($SD_{Ex.2}$ = .032), ($SD_{Ex.3}$ = .044), ($SD_{Co.}$ = .044), respectively.

To examine the significance of the mean differences among the four groups in terms of writing CAF and lexical variety, before the intervention, on three types of task planning, a one-way ANOVA was run. First, the homogeneity of the variances was checked via computing Levene's test. Levene's statistics showed that the group variances were similar in pre-test scores ($F_{accuracy}$ 3, 76 = .282; $P_{accuracy}$ (.839) ≥.05), ($F_{complexity}$ 3, 76= .876; $P_{complexity}$ (.458) ≥ .05) and ($F_{lexical\ variety}$ 3, 76= 1.136; $P_{variety}$ (.340) ≥ .05).

© Shiraz University

© Shiraz University

Levene's statistics supported the hypothesis that the group variances were the same.

As for the normality assumption test, the skewness and kurtosis analyses were calculated. It was revealed that the skewness as well as kurtosis values were all within the range of ± 2, supporting that the distributions were normal (George & Mallery, 2010). These values were (skewness $_{\text{accuracy}}$ = .034, kurtosis $_{\text{accuracy}}$ = -.567; skewness $_{\text{complexity}}$ = .469, kurtosis $_{\text{complexity}}$ = 1.529; skewness $_{\text{lexical variety}}$ = .137, kurtosis $_{\text{lexical variety}}$ = -.843). The significance values of the F test were greater than (.05) for the writing pre-tests. Thus, the average assessment scores for the L2 argumentative writing tests were equal across the four groups at the beginning of the study ($F_{\text{accuracy}}(3, 76) = .667$, $p = .575 > .05$), ($F_{\text{complexity}}(3, 76) = 2.620$, $p = .057 > .05$), and ($F_{\text{lexical variety}}(3, 76) = 1.856$, $p = .144 > .05$) and none of the differences reached the statistical significance.

To provide the answer to the research questions, a one-way MANOVA was run on the post-test the results, which are presented in two sections, namely, descriptive and inferential statistical analyses. The descriptive statistics for the posttest writing scores revealed that the mean scores of the *fluency*, as measured by Wigglesworth and Storch's (2009) definition, were ($M_{Ex.1}$ = .3820), ($M_{Ex.2}$ = 3345), ($M_{Ex.3}$ = .2871), ($M_{Co.}$ = .2198), respectively. The standard deviation was ($SD_{Ex.1}$ = .0338), ($SD_{Ex.2}$ = .0306), ($SD_{Ex.3}$ = .0297), ($SD_{Co.}$ = .0302), respectively. Therefore, the rehearsal group reported a higher number of utilized words, the average number of T-units, and the number of clauses indicating higher fluency scores compared to the other three groups. In contrast, the control group had the lowest mean for the aforementioned factors.

© Shiraz University

Following Wigglesworth and Storch (2007) for the measure of *accuracy*, the mean scores of the *accuracy* were ($M_{Ex.1}$ = .3720), ($M_{Ex.2}$ = 3245), ($M_{Ex.3}$ = .2771), ($M_{Co.}$ = .2098), respectively. Furthermore, the standard deviations were ($SD_{Ex.\,1}$ = .0238), ($SD_{Ex.2}$ = .0206), ($SD_{Ex.3}$ = .0207), ($SD_{Co.}$ = .0202), respectively. The rehearsal group, therefore, reported a higher number of error-free T-units. In contrast, the control group had the lowest mean for the number of error-free T-units. Nevertheless, the total number of T-units used by the online task planning group was higher than the other experimental groups. Overall, the first experimental group who worked on rehearsal task planning reported higher accuracy scores than the other three groups.

