

Integrating Questioning and Task Planning: Does it work for EFL Oral Production?

Mohamad Reza Mollahosseiny¹
Davood Mashhadi Heidar*²

Received: 2019-07-22 | Revised: 2019-12-04 | Accepted: 2019-12-29

Abstract

Although the role of task planning conditions affecting EFL learners' oral performance has been of investigators' enduring concern, to the best of our knowledge, no study so far has investigated the influence of planning time in combination with a cocktail of questioning instruction on the oral performance in EFL settings. The present study thus examined the combined impact of pre-task planning and questioning instruction on the accuracy and complexity of Iranian EFL learners' oral presentation of intermediate learners of English at a private language institute in Mazandaran. Drawing on Bloom's taxonomy of higher order questions (HOQs), a pretest posttest quasi-experimental design was utilized. Initially, three groups including a control and two experimental groups engaged in pre-intervention oral tasks and performed the post-intervention oral tasks ensuing rehearsal planning plus HOQs (EX1), strategic planning in addition to HOQs (EX2), and no planning condition along with HOQs (CG) instructions. Results pointed to the positive impact of pre-task planning integrated with HOQs on the complexity as well as accuracy of oral narra-

¹ PhD Candidate, Department of English, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran; reza.mollahosseiny@gmail.com

² Assistant Professor, Department of English, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran; (Corresponding author); davoodm_tarbiatmodares@yahoo.com

tive tasks. More specifically, engaging in rehearsal planning condition plus HOQs proved to be optimal for inducing better results in EFL learners' oral production in terms of all three measures of performance. Accordingly, the findings rejected a conflict between complexity and accuracy in that both of them were found to have increased simultaneously. Finally, drawing on the outcomes of the investigation, it is proposed that in order for task-based instruction to more beneficially contribute to EFL learners' task performance, a more flexible view of implementing TBLT needs to be merged with other influential educational factors including thinking skills which have proved to be fruitful in the present study.

Keywords: Rehearsal Planning, Strategic Planning, Speaking, Accuracy, Complexity.

Introduction

Achieving proficiency in speaking skill is the primary phase toward complete acquisition of English in EFL/ESL contexts. Despite this significance, unfortunately the oral skill and its associated activities are almost overlooked in EFL classrooms (Bora, 2012). As a result, mastering speaking continues to be a challenging endeavor for the majority of EFL learners (Diaab, 2016; Zhang, 2009). In this regard, task planning is anticipated to be a principal component in stimulating EFL learners' oral competence (Ellis, 2005; Foster & Skehan, 1996). In general terms, task planning is defined as the allocation of time in advance or during performing a task (Mochizuki & Ortega, 2008). Depending on the span of time offered either prior to or during performance of tasks, Ellis (2009) catalogs task planning in a dichotomy of pre-task and within-task planning. Pre-task planning includes strategic and rehearsal planning. In rehearsal planning, learners have the chance to carry out the task in advance of the actual task presentation (Ellis, 2005). Strategic planning, alternatively, involves learners' preparation of the content of the task they are going to perform. Furthermore, it is Skehan's (1998) three-dimensional differentiation amongst fluency, accuracy and complexity (CAF) of linguistic performance that forms the foundation for measuring learners' language production (Larsen-Freeman, 2006; Rosmawati, 2014). However, to explain the influence of planning on various components of language performance, scholars who believe that humans possess a restricted attention aptitude and processing capability (Skehan, 2009; Skehan & Foster, 1997) proposed a trade-off effect. In this sense, directing attention toward one dimension is likely to be at the expense of jeopardizing others if the performer is not aided through deploying performance circumstances such as the planning time offered. Particularly, there has been a tension between accuracy and complexity during oral task practices (Garcia-Ponce, 2017). Furthermore, numerous studies have evidenced that planning enhances fluency and complexity of production in instructional settings (Abdi et al., 2012; Ahangari & Abdi, 2011; Skehan & Foster, 1997). Among the components of performance, fluency effects seem to be the clearest and most stable (Tavakoli & Skehan, 2005). Nonetheless, results concerning the impact of planning on the second of these components (i.e. accuracy) are inconsistent and mixed, which

has led researchers to question the contribution of pre-task planning to language performance in this regard.

Besides planning time, thinking skill is a broadly accepted element among investigators as it plays a fundamental role in the process of second or foreign language speaking (Naghdipuor & Emeagwali, 2013). Acknowledging thinking as an essential requirement to effective learning, the educational process in the modern era calls for the improvement of higher-order thinking skills as one of the paramount goals to practice both intellectual and affective spheres so as to attain or provide information, resolve problems, or make decisions for the different vital activities (Boa et al., 2018; Hashim & Yaakub, 2004; Mohd & Hassan, 2005; Yee et al., 2015).

Most of the studies conducted on higher order thinking skills (HOTS) drew on Bloom's Taxonomy (1956) to structure classroom activities and foster thinking skills from basic knowledge recall (lower order thinking skills) all the way through to higher order thinking skills. Despite the array of studies addressing the role of higher order thinking skills and language achievement, it seems that thinking skill has not sufficiently been employed to augment EFL learners' speaking performance. According to Richland and Begolli (2016), effective instruction that emphasizes higher order thinking is challenging. Wilen (1991) and Chen (2016) put forward higher order questioning as a way for activating learners' thinking. Bloom's taxonomy (1956) is commonly used to classify cognitive dimensions of learning and differentiates between lower and higher order questions (Richland & Simms, 2015). Questions can be categorized in relation to the category of thinking skill essential for the response. Higher-order questions require learners to manipulate information while utilizing higher-order thinking, whereas lower-order questions are those that call for the engagement of lower-order thought as they merely evoke the prearranged data from memory (Bernadowski, 2006). Inquiry on the prominence of questioning as a teaching and learning instrument is well recognized (Albergaria Almeida, 2010; Chin & Osborne, 2008). Several authors advocate the view that the development of the students' questioning has the potential to develop higher order thinking skills (Hofstein et al., 2005).

