

Fostering Speaking via Integrated Questioning: Does it Count for Transcontextual Transferability?

Mohammadreza Mollahosseiny¹, Davood Mashhadi Heidar^{2*}

^{1,2} Department of English, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran

*Corresponding author: davoodm_tarbiatmodares@yahoo.com

(Received: 2021/5/26; Accepted: 2021/8/1)

Online publication: 2021/8/21

Abstract

Given the significance and problematic nature of speaking in EFL instruction as well as lack of orthodoxy of views on acceptance or rejection of language learning transferability, this study proposed an inventive model to teaching thinking via questioning as a way to trigger speaking. It comprises six components, and organizes a rich interactive teaching/learning environment to gauge its usefulness in the development of EFL learners' oral skill as well as its inter-domain transferability effects. To this end, we compared the performance of 60 participants on pre-post-treatment oral narrative tasks over thirteen treatment sessions with a time allocation of 45 minutes for each session: One group with isolated questioning instruction and the other with integrated one. The results of independent samples t-test indicated that instructional treatment assisted the experimental group participants to outperform the control group learners solely in terms of measures of lexical and grammatical complexity. It is also suggested that the merits of questioning intervention transfer to learners' performance on a delayed task in a new social domain. As proposed, implementing the integrated approach of HOQs assisted EFL learners to successfully accomplish demanding oral tasks initially in academic setting and in later delayed novel transcontextual settings.

Keywords: integrated questioning, speaking, transferability, complexity, accuracy

Introduction

Given the importance of speaking in EFL instruction, several pedagogical methods including audio-lingual and communicative language teaching unanimously highlight the spoken language as an ambition for most of learners in various EFL settings (Diaab, 2016). However, speaking ability has been regarded as a demanding skill in EFL learning (Dörnyei, 2001). Amid various causes for this issue, one way to deal with the problem is to examine the link between speaking and thinking skills which has been an ongoing concern in the field of linguistics (Dipper, Black, & Bryan, 2005). As an essential requirement to effective learning, the educational process in the modern era calls for the advancement of higher order thinking abilities as one of the paramount goals (Boa, Wattanatorn, & Tagong, 2018). Higher order thinking skills (HOTS), which prevailed the scientific and technical fields, are now intensely emphasized in the language courses (Jaganathan & Subramaniam, 2016).

Whereas development of students' thinking skills has been a key educational concern (Zohar & Schwartz, 2005), little substantial knowledge is available on effective instructional approaches to higher level thinking skills (Tsui, 2002). Delving into the extant research corpus on thinking skills pedagogy (Zohar, 2013; Zohar & Schwartz, 2005) leads us to the dichotomy of integrated vs. isolated-subject methodologies that involves either incorporating or separating HOTS into/from the subject matter to be taught, stressing HOTS as domain-specific