Next, following Foster and Skehan (1996), the *complexity* of the L2 writings was estimated by obtaining the proportion of clauses to T-units for the post-test, too. The mean scores of the *complexity* were ($M_{Ex.1}$ = .918), ($M_{Ex.2}$ = .845), ($M_{Ex.3}$ = .699), ($M_{Co.}$ = .621), respectively. The degree of the variation of scores were ($SD_{Ex.1}$ = .020), ($SD_{Ex.2}$ = .034), ($SD_{Ex.3}$ = .022), ($SD_{Co.}$ = .024), respectively. The results revealed that although the total number of clauses used by the strategic group was higher than the other groups, there were more clauses per T-unit in paragraphs of the rehearsal group than those of the other three groups. Therefore, the rehearsal group scored higher complexity scores than the other groups.

For the *lexical variety*, following McKee et al. (2000) vocd-D model, the D value was computed by adjusting D to the equation converges on the TTR value. The mean scores of the *lexical variety* were ($M_{Ex.1}$ = .888), ($M_{Ex.2}$ = .862), ($M_{Ex.3}$ = .831), ($M_{Co.}$ = .770), respectively. Moreover, the degree of the variation of scores were ($SD_{Ex.1}$ = .038), ($SD_{Ex.2}$ = .027), ($SD_{Ex.3}$ = .028), ($SD_{Co.}$ = .047), respectively. Therefore, the experimental group who worked on rehearsal task planning used a higher number of content and orthographic words and thus had a higher lexical variety score than the other three groups.

Inferential Statistics

To provide the answer to the research questions, a one-way MANOVA was run on the post-test results after ensuring the assumptions were met. A Shapiro-Wilk test was run to assess the normality of distribution of accuracy, complexity, and lexical variety for the post-test scores (See Table 1).

© Shiraz University

Table1

Test of Normality for the Post-test Scores

	Groups	Shapiro-Wilk		
		Statistic	df	Sig.
Fluency posttest	Rehearsal	.987	20	.701
	Strategic	.973	20	.299
	Online	.899	20	.029
	Control	.957	20	.471
Accuracy posttest	Rehearsal	.967	20	.691
	Strategic	.943	20	.271
	Online	.871	20	.012
	Control	.955	20	.451
Complexity Posttest	Rehearsal	.924	20	.120
	Strategic	.943	20	.272
	Online	.947	20	.330
	Control	.970	20	.749
Lexical variety posttest	Rehearsal	.954	20	.428
	Strategic	.971	20	.783
	Online	.136	20	.200
	Control	.188	20	.063

Table 1 reveals that the p-values for the post-test scores were higher than (.01) for the Shapiro-Wilks test indicating that the writing scores were normally distributed for the post-test, too. The homogeneity assumption of covariance matrices for the post-test writing scores was checked through computing Box's M. The results are given in Table 2.

Table 2
Box's Test (Equality of Covariance Matrices)

Box's M	20.333
F	1.048
df1	18
df2	20410.896
Sig.	.400

Box's M significance value was .400 with F = 1.048 for the post-test scores that is higher than .05, confirming the homogeneity of covariance matrices. Thus, a one-way between-groups MANOVA test was run on the post-test results to model the values of argumentative writing tests based on their relationships to the categorical predictor, namely, types of task planning. The multivariate tests displayed four tests of significance for model effect (See Table 3).

Table 3
Multivariate Tests for the Post-test Scores

Effect		Value	F	H. df	Error df	Sig.	Partial Eta Squared
Group (types of	Pillai's Trace	1.247	18.02	9.00	228.00	.00	.416

© Shiraz University

Effect		Value	F	H. df	Error df	Sig.	Partial Eta Squared
task planning)	Wilks' Lambda	.028	66.85	9.00	180.24	.00	.696
	Hotelling's Trace	24.88	200.94	9.00	218.00	.00	.892
	Roy's Largest Root	24.49	620.58	3.00	76.00	.00	.961