However, asking higher-level questions is not a habitual practice of teachers and students (Albergaria Almeida, 2010). Moreover, with regard to the impact of higher order questions on speaking in Iranian EFL context, although scholastic policy handouts from all over the world stress the necessity of training higher order thinking skills as the pivotal and pillar component of 21st century skills in several disciplines, unfortunately teachers and learners in Iranian EFL settings seem to be lagging behind (Ketabi et al., 2012). Iranian EFL classrooms are still mainly described by pedagogy of knowledge transference through memorization and concentrating on lower-order thinking skills (Sanavi & Tarighat, 2014).

Literature Review

The related literature is abundant with studies conducted on the role of planning on EFL/ESL task performance. In a pioneering piece of research, Foster and Skehan (1996) studied the impact of pre-task planning on oral task presentation. Implementing various task types including individual information, storyline, and decision-making tasks, the researchers found that when learners are offered with a chance to plan a task prior to accomplishing it, the language they generate is more complex and fluent than the time when no planning is feasible (Ellis, 2005). In a subsequent study, Skehan and Foster (1997) postulated once more that complexity and accuracy are in tension for attentional aptitude when task burdens are amplified along the planning time in that effective performance in task-based milieus has often been considered as creating more advanced language, resulting in complexity as well as an apprehension to avoid error, leading to higher accuracy (Skehan, 2009). Accordingly, it was concluded that planning time can merely be directed toward one of the aspects of performance (either accuracy or complexity), and not to the two dimensions concurrently. In favor of these claims, Ahangari and Abdi (2011) have found that pre-task planning favors complexity while no positive influence was found on the accuracy of learners' oral performance. In another study in an Iranian milieu, Mehrang and Rahimpour (2010) tried to probe into the effect of planning circumstances on the performance of the EFL learners while carrying out structured vs. unstructured oral tasks. Outcomes revealed that planning time had no influence on the accuracy and fluency of the learners' performances, but brought about more complex performances when participants accomplished the unstructured task. Iwashita et al. (2001) examined the distinctive influences of implementation of planning time on oral performance. Measures of accuracy and complexity did not show any difference in the planning conditions. These results are in contrast to the extant findings and have led researchers to question the contribution of pre-task planning to language performance.

As the review of related literature reveals, numerous studies have evidenced that planning works in favor of fluency and complexity of production in scholastic settings. However, results with regard to the effect of planning on accuracy are inconsistent and mixed. In the meantime, a number of studies are addressing HOTS (Chen, 2016; Jensen et al., 2014; Miri et al., 2007; Zohar & Cohen, 2016).

Yee et al. (2015) suggest that learners require to learn higher order thinking skills to address the problems they encounter in creating ideas in that such problems can lead to poor performance on the task. The results of Jensen et al.'s (2014) study suggest that learners who are assessed all through the term with high-order questions gain a deeper theoretical comprehension of the material and an improved memory for the course material, and this lends support to the proposed hierarchical core of Bloom's taxonomy. Moreover, as recommended by Miri et al. (2007), if teachers deliberately and insistently exercise higher order thinking tactics, there is a good opportunity for a consequent advancement of critical thinking capabilities. Despite the abundance of empirical

evidence concerning the effectiveness of teaching higher order thinking skills, controversies still linger on the issue as they derive from notions for and against HOTS instruction. As Smith (2015) puts it, higher order thinking is the most problematic skill to develop. In that sense, the majority of classrooms worldwide are still predominately characterized by the pedagogy of knowledge transference, centring on lower-level thinking skills. This incongruity necessitates studying challenges on the implementation of HOTS (Zohar & Cohen, 2016), challenges that, *inter alia*, are involved in weaving HOTS into specific educational contents.

Theoretical Underpinning

Bloom's hierarchy (1956) of thinking skills was taken as the theoretical framework of the present study to categorize and analyse question types (Bernadowski, 2006). The initial three levels of Bloom's taxonomy, *i.e.* knowledge, comprehension, and application, are the lower order thinking skills (Ennis, 1987), whereas the remaining top categories, *i.e.* analysis, synthesis, and evaluation, belong to higher order thinking skills. It is generally believed that each behaviour has to be mastered before the next one can take place (Collins, 2014). That is, lower-order thinking is a necessity in the process of employing higher-order thinking. Bloom's pyramid classifies questions from lower levels to higher order ones (Richland & Simms, 2015). It also describes the basic words and associated questions of each level as the primary features in developing teacher's questions. Another premise which forms the theoretical foundation of this study is Skehan's Trade-off Hypothesis (1998). It maintains that complexity and accuracy are in tension and typically they improve to each other's detriment due to limited attentional capacity unless this limited capacity is mitigated by means of ideal choice of planning (Nation & Newton, 2009). Planning as a task implementation variable is supposed to influence language production which assists learners to catch up for their performance limitations while completing the task (Ortega, 1999).

Considering the significance of oral competence as well as the demanding nature of oral tasks in EFL context (Karatas et al., 2016) and drawing on the extant controversies regarding the impact of both task planning and thinking skills instruction on EFL learners' speaking and language achievement, it is believed that further investigation is essential to examine whether a combination of these variables may contribute to EFL oral performance. Therefore, this study as a part of a doctoral dissertation project will seek to adopt Bloom's taxonomy (1956) as the theoretical framework to uncover how HOTS instruction coupled with rehearsal, strategic, and no planning conditions affect accuracy and complexity of EFL learners' oral performance. The significance of this study first and foremost lies in its novelty to probe the combined effect of two separate variables which was not attempted in any other study to date. More importantly, the outline for HOTS instruction focused on infusing thinking skills into the language via higher order questions (HOQs). Furthermore, considering that TBLT is mainly a meaning-centered methodology which tries to involve

learners in meaning-dedicated language use (Shabania & Ghasemi, 2014), we attempted to cover the focus-on-form dimension as an essential feature of TBLT (Ellis, 2009) in our instructional treatment via provision of oral feedback as a supportive strategy for rehearsal planners only. However, to make error correction more effective and avoid the interruption of the learners' flow of thought and speech production, corrective feedback concentrating on the participants' grammatical and structural errors were mainly provided via clarification request, repetition, and scaffolding strategies along with some explicit feedback merely at the end of task implementation. Furthermore, adopting a cognitive approach of TBLT, we included a lexical measurement component in our measurement indices to cover the lexical aspects of performance. In doing so, the enrichment and validity of our measurement criteria is guaranteed (Skehan & Foster, 2007). In line with the goals of this study, answers to the following questions were sought:

1. Do *rehearsal* pre-task planning and HOQs instruction have any meaningful effect on accuracy and lexical and grammatical complexity of intermediate EFL learners' oral production?
2. Do *strategic* pre-task planning and HOQs instruction have any meaningful effect on accuracy and lexical and grammatical complexity of intermediate EFL learners' oral production?

Method

Participants

All 45 prospective participants (24 males and 21 females) nominated for this study were intermediate EFL learners who participated in a speaking course in Shokouh institute for English as a foreign language in Mazandran province, Iran. They were all from an EFL setting aged 18 to 26. In order to ensure the homogeneity of the sample, a proficiency test serving as a placement test was used. Students were positioned at this level based on their scores on the Nelson English Language Tests (NELT scores 35-50). Afterwards, the classes were randomly assigned as two planning groups and a no planning group (control group) including 15 students each. In groups 1 and 2, the students were requested to perform the speaking task under the rehearsal pre-task planning (RP) and strategic planning conditions (SP), respectively. The third group was asked to perform the speaking task with no planning circumstance (NP) which served as the control group.

Instruments

The Nelson English language proficiency test (NELT) was administrated to assess the proficiency level of the groups. Second, the teaching materials utilized included oral tasks as well as *HOQS* handouts comprising a classification of higher order questions, with each level defined and exemplified. The oral tasks serving as the instructional material were characterized with features of

streamlined literature together with authentic life stories and conditions focusing on familiar topics provided by the learners. In addition, two oral narrative tasks were utilized as the pre-test and post-test. Moreover, cassettes and tape recorders were employed to record the participants' oral performance.

HOQs Handouts

The 5Ws Questioning framework (five Wh question words) developed by Morgan and Saxton (2006) was adopted as the groundwork of HOQs training. It included (based on Bloom' taxonomy) formulated series of questions for classroom usage and elucidated the six thinking skills executed at each level:

1. *Knowledge*: centered on Rote memory skills.
Qs words: Who? What? When? Where? List.... How do you mention... in English?
2. *Understanding*: focused on the skill to render, rewording, or deduce material.
Qs words: What do you meant by...? Can you restate...? Can you explain...? Clarify... Can you elucidate...?
3. *Application*: centered on the ability to handover knowledge from one context to another.
Qs words: What would take place if..? If you were...? What is a novel example of...? How is...connected to..?
4. *Analysis*: aiming at the ability to find out and discriminate the constituents of a superior whole.
Qs words: Why? What inferences can you make about...? What is the dissimilarity between... and...?
5. *Synthesis*: concentrating on the skill to merge constituents into a comprehensible whole.
Qs words: How could you...? What would occur if...? What is a probable solution to...?
6. *Evaluation*: focused on the capability to decide on the significance or use of information by means of a set of values.
Qs words: Which one is superior? Would you approve that...? What is your judgment...? Is it a better answer to...?

Oral Narrative Task as the Pre-test

Following previously conducted studies including Tavakoli and Foster (2008) and Park (2010) who initially took the images from Heaton (1975), this study employed an oral narrative task, entitled "*The School Bus*" with six arrays of pictures to scrutinize L2 learners' oral performance. Narrative tasks are common in this line of research and it was supposed to assist comparison with previous conclusions. The general guidelines for the oral narrative task were given both in English and Persian. Then the participants were given pictures based on

the story and asked to narrate or explain at least three sentences for each picture describing what was occurring in the pictures. Such a picture-based oral narrative task was utilized assuming that it could guarantee that the task was practically challenging for the learners and would prompt their linguistic aptitudes (Ellis & Yuan, 2004).

Oral Narrative Task as the Post-test

To make a reliable comparison with former investigations in the related domain, an oral narrative task in the form of picture-grounded story narration named “*The Lost Package*” was used as the post-test to gather quantitative data. The set of pictures employed for this task was adopted from Piri et al. (2012), who initially took the pictures from Heaton (1975). The general guidelines for the task were given both in English and Persian. All the participants were asked to narrate at least three sentences about each picture recounting what occurred in the pictures.

Procedures

Primarily, an oral narrative task was administered to further ensure the participants' proficiency level. Later, the dual instructional intervention commenced. A control and two planning conditions were designed to elicit the learners' oral production. Prior to task performance, all subjects were instructed on how to accomplish the task in order to ensure that they would implement the task in the trained way. The rehearsal planners executed the same task twice with a time lapse of twelve days between the two presentations. It has to be noted that the participants had not been told previously about the replication of the task in order to reduce the rehearsal effect. In fact, the participants in the RP group had an opportunity to plan and perform the tasks prior to the main task performance. They were given 10 minutes to arrange for the rehearsal task planning (performance) and 8 minutes for the main task performance. For the rehearsal phase, the participants in this group were given papers to jot down notes throughout the pre-task planning phase along with the set of pictures representing the story of the narrative task. Once the planning time was finished, the notes were taken away as they initiated carrying out the task.

On the other hand, the strategic pre-task planners were assigned 10 minutes to arrange for the task and plan their explanations before the main performance. They were told to look at the pictures in which clues were given, and individually plan what they wanted to say. As proposed by Lavolette (2013), strategic planning is executed by allocating learners time (usually about 5-10 minutes) to plan what they want to say in advance of being asked to perform a task actually. They were also requested to finish the main task performance within a time limitation of about 8 minutes. Moreover, participants could jot down notes on pieces of paper which were later taken away prior to the main task performance phase.

Accordingly, the only prominent difference between the rehearsal and strategic planning group is whether actually perform the designated oral narrative task or are just engaged in planning and getting prepared for the main task performance.

The second component of experimental intervention entailed teaching HOQs through which each thinking skill category (based on Bloom's taxonomy) was explicitly explicated followed by opportunities for exemplifying and practicing thinking skills. Following (Swartz & Parks, 1994; Swartz et al., 2007), teaching HOTS involved the researchers' designed lessons in which the thinking skills and the curriculum content were taught concurrently. The students were introduced overtly to tactics for more skillful thinking (via HOQs), and then encouraged to use these strategies to reflect on the content they were learning via oral tasks. By highlighting higher-order thinking (via HOQs) in content teaching, deeper comprehension is assumed to be gained. It is believed that, when this integration is combined with a highly scaffolded assistance on the part the instructor along with planning by the students (about how they will be involved in the same type of thinking activity competently), a very powerful learning environment is created next time performing that activity (Swartz & Parks, 1994; Swartz et al., 2007).

In the no planning group (NP) which acted as the control condition, students performed the task while being extremely pressured in that they had to initiate their oral production immediately after a short look at the picture series (0.5 minutes) and were asked to perform the oral narrative task within the time limit (8 minutes). They were also urged to say at least 10 sentences. The justification for this time limit lies in the researchers' expectation to create an opposing condition so that it can pave the road for a better comparison with two the planner groups. Moreover, unlike the experimental conditions, higher order questioning was taught implicitly.

The experimental instruction ran for 12 sessions. Finally, the participants in all groups did the same oral task as the post-test at the end of course of instruction. The content of the test was based on the material taught as the treatment. Having recorded each participant's speech, the researchers transcribed, segmented, and analyzed the data in terms of the three production measures of accuracy, and lexical, and grammatical complexity (see the following section for the operational definitions of these variables). To further warrant that the subdivision and scoring of the transcripts were accomplished consistently, fifty percent of the data were double checked (subdivided, coded and scored) by an independent expert colleague. Intercoder/inter-rater reliability coefficient magnitudes were above .92 for all measures (with a mean of .90). The scores were then entered into SPSS version 18.0 and checked in terms of normality of distribution using indices of skewness and kurtosis. Finally, one-way ANOVAs followed by Scheffe post hoc tests were run to answer the research questions of the study.

Variable Measures

In order to score the collected data, the measures used by Abdi et al., (2012) and Khan (2011) were adopted for assessing the accuracy and complexity of the participants' performance. Accuracy measurement was attained by calculating the proportion of error-free clauses in the whole number of clauses created and the resulting number was then multiplied by 100. Grammatical complexity measurement was operationalized as the number of clauses per T-unit. A T-unit, or minimal terminal unit, is described as "one main clause with all subordinate clauses attached to it" (Hunt, 1965: 20). In order to measure "Grammatical complexity", the percentage of clauses to T-units in the learners' narratives was evaluated by dividing the number of clauses by the number of T-units in each narrative. Lastly, lexical complexity was measured via type-token ratios (TTR), i.e. the entire number of diverse words used (types) divided by the entire number of words in the text (token) (Robinson, 1995).

Results

Testing Assumptions

According to (Field, 2009), interval data, independence of subjects, normality, and homogeneity of variances assumptions have to be met before one attempts to employ parametric tests. The first assumption is fulfilled in that the present data are measured on an interval scale. The notion of *independence* of subjects is attained as the performance of any given subject is not dependent on that of other individuals (Bachman, 2005). The third assumption relates to the normality of pre-test scores which are tested by means of skewness and kurtosis ratios over their relevant standard errors. Table 1 indicates that the pre-test scores for the three measures of performance in the rehearsal, strategic, and no planning group (control group) have a normal distribution as the skewness and kurtosis ratios over their relevant standard errors are within the ranges of +/- 1.96.

Table 1.
Normality Tests for Three Measures of Performance

Measure	Group	N	Skewness		Kurtosis	
			Statistic	Std. Error	statistic	Std. Error
Accuracy	Rehearsal	15	-.282	.580	1.401	1.121
	Strategic	15	-1.112	.580	1.322	1.121
	No planning	15	1.176	.580	-.734	1.121
L. complexity	Rehearsal	15	-1.55	.597	1.69	1.154
	Strategic	15	-.952	.597	-.228	1.154
	No planning	15	-1.566	.597	.501	1.154
G.Complexity	Rehearsal	15	-.282	.580	1.401	1.121
	Strategic	15	-1.672	.580	.897	1.121
	No planning	15	-1.176	.580	-.734	1.121

The last assumption concerning the homogeneity of variances is to be considered when representing the outcomes of the inferential statistics.

Investigating Research Questions of the Study

Before discussing the results of the study, it is essential to consider the last assumption i.e. the homogeneity of variances which is the most important notion among the set of assumptions that must be met before the test can be used properly. Levene's Test was utilized for this aim. A quick look at Table 2 reveals that the assumption of homogeneity of variance has not been violated for the three groups' performance scores in that the Sig. value for the Levene's test for the three measures of performance (i.e. accuracy, lexical complexity, and grammatical complexity) are .125; .101; .597, respectively, which are greater than .05.

Table 2.

Test of Homogeneity of Variances for Performance Scores in Three Groups

Measure	Levene Statistic	df1	df2	Sig.
Accuracy	2.184	2	42	.125
L.Complexity	2.419	2	42	.101
G.Complexity	.522	2	42	.597

Initially, prior to the treatment and in an attempt to guarantee the homogeneity of participants at the onset of study, a set of one way ANOVAs were run to measure the homogeneity of the obtained pre-test scores in terms of the three measures of performance, the results of which are presented in Table 3 below.

Table 3.