© Shiraz University

The increased and positive value of Pillai's trace statistic for the "types of task planning" (Pillai's trace = 1.247) indicated that this effect contributed to the model. The significance value of the main effect, *types of task planning*, was less than (.05), indicating that the effects contributed to the model. Using general rules of thumb given by Cohen, Miles, and Shevlin (2001), the Partial eta squared for *types of task planning* was large ($\eta 2$ partial $_{\text{Pillai's}}$ = .416). The multivariate test documented the significant effect of task planning (rehearsal, strategic, and online) on EFL learners' L2 argumentative writing. Significant differences were found in the dependent variables (p < .05). One-way MANOVA was run to determine whether the task planning type affected the study's dependent variables (accuracy, complexity, and lexical variety). The test results revealed that types of task planning had a statistically significant effect on the three writing features (Pillai's trace = 1.247, Wilks' Lambda = 0.28, $F_{(9, 180.247)} = 66.859$, $p < 0.01$, $\eta 2$ partial $_{\text{Wilk}}$ = 0.696.

Multiple comparisons were made after documenting the significance of the differences among the groups. The Scheffe test made multiple comparisons among the groups concerning their post-test performance. The results of the Scheffe test revealed that in the argumentative post-test writing, for the fluency measure, the highest mean difference was between the first experimental having rehearsal task planning and the control group (mean difference= .170). In comparison, the lowest mean differences were reported between the rehearsal group and the strategic group (mean difference = .045) as well as the online group and the strategic group (mean difference = .046). In addition, the differences among the three experimental groups working on three forms of task planning were recorded to be statistically significant ($p \leq .05$).

The results also revealed that in the L2 argumentative post-test writing, for the accuracy measure, the largest difference was between the first experimental group having rehearsal task planning and the control group (mean difference= .162). In comparison, the lowest mean differences were reported between the rehearsal group and the strategic group (mean difference = .047) as well as the online group and the strategic group (mean difference = .047). In addition, the differences among the three experimental groups were statistically significant ($p \leq .05$).

It was also found that for the complexity measure, the largest difference was between the first experimental group having rehearsal task planning and the control group (mean difference = .223). In contrast, the lowest mean differences were reported between the rehearsal and the strategic groups (mean difference = .074). Similar to the accuracy scores, significant differences were found among the three experimental groups in terms of their complexity in argumentative writing ($p \leq .05$).

The results also revealed that the highest mean difference was reported between the first experimental group that worked on rehearsal task planning and the control group (mean difference= .118). In contrast, the lowest mean

© Shiraz University

© Shiraz University

differences were reported between the rehearsal and the strategic groups (mean difference= .026). The difference between the rehearsal and strategic groups was not significant concerning the lexical variety scores. Nevertheless, significant differences were reported between each of the three experimental groups in terms of their lexical variety in argumentative writing and the control group ($p \leq .05$).

In general, the findings unveiled the statistically significant effect of task planning on the three features of argumentative writing. The comparison of the mean differences indicated that rehearsal task planning affected the learners' fluency in argumentative writing more than their accuracy, complexity, and lexical variety in writing. In addition, rehearsal planning had the lowest effect on the participants' lexical variety in writing. For the L2 argumentative post-test writing, the factor levels of *types of task planning* were shown along the horizontal axis for the accuracy scores. Overall, the three experimental groups surpassed their counterparts in the control group, as the line for the control groups sloped downward, but for the experimental groups, it sloped upward. Concerning the argumentative writing complexity, the experimental groups outdid the control group. The rehearsal group had the highest complexity score among the experimental groups. Concerning the lexical variety, the experimental groups outperformed the control group as well. The experimental group working on rehearsal task planning reported the highest lexical variety score among the experimental groups. The experimental groups reported higher accuracy, complexity, and lexical diversity scores than the control group for the argumentative post-test writing.