One-way ANOVA for Three Measures of Performance in Pretest

Measure	Mean (SD)			F. value	Sig
	RP	SP	NP		
Accuracy	13.86(.516)	13.93(.258)	14.13(.351)	1.896	.163
L. complexity	14.06(.258)	13.86(.351)	13.80(.414)	2.395	.104
G. complexity	14.00(.377)	14.13(.351)	14.06(.457)	.420	.660

As can be seen in the Table 3, the results of the ANOVA for the RP ($M = 13.86$, $SD = .516$), SP ($M = 13.93$, $SD = .258$), and control group (NP) ($M = 14.13$, $SD = .351$) in terms of accuracy show no substantial difference in the three groups' performance, $F(2, 42) = 1.89$, $p = .163$, $p > .05$.

Regarding the lexical complexity of oral performance, the obtained results, as presented in Table 3, indicate no significant difference between three groups' performance $F(2, 42) = 2.39$, $p = .104$, $p > .05$. Furthermore, the results

of the one way ANOVA (Table 3) failed to find any substantial difference among the three groups in terms of grammatical complexity of oral production, $F(2, 42) = .420, p = .660, p > .05$.

Accordingly, drawing on the outcomes of the one way ANOVA performed on the pre-test scores, we can confidently report the homogeneity of participants at the very beginning of the study. To answer the research questions posed, a set of one-way ANOVAs followed by Scheffe post hoc tests were run.

Table 4.
One-way ANOVA for Three Measures of Performance in Post-test

Measure	Mean (SD)			F. value	Sig
	RP	SP	NP		
Accuracy	16.33(.975)	14.06(.258)	14.20(.861)	41.319	.000
L. complexity	16.60(.736)	16.40(.507)	13.80(.414)	113.029	.000
G. complexity	16.93(1.09)	16.66(1.04)	14.40(.632)	32.239	.000

ANOVA results, as observed in Table 4, illustrate a meaningful difference among the three groups in terms of accuracy scores at the $p < .05$ level, $F(2, 42) = 41.31, p = .000, p < .05$. Luckily, the p value (.000) was less than .05, and our F value, 41.31 was more than F critical (4.38).

Likewise, the ANOVA results shown in Table 4 reveal a substantial difference in lexical the complexity scores of the three groups at the $p < .05$ level, $F(2, 42) = 113.02, p = .000, p < .05$. Luckily, the p value (.000) was less than .05, and our F value, 113.02 was above the F critical (4.38). Furthermore, a quick look at the ANOVA results depicted in Table 4 reveals that there is a substantial difference in grammatical complexity scores among the three groups at the $p < .05$ level, $F(2, 42) = 32.23, p = .000, p < .05$. Our p value (.000) was lower than .05, and our F value, 32.23 was above the F critical (4.38). Since ANOVA does not inform us about the precise location of the differences among the groups, Scheffe's post hoc test was run subsequently. The results of the Scheffe test are presented in Table 5 below. The first hypothesis predicted that RP condition merged with HOQs will result in more grammatically and lexically complex oral production, but not a more accurate performance.

The results of the Scheffe post-hoc test (see Table 5) revealed a meaningful difference in accuracy scores between the rehearsal group ($M = 16.33, SD = .975$) and the no planning group (control group) ($M = 14.20, SD = .861$) with the mean difference of 2.13, $p = .000, p < .05$, in which p value, .000, was less than .05.

The Scheffe's post-hoc test also (Table 5) found a meaningful difference in accuracy scores between the rehearsal planning group ($M = 16.33, SD = .975$) and strategic planning group ($M = 14.06, SD = .258$) with the mean difference of 2.26, $p = .000, p < .05$, in which the p value, .000 was below .05.

Table 5.
Scheffe Post-hoc Test for three Measures of Performance in Post-test

Measure	Groups	Mean Dif- ference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Accuracy	RP					
	SP	2.26667*	.27983	.000	1.5565	2.9768
	NP	2.13333*	.27983	.000	1.4232	2.8435
	SP					
	RP	-2.26667*	.27983	.000	-2.9768	-1.5565
	NP	-.13333	.27983	.893	-.4538	.5768
Grammatical Complexity	NP					
	RP	-2.13333*	.27983	.000	-2.4872	-1.4232
	SP	.13333	.27983	.893	-.5768	.8435
	NP					
	RP	.26667	.34672	.745	-.6132	1.1465
	SP	2.53333*	.34672	.000	1.6535	3.4132
Lexical Complexity	NP					
	RP	-.26667	.34672	.745	-1.1465	.6132
	SP	2.26667*	.34672	.000	1.3868	3.1465
	NP					
	RP	-2.53333*	.34672	.000	-3.4132	-1.6535
	SP	-2.26667*	.34672	.000	-3.1465	-1.3868
Accuracy	RP					
	SP	.20000	.20778	.632	-.3273	.7273
	NP	2.80000*	.20778	.000	2.2727	3.3273
	SP					
	RP	-.20000	.20778	.632	-.7273	.3273
	NP	2.60000*	.20778	.000	2.0727	3.1273
Grammatical Complexity	NP					
	RP	-2.80000*	.20778	.000	-3.3273	-2.2727
	SP	-2.60000*	.20778	.000	-3.1273	-2.0727
	NP					

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

In short, the outcomes demonstrated that the rehearsal planning group outperformed the other two groups in terms of accuracy of oral performance. Thus, it is confirmed that HOQs intervention has a substantial impact on the accuracy of performance under the RP condition. Accordingly, hypotheses 1 regarding accuracy suggesting that "RP condition merged with HOQs will lead to accurate oral production" is confirmed.

Scheffe's post-hoc test (see Table 5) showed a meaningful difference in the grammatical complexity scores of the rehearsal group ($M = 16.93$, $SD = 1.09$) and control group ($M = 14.40$, $SD = .632$) with the mean difference of 2.53, $p = .000$, $p < .05$, in which p value, .000, was less than .05, proving the usefulness of experimental instructions in the grammatical complexity. The results of the

Scheffe's post-hoc test, as represented in Table 5, also revealed that there was a significant difference in the lexical complexity scores of the rehearsal group ($M = 16.60, SD = .736$) and control group ($M = 13.80, SD = .414$) with the mean difference of 2.80, $p = .000, p < .05$, in which p value, .000, was below .05. Thus, we can confidently conclude that the research hypothesis stating that "HOQs instruction in conjunction with RP condition will have a substantial influence on lexical complexity of EFL learners' oral performance" is strongly confirmed.

The second hypothesis anticipated that SP condition combined with HOQs will bring about a more grammatically complex performance, but will not lead to a more accurate or lexically complex speech. A quick look at the ANOVA results as depicted in Table 5 reveals that there was not a substantial difference in the accuracy scores of the strategic group ($M = 14.06, SD = .258$) and control group ($M = 14.20, SD = .861$) with the mean difference of .13, $p = .893, p > .05$, in which p value, .893, exceeded .05; consequently, the second null hypothesis stating that "The strategic task planning combined with HOQs does not result in accuracy in L2 oral performance" was retained. Accordingly, it is concluded that the strategic task planning merged with HOQs does not bring about accuracy in second language oral performance.

As for the lexical complexity measure, as shown in Table 5, the results of Scheffe post hoc test indicate that there was a meaningful difference in the lexical complexity scores of the strategic group ($M = 16.40, SD = .507$) and control group (NP) ($M = 13.80, SD = .414$) with the mean difference of 2.60, $p = .000, p < .05$, in which the p value, .000, was less than .05. Therefore, the obtained results lead us to reject the second part of the second hypothesis that takes lexical complexity into consideration. This hypothesis states that "The rehearsal task planning merged with HOQs does not lead to lexical complexity in L2 oral production". Consequently, we could assert that the strategic task planning integrated with HOQs can positively influence lexical complexity in second language oral production.

Regarding the grammatical complexity measure, the results of Scheffe's post-hoc test (see Table 5) displayed a meaningful difference in the grammatical complexity scores of the strategic group ($M = 16.66, SD = 1.04$) and control group ($M = 14.40, SD = .632$) with the mean difference of 2.26, $p = .000, p < .05$, in which p value, .000, was less than .05. Therefore, it is safely concluded that the second hypothesis is solidly approved in terms of the grammatical complexity of oral performance. Thus, we could declare that the strategic task planning combined with HOQs can enhance grammatical complexity in second language oral performance.

Discussion

The present study has examined the impact of three different kinds of task planning (rehearsal, strategic and no planning) combined with HOQs instruction on Iranian intermediate EFL learners' oral production during an oral task performance. In this part, we will go over the findings of the study and discuss

the conclusions comparing them with other studies. The first research question of this study asked "Do *rehearsal* pre-task planning and HOQs instruction have any meaningful effect on accuracy and lexical and grammatical complexity of intermediate EFL learners' oral performance?"

Results of this investigation revealed a meaningful difference between the rehearsal group and the control group in terms of all three measures of oral performance. Also, the participants in the rehearsal group were found to have outperformed both the strategic and no planning groups in terms of the accuracy measure. In view of that, we maintain that the rehearsal task planning integrated with HOQs leads to accuracy, lexical complexity and grammatical complexity in L2 oral production. Bearing in mind the relationship between rehearsal planning condition and accuracy of L2 oral performance, the findings of this study are in line with Qin's (2019). He highlights the crucial role of rehearsal in pre- task planning which enables learners to gain control over accurate use of particular structures attempted in earlier repeated attempts. In fact, meaning-focused activities are supposed to direct learners' attention mainly toward meaning (Ellis, 2003). Therefore, based on the methodological principles of TBLT, it is vital to embed a focus-on-form component in our teaching methodology (Long, 2015). According to the results of this study, one way to achieve this aim is via provision of feedback both by peers as well as teachers. Furthermore, regarding the provision of oral feedback merely for rehearsal planners, which necessarily aims at fulfilling the focus-on-form aspect of TBLT (Ellis, 2009), and according to previous research outcomes (Chu, 2011), it seems to be safe to conclude that oral feedback delivered to rehearsal planners is likely to have a great effect on oral accuracy as it expedites the learners' self-repair and enables them to reanalyze, revise, and rectify their errors by directing their attention to form through interaction. We can further argue that rehearsal planners' improved accurate performance may be partly due to the element of task familiarity gained via practice on the narrative task presented with well-ordered sequence of pictures.

This observation is in line with that of Garcia-Ponce et al. (2018), proposing that the dearth of acquaintance with narrative tasks imposes greater processing load on students which in turn demands more attention from the students to perform the task and obliges them to focus more on fluency and complexity than on accuracy in task accomplishment.

Regarding the lexical complexity measure of oral production and considering the obtained results, the combined treatment proved to be beneficial for rehearsal planners as the post-intervention results demonstrated that they managed to produce lexically complex utterances. This enhancement can be justified by the fact that while performing the target task in reality, rehearsal planners had a second opportunity to reproduce their task. This could have led to their greater lexical complexity as they could use some words utilized in their first attempt. This inference is in alignment with Ansarin and Bayazidi's (2016) study, stressing that lexis, once coming upon and practiced via task rehearsal/performance, can be easily recalled in latter encounters. Additionally, in line with the findings of Soodmand Afshar and Rahimi (2014), to improve