The result of qualitative content analysis was quantified through the use of descriptive statistics and MANOVA as presented in the following tables. The descriptive statistics revealed that fluency (*SD* = .99) received higher allocated attention in comparison to other features mostly in the rehearsal group. The allocated time to the number of the T-units (*M* = 39.75) received the focal attention in the rehearsal group followed by the allocated time to the number of clauses (*M* = .3812) revealing that recording the idea they have in their mind either in the form of sentence or clause is a priority for them. The allocated time to the number of the words (*M* = 15.40) was ranked the third in rehearsal, strategic, online, and control groups, respectively. Accuracy was the second feature (*SD* = .98) that captured the allocated attention of the learners. Again, as in fluency, the rehearsal group superseded the rest (*SD* = .98), followed by the strategic counterpart (*SD* = .97). The allocated time to the total number of T-units (*M* = .39.50) received noticeable attention in comparison to the other sub-scale, namely, error-free clauses (*SD* =. 69). The findings revealed that the learners' primary concern was capturing the ideas through t-units, while the accuracy of the tabulated sentences was ranked second. Complexity (*SD* = .035) stood the third after complexity, in which the allocation to the total number of T-units (*M* = 39.57) superseded the allocated time to the total number of clauses (*M* = 36.35) in the rehearsal group. The rehearsal group reported a higher rate of allocated attention to the complexity (*SD* = .035) than the other three groups. The strategic group (*M* = .031) and online planning group (*M* = .024) ranked second and third, respectively. In contrast, the control group had the lowest mean for the number of error-free T-units. Lexical diversity (M = .83) received the rehearsal group's lowest attention. Before running the MANOVA, a Shapiro-Wilk test was run to assess the normality of distribution of allocated attention to accuracy, complexity, and lexical variety for the post-test scores (See Table 4).

© Shiraz University

Table 4

Test of Normality for the Allocated Attention

	Group	Allocated Attention	Shapiro-Wilk		
			Statistic	df	Sig.
Fluency	Rehearsal	.0298	.998	20	.701
	Strategic	.0280	.980	20	.281
	Online	.0279	.972	20	.222
	Control	.0262	.966	20	.461
Accuracy	Rehearsal	.0238	.972	20	.701
	Strategic	.0210	.931	20	.281
	Online	.0209	.881	20	.322
	Control	.0202	.960	20	.461
Complexity	Rehearsal	.035	.920	20	.120
	Strategic	.033	.953	20	.282
	Online	.031	.955	20	.340
	Control	.024	.988	20	.759
Lexical variety	Rehearsal	0.34	.963	20	.438
	Strategic	0.29	.978	20	.793
	Online	0.26	.146	20	.260
	Control	0.46	.199	20	.073

As the P-values were higher than (.01), it could be concluded that the scores were normally distributed. Homogeneity of Covariance Matrices was checked by computing Box's M (Sig = .450 with F = 1.08) higher than .05. The covariance matrixes of the dependent variables were equal across the four groups. Next, a one-way between-groups MANOVA was run (See Table 5).

© Shiraz University

Table 5

Multivariate Tests for the Allocated Attention under Three Planning Conditions

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Allocated Attention	Pillai's Trace	1.37	18.02	9.00	228.00	.00	.435
	Wilks' Lambda	.038	67.88	9.00	180.24	.00	.700
	Hotelling's Trace	25.90	25.99	9.00	218.00	.00	.902
	Roy's Largest Root	25.51	66.58	3.00	791.00	.00	.990

The increased and positive value of Pillai's trace statistic for the *allocated attention* (Pillai's trace = 1.37) indicated that this effect contributed to the model. The significance value of the main effect was less than (.05), indicating that the effects contributed to the model, too. Using general rules of thumb given by Cohen et al. (2001), the Partial eta squared was large ($\eta 2$ partial$_{\text{Pillai's}}$= .435). The multivariate test showed a significant main difference of task planning (rehearsal, strategic, and online) on EFL learners' allocated attention. Significant differences were found in the dependent variables among the groups ($p < .05$).

One-way MANOVA examined whether the time planning type affected the allocated time to the accuracy, complexity, and lexical variety. The results revealed that types of time planning had a significant effect on the allocated attention to the three writing features (Pillai's trace = 1.37, Wilks' Lambda = 0.38, $F_{(9,\ 180.247)} = 67.88$, $p < 0.01$, $\eta 2$ partial $_{\text{Wilk}}$ = 0.990. Concerning

© Shiraz University

accuracy, complexity, and lexical variety, the results of the Scheffe test uncovered that the largest difference was between the rehearsal and control groups with the mean difference (M = 34.5100) and (M = 20.89). In contrast, the lowest differences were reported between the rehearsal group and the strategic group with the mean difference (M = 34.51) and (M = 29.1) as well as the online group and the strategic group (M = 29.1) and (M = 25.98), respectively.