learners' speaking ability, instructors should prompt their thinking skills in speaking classes by posing questions and giving challenging speaking tasks, so that they provide the learners with further opportunities to communicate. Accordingly, oral questioning to provoke thinking skills has proved to be an effective way to stimulate learners' willingness to participate in oral tasks. With respect to grammatical complexity, the current results provide evidence for an advantageous effect of RP plus HOQs training on the complexity of foreign language learners' oral production. This is in line with a number of preceding studies regarding the positive impact of planning time on grammatical complexity (Guará-Tavares, 2011; Sangarun, 2005). The rehearsal pre-task planning coupled with questioning led the participants to employ a wide range of various clauses and T-units to convey the intended meaning, which assisted them to produce grammatically complex utterances and outperform the subjects of other groups. In short, the concurrent advancement across the three dependent variables in the rehearsal planners' performance (interestingly including accuracy as opposed to the mainstream research outcomes) may be attributed to the measures taken to counterbalance the learners' predetermined tendency for focusing on meaning. Due to the communicative essence of picture-based narrative tasks, planners are mainly oriented toward a focus on meaning (Skehan, 1998). Considering the obtained results, however, compensatory focus-on-form strategies adopted for rehearsal planners proved to be fruitful in setting an equilibrium between focus-on-form/meaning dichotomy. All in all, the rehearsal planners' success can be justified in terms of their involvement in a type of double thinking practice entailing planning and acting on tasks both via the allocated task planning time and HOQs instruction. This aligns with the findings of Moseley et al., (2004) who suggested that 'thinking skills' entail actual capability in the processes of thinking including planning what to do and say as well as generating new ideas.

The second research question of the study asked "Do *strategic* pre-task planning and HOQs instructions have any statistically significant effect on accuracy and lexical, and grammatical complexity of intermediate EFL learners' oral production?" The results of the current study revealed that, whereas strategic planning combined with HOQs did not bring about any advancement in the participants' L2 oral production in terms of accuracy, it proved to be beneficial in creating more lexically as well as grammatically complex oral production. The outcomes vis-a-vis accuracy and complexity are in accordance with the findings of preceding studies pointing to the strong effect of engaging in pre-task strategic planning on complexity but not accuracy of oral production (Saeedi, 2013). This inquiry, being in agreement with Ellis and Yuan's (2004) study, proposes that SP has a progressive impact on the complexity of the oral output. In terms of accuracy, calculated by number of error-free clauses, the reported results indicated no significant effect on the part of SP plus HOQs teaching. In support of these claims, a number of studies (Gilabert, 2007; Wendel, 1997) have found that strategic planning had no effect on accuracy of oral task performance. One probable justification for strategic planners' lack of success in producing accurate utterances is that, unlike rehearsal planners, they had no chance to actually perform the task prior to the main performance unless they thought about the

possible content of what they would say in their later performance. In view of that, they could not practice, revise or rectify their structural and linguistic forms. Accordingly, they were given no mentoring comment on their performance. These findings highlight the importance of a principal variable in elucidating how strategic planning influences task performance which concerns the focus of learners' attention during planning, specifically whether it is on form, meaning or a combination of the two. However, the mainstream task-based research challenges the dichotomy of 'form' and 'meaning', arguing that what matters is attention to 'form-in-meaning' (Ortega, 1999) and having a primary focus on meaning (Ellis, 2003), suggesting that this is precisely what strategic planning helps learners to achieve when they perform a task.

Conclusion and Implications

This study investigated the potential of task planning for executing a speaking course in which pre-task planning informed by the principles of task-based instruction is combined with the integrated questioning for Iranian EFL learners task performance. Generally, the efficiency of intervention in rehearsal and strategic planner groups may be attributed to an interaction between methodological codes and the underlying practices implemented in the speaking classrooms as both have undergone fundamental changes.

Accordingly, a successful TBLT calls for radical adjustments in terms of theoretical assumptions underpinning the approach as well as the methodological practices to be implemented. Therefore, a main concern in TBI is to uncover how feasible it is to settle the form-meaning tension and set a balance between various areas of performance (Ellis, 2009). One way to attain this goal is to create conditions for focus-on-form achievement within a focus-on-meaning context. This is mainly done via adopting compensatory strategies including direct teaching of some structures as well as deploying supportive strategies such as scaffolding along with corrective feedback of both explicit and implicit nature (Brown, 2008; Ellis, 2008)

Hence, in view of that and in line with Brown's (2008) list of compensatory strategies for effective teaching of speaking, provision of corrective feedback proved to be beneficial in that it helped learners in the rehearsal planning group to rectify their errors committed in initial performance which in turn led to production of more accurate utterances in their later task performance. Furthermore, a beneficial implementation of TBI urges alterations in teachers' as well as learners' roles. Teacher's role within a task-based instruction milieu entails a demanding role targeted at deciding on, adjusting, and designing tasks in a way to meet the requirements of EFL learners (Willis & Willis, 2008).

Similarly, the teacher not only monitors the learners' task implementation but also directs the sequences of performance and assists students when necessary (Van den Branden, 2016). Additionally, the teacher employs an array of form-focusing strategies including explicit and implicit feedback along with scaffolding techniques to enrich the focus-on-form portion of TBI. In fact, the

teacher's role has been modified from an instructor to a creator of learning circumstances, a supporter, a director, and an advocate of the learning process. Regarding the learners' role and drawing on the challenging nature of oral narrative tasks, learners are supposed to play a more active role as they are expected to accomplish tasks for which they may not have the comprehensive essential linguistic resources. This last outcome is also vindicated by Nunan (2004). In short, task achievement in TBI setting requires a dynamic participation on the part of learners. To conclude, in line with Iizuka's (2019) finding, to better know and overcome the methodological impediments associated with TBLT, we should bear in mind that TBI should not be viewed as a single-size-fits-all method and cutting corners for instructors and students is a wrong policy. On the contrary, an effective TBLT methodology needs to be tailored to particular educational settings and goals in terms of both teacher and learners' roles.

Moreover, whereas the mainstream research on task performance predominantly lingers on the complexity, accuracy, and fluency triple and is still yielding mixed outcomes, we advance the proposal that it is ideal to sharpen up our measurement indexing via embedding a lexical component thereby we are capable to enrich the precision of our measurement scale to better differentiate task language assessment, which in turn enhances the validity of our task performance measurement. Another point to make here is that we have a more vigorous outlook concerning the impact of task planning as a result of traversing beyond the characteristics covered in preceding studies within the TBI avenue of research as we have examined planning conditions in conjunction with another influential factor, i. e. thinking skills, an amalgamation which proved to be drastically beneficial in that it yielded simultaneous outperformance in all areas of task implementation. To sum up, in pursuance of Hawks' (2019) suggestion, EFL instructors are recommended to adopt a more flexible view of implementing TBLT and take on more practical and context-specific ways out of difficulties they may encounter as practitioners of TBI.