In addition, the differences among the three experimental groups concerning the allocation of attention were statistically significant ($p \leq .05$). In general, statistically significant differences were reported between each of the three experimental groups working on three types of task planning in terms of allocated attention to the three levels of argumentative writing and the control group ($p \leq .05$). Moreover, the comparison of the mean differences indicated that the allocation of attention to the rehearsal time planning was more oriented toward fluency than their accuracy, complexity, and lexical variety in writing. This means that while capturing the ideas through sentences and clauses was more susceptible to attentional control, syntactic complexity and lexical variety were not much affected by learners' planning in advance. Overall, the experimental groups outperformed the control group.

Discussion

The first purpose of the study was to explore the difference in the fluency, accuracy, complexity, and lexical variety of undergraduate EFL learners' argumentative writing task performance under three types of task planning (rehearsal, strategic, and online) conditions. The findings revealed that certain aspects of learners' argumentative writings were affected under different planned conditions; however, rehearsal task planning yields higher

© Shiraz University

complexity, accuracy, fluency, and lexical variety than strategic or online task planning. The findings tally with previous findings that the pre-task planning condition leads to greater gains in complexity and fluency than in accuracy (Ellis, 2005; 2009; Fazilatfar et al., 2020; Tavakoli & Rezazadeh, 2014). The findings are also congruent with those of Ellis (2009), who considered the influences of three types of planning time conditions on the CAF of L2 learners' oral production and documented the beneficial effects of all three planning types on the pertinent variables. The results are also supported by Ong's (2014) and Ong and Zhang's (2010) studies which revealed the benefits of appropriate planning time in writing fluently and appropriately.

The findings unveiled that the rehearsal group had greater error-free T-units. The rehearsal group had slightly higher mean scores in comparison. Yalaoui and Rabahi (2017) also indicated the significant effect of rehearsal planning on reducing the number of errors and contributing to enhancing the writing accuracy of narrative compositions. Ellis (2009) also testified that rehearsal planning resulted in greater fluency, complexity, and to a lesser extent, accuracy in learners' oral performance. The rationale behind the progressive improvement of accuracy, complexity, and lexical variety in rehearsal groups could be traced back to the adequate chances given to the organization of reiterative processes of writing, lexical retrieval, the emergence of ideas, and sentence writing in rehearsal planning time condition.

Concerning fluency, the results advocated that fluency or the length of production was significantly privileged by rehearsal planning, as acknowledged by Ahmadian et al. (2015) and Ellis and Yuan (2004). They confirmed that when pre-task planners, especially the rehearsal group, had further access to the processing resources, they could compose more fluent texts. Moreover, taking Ellis and Yuan's (2004) dispute into account, it can

© Shiraz University

be assumed that pre-task planning privileged fluency in two conceivable ways. One is through enabling individuals to process and plan the writings concerning the content as well as the organization. Those with organized viewpoints on the type of the argumentation, the recognition of the core argument, the arrangement of the premise, and the organization of the evidence and supports had less burden on working memory in performing the task. The second one is its contribution to L2 learners' writing development more fluently and being less involved in in-depth monitoring.

As for complexity, the findings confirm those of Mohazabieh, Sahragard, Rassaei, and Zamanian (2020) regarding the planning conditions. They also documented that the pre-task planning group exceeded all other groups concerning complexity and fluency. Concerning pre-task planning in general, the findings of complexity are consistent with those attained in both written and oral language task performance (Abdi Tabari, 2017; Ellis & Yuan, 2004), almost all of which unveiled that pre-task planning yielded greater complexity compared to no planning condition. The results are in accordance with those of the earlier studies (e.g., Ellis & Yuan, 2004; Ahmadian et al., 2015).

With respect to the accuracy, it was found that rehearsal planners engendered more precise language than the task performers under the online and no planning conditions. This asset accentuates the discoveries of some of the preceding studies (Foster & Skehan, 1996; Haghverdi, Biria, & Khalaji, 2013). One justification that counts for the positive stimulus of the rehearsal group is their benefit from the implicit knowledge that facilitated the formulation of the views in the execution stage of the performance. The result documented by Piri, Barati, and Ketabi (2012) also showed that once students produce language twice in pre-planning mode, they would most probably access their implicit knowledge.

© Shiraz University

© Shiraz University

As for lexical variety, the results confirmed the production of samples of writings with more lexical variety under the rehearsal condition than those with strategic or online planning. A statistically significant difference was spotted among the three groups. However, the rehearsal planning had the lowest effect on the participants' lexical variety in comparison to fluency, accuracy, and complexity in writing. The results are congruent with those of Ellis and Yuan (2004), who documented that the pre-task and online groups did not advance lexical variety significantly. The results are not in line with those of Abedi Tabari (2017), who found that the online group had higher accomplishments than the pre-task group regarding lexical variety; however, the difference was not statistically significant.

The second research question concerned EFL learners' allocation of attention to lexical variety and the CAF triad under different time planning conditions. The analyses of the data revealed that the learners' attention to fluency, complexity, accuracy and lexical variety varied under three planning conditions. It was shown that the learners paid the most attention to the production of clauses to conceptualize ideas during pre-task planning. This allowed the participants to provide the necessary clauses to accomplish the task before the main performance. Deliberate attention to time planning required more attentional resources throughout the preparation and formulation process in the rehearsal group, whereas the control group engaged more in continuous writing, in which they drafted without any time planning consideration that would result in the enrichment of their written complexity. The merits of planned conditions due to providing monitoring processes and drawing the learners' allocated attention to both form and meaning facilitate the production of clauses to T-units. This might justify why pre-task planning often contributed to greater complexity but not greater accuracy. The study provided pedagogical insight for the teachers in developing their writing quality, particularly the accuracy, complexity, and lexical variety triad. EFL learners can also practice varied planning time before writing and gain further insights into how to organize planning time in the process of L2 writing to develop the quality of their writing.

Conclusion

The study addressed how linguistic features (i.e., fluency, accuracy, complexity, and lexical variety) were affected by the three types of task planning (i.e., rehearsal, strategic, and online). Unlike other studies on planning conditions, the present study attempted to underscore the essential function of modality and measure lexis as a distinct dimension of L2 written performance attesting to Skehan's (1998) assertion in considering lexical variety as an essential aspect of language performance. Moreover, it pursued EFL learners' allocation of attention in terms of CAF and lexical variety across different time planning conditions. The main theoretical implication of this study is that it expands the literature by extending Skehan's (1998) Trade-off Hypothesis from speaking to writing, specifically in the argumentative genre. The pattern obtained from the findings provided further evidence in support of trade-off effects between complexity and accuracy under three planning conditions. Due to the paucity of research on written language production from the limited attentional capacity model and the Trade-off Hypothesis concerning the allocation of attention and lexical variety, the findings of this study open up a range of possibilities for future research. What emerged in this study confirmed the underlying theoretical rationale that learners' attentional capacity is restricted. Therefore, paying attention to one dimension of language may reduce the attention to other features of

© Shiraz University

performance. As it happens in this study, the participants were performing the tasks under some information-processing pressure after planning in which they had to allocate attention to fluency at the expense of other features such as complexity and accuracy. The limited attention to complexity and accuracy might also be due to the task type. Producing argumentative compositions may impose additional cognitive loads which bring about an attentional shift to focus on other aspects of writing like fluency (Tabari, 2017). Moreover, planning is argued to alleviate the learners' cognitive processing load and ease the recall of all the relevant background knowledge. The results indicate that by engaging in such pre-task planning, the learners can pay more attention to how they carry out the tasks. However, instructors should also pay close attention to the sequence and grading of the tasks selected for instruction to develop learners' writing ability in an appropriate manner. The results of this study have some pedagogical implications, too. As Ellis and Yuan (2004) proposed, the use of planning can be a potent pedagogical means for language learners to develop their writing ability. Accordingly, it may be beneficial for teachers to promote a variety of planning types and activities in L2 writing classes. EFL Teachers could train learners on how to utilize planning types in performing writing tasks in classes to help them organize and arrange information according to available time as well as the nature of the writing tasks.

Despite the positive effect of planning on L2 argumentative writing, some limitations are to be considered in future studies. The first potential limitation is the genre under study. The task type was delimited to argumentative writing to control extraneous effects on the findings, and consequently, the results might not be generalizable to other studies with other types of task performance. Thus, more research can be conducted to assign tasks to other

© Shiraz University

genres in writing. The next limitation was the operationalization of the CAF measures and lexical variety that was limited to the aforementioned scales; however, there are other scales to be used in future studies, as complexity, for example, was challenged by some scholars to comprise noun-phrase complexity, lexical density, lexical complexity, and syntactic complexity (Lu, 2011). Therefore, the recently added complementary measures such as noun-phrase complexity and lexical variety in line with recent conceptualizations of complexity should be regarded in future research. Since the present study mainly followed a quantitative design and the qualitative data were also analyzed quantitatively, future studies can employ a mixed-method design to work on the participants' feelings and attitudes while carrying out written tasks. In addition, although the type of data analysis run was in line with that of the previous studies, it might affect the generalizability to some extent. As such, the results are to be taken more cautiously. The last limitation was the study's context, which was limited to the Iranian EFL learners at the upper-intermediate level. Other studies can consider the comparative effect of language proficiency. Other studies can also work longitudinally to collect rich data on how operating the planning time conditions might stimulate L2 written output as measured in terms of the CAF dimensions and lexical variety.

© Shiraz University

References

Abdi Tabari, M. (2017). Investigating the effects of planning time on the complexity of L2 argumentative writing. *The Electronic Journal for English as a Second Language, 21*(1), 1–24.

Ahmadian, M. J. (2012). The effects of guided careful online planning on complexity, accuracy, and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research, 16*(1), 129–149.

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*(2), 35–59.

Baleghizadeh, S., & Nasrollahi Shahri, M. S. (2013). The effect of online planning, strategic planning and rehearsal across two proficiency levels, *The Language Learning Journal*, *45* (2), 171–184.

Cohen, J., Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Styles shifting in the use of past tense. *Studies in Second Language Acquisition*, *9*(2), 1–20.

Ellis, R. (2005). *Planning and task performance in a second language*. Philadelphia: John Benjamins.

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.

Fazilatfar A. M., Kasiri, F., & Nowbakht, M. (2020). The comparative effects of planning time and task conditions on the complexity, accuracy, and fluency of L2 writing by EFL learners. *Iranian Journal of Language Teaching Research* *8*(1), 93–110.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(4), 299–323.

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

Gass, S., A. Mackey, M. J. Alvarez-Torres, & M. Fernandez-Garcia. (1999). The effects of task repetition on linguistic output. *Language Learning* *49*(4), 549–81.

George, D., & Mallery, P. (2010) *SPSS for Windows step by step: A simple guide and reference 17.0 update*. Pearson: Boston.

Haghverdi, H., Biria, R., & Khalaji, H. (2013). The impact of task-planning and gender on the accuracy of narrations composed by Iranian EFL learners. *Journal of Language Teaching and Research*, *4*(1), 74–83.

© Shiraz University

Jacobs, H. L., Zinkgraf, S.A., Wormouth, D.R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach.* Rowley, MA: Newbury House.

Johnson, M. D. (2014). Does planning really help? Effectiveness of planning in L2 writing. *Journal of Second Language Teaching and Research*, *3*(1), 107–118.

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing, 21*, 264–282.

Kargozari, H. R., Soleimani, H., Jafarigohar, M., & Hemmati, F. (2016). Comparing the effects of two types of planning on Iranian EFL learners' writing performance in terms of fluency, accuracy, and complexity. *Journal of Language & Translation Studies*, *48*(4), 29–33.

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, *103*, 327–342.

Khomeijani Farahani, A., & Faryabi, F. (2016). The effects of pre-task planning on Iranian EFL undergraduates' argumentative writing task performance. *Journal of Recent Research in English Language Studies, 3*(2), 97–116.

Khomeijani Farahani, A., & Meraji, S. (2011). The effect of length of pre-task planning time on discourse-analytic measures and analytic ratings in L2 written narratives. *Journal of English Studies*, *1*(2), 21–37.

McCarthy, P. M. & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392.

Malvern, D. D., & Richards, B. J. (2000). Validation of a new measure of lexical diversity. In Beers, M, v. d. Bogaerde, B., Bol, G., de Jong, J., & Rooijmans, C. (Eds.), *From sound to sentence: Studies on first language acquisition.* Groningen: University of Groningen, Centre for Language and Cognition.

Marzban, A., & Norouzi, M. (2010). The effect of strategic planning focus and time on writing fluency and accuracy. *Journal of English Language Pedagogy and Practice*, *4*(8), 122–145.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*(3), 323–337.

Mohazabieh, S., Sahragard, R., Rassaei, E., & Zamanian, M. (2020). The potential combined effects of task complexity and planning types on Iranian EFL

© Shiraz University

learners' oral production performance. *Applied Research on English Language*, *9*(4), 503–538.

Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

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

Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, *23*, 17–30.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*, 218–233.

Piri, F., Barati, H., & Ketabi, S. (2012). The effects of pre-task, online, and both pre-task and online planning on fluency, complexity, and accuracy – The case of Iranian EFL learners' written production. *English Language Teaching*, *5*(6), 158–167.

Rahmanian, M. (2004). *The relationship between pre-task and online planning and fluency, accuracy and complexity of Iranian EFL student's production in written narrative and descriptive tasks*. Unpublished master's thesis, Tarbiat Modares University, Tehran, Iran.

Rahimpour, M., & Safarie, M. (2011). The effects of online and pre-task planning on descriptive writing of Iranian EFL learners. *International Journal of English Linguistics*, *1*(2), 274–280.

Rostamian, M., Fazilatfar, A. M., & Jabbari, A. A. (2018). The effect of planning time on cognitive processes, monitoring behavior, and quality of L2 writing. *Language Teaching Research*, *22*(4), 1–21.

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 thesis, Graduate School of Seoul National University, Seoul, South Korea). Retrieved from https://hdl.handle.net/10371/127538

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

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

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

© Shiraz University

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.

Soleimani, H., & Kargozari, H. R. (2014). Uncovering new trends of investigation in writing task planning. *International Journal of Asian Social Science*, *4*(9), 991–1002.

Storch, N. & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157–177). London, UK: Multilingual Matters.

Storch, N., & Wigglesworth, G. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445–466.

Tavakoli, M., & Rezazadeh, M. (2014). Individual and collaborative planning conditions: Effects on fluency, complexity, and accuracy in L2 argumentative writing. *The Journal of Teaching Language Skills*, *5*(4), 85–110.

Wigglesworth, G., & Storch, N. (2009). versus individual writing: Effects on fluency, complexity, and accuracy, *Language Testing*, *26*(3) 445–466.

Yalaoui, F. Z., & Rabahi, K. (2017). *The effect of rehearsal planning on the accuracy of narratives composed by secondary school students: The case of third year-foreign languages stream at NEDJAR HOCINE Secondary School* (Master's thesis, University of Tlemcen, Sidi Amar, Algeria). Retrieved from dspace.univtlemcen.dz/handle/112/11317

Yuan, F. (2001). *The effects of planning on language production in task-based language teaching*. (Doctoral thesis, Temple University, Philadelphia, PA).

© Shiraz University
پرتال جامع علوم انسانی