Suggestions for Further Research

This investigation reported on the design and implementation of an integrated TBLT that merged classroom instruction with higher order thinking skills for oral tasks in a speaking class, for intermediate students in an Iranian EFL context. This application was found to be beneficial in enhancing students' speaking abilities by offering an innovative learning experience to students who were able to engage in accomplishing oral narrative tasks via meaningful interaction, and improve in three areas of oral performance including accuracy and lexical, and grammatical complexity. However, given that the study is small-scale conducted with only 45 EFL learners, and that the context is unique, what is documented here is not generalizable. Therefore, conducting further research into TBLT incorporating the higher order thinking skills with a larger number of students, with diverse planning lengths, additional task varieties, and with participants at other proficiency levels might be more influential. To be able to de-

- Ellis, R. (2008). A typology of written corrective feedback types. *ELT J*, 63(2) 97-107.
- 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.
- Ennis, R. H. (1987). Taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. S. Sternberg (Eds.). *Teaching thinking skills: Theory and practice* (pp.9-26). W. H. Freeman.
- Field, A. (2009). *Discovering statistics using SPSS*. 3rd Ed. Sage.
- 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.
- Garcia-Ponce, E. E. (2017). Effects of task characteristics on learners' fluency, complexity and accuracy during EFL interactions: Implications for speaking practice. *The Journal of Language Learning and Teaching*, 7(2), 70-89.
- García-Ponce, E. E., Mora-Pablo, I., Lengeling, M. M., & Crawford, T. (2018). Task design characteristics and EFL learners' complexity, accuracy and fluency during uncontrolled pair interactions: A naturalistic perspective. *Iranian Journal of Language Teaching Research*, 6(1), 75-92.
- Gilbert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45(3), 215-240.
- Guará-Tavares, M. (2011). Pre-task planning, working memory capacity, and L2 speech performance. *Organon, Porto Alegre*, 51, 245-266.
- Hashim, S., & Yaakub, R. (2004). *Psikologi pembelajaran dan personaliti*. PTS Publication Sdn Bhd.
- Hawks, M. (2019, 01). Review of the book *Reflections on task-based language teaching* by R. Ellis. *Elsevier System*. <https://doi.org/10.1016/j.system.2019.01.011>.
- Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Longman.
- Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of research in science teaching*, 42(7), 791-806.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (No. 3). National Council of Teachers of English.
- Iizuka, T. (2019). Task-based needs analysis: Identifying communicative needs for study abroad students in Japan. *System*, 80, 134-142.
- Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language learning*, 51(3), 401-436.
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307-329.
- Karatas, H., Alci, B., Bademcioglu, M., & Ergin, A. (2016). An investigation into university students' foreign language speaking anxiety. *Procedia - Social and Behavioral Sciences* 232, 382-388.
- Ketabi, S., Zabihi, R., & Ghadiri, M. (2012). Critical thinking across the ELT curriculum: A mixed methods approach to analyzing L2 teachers' attitudes towards critical thinking instruction. *International Journal of Research Studies in Education*, 2(3), 15-24.

- Sanavi, R. V., & Tarighat, S. (2014). Critical thinking and speaking proficiency: A mixed-method study. *Theory and Practice in Language Studies*, 4(1), 79-88.
- Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.111-141). Benjamins.
- Shabani, M. B., & Ghasemi, A. (2014). The effect of task-based language teaching (TBLT) and content-based language teaching (CBLT) on the Iranian intermediate ESP learners' reading comprehension. *Procedia-Social and Behavioral Sciences*, 98, 1713-1721.
- Skehan, P. (1998). *A Cognitive approach to language learning*. Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, 30(4), 510-532.
- 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. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning & teaching* (pp. 207-226). KVAB.
- Soodmand Afshar, H., & Rahimi, M. (2014). The relationship among critical thinking, emotional intelligence, and speaking abilities of Iranian EFL learners. *Procedia-Social and Behavioral Sciences*, 136, 75-79.
- Swartz, R. J., Costa, A., Kallick, B., Beyer, B., & Reagan, R. (2007). *Thinking-based learning: Activating students' potential*. Christopher-Gordon Publishers.
- Swartz, R. J., & Parks, D. (1994). *Infusing the teaching of critical and creative thinking in elementary instruction*. Critical Thinking Books & Software.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439-473.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance* (pp. 239-273). John Benjamins Publishing Company.
- Van den Branden, K. (2016). The role of teachers in task-based language education. *Annual Review of Applied Linguistics*, 36, 164-181.
- Wendel, J. (1997). *Planning and second language narrative production* (Unpublished Ph.D. thesis). Temple University, Japan.
- Wilensky, W. W. (1991). *Questioning Skills for teachers*. (3rd ed.). National Education Association of the United States.
- Willis, D., & Willis, J. (2008). *Doing task-based teaching*. Oxford University Press.
- Wu, K. Y. (1993). Classroom interaction and teacher questions revisited. *RELC journal*, 24(2), 49-68.
- Yee, M. H., Yunos, J. M., Othman, W., Hassan, R., Tee, T. K., & Mohamad, M. M. (2015). Disparity of learning styles and higher order thinking skills among technical students. *Procedia-Social and Behavioral Sciences*, 204, 143-152.
- 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.
- Zhang, S. (2009). The role of input, interaction and output in the development of oral fluency. *English Language Teaching*, 2(4), 91-100.
- Zohar, A., & Cohen, A. (2016). Large scale implementation of higher order thinking (HOT) in civic education: The interplay of policy, politics, pedagogical leadership and detailed pedagogical planning. *Thinking Skills and Creativity*, 21, 85-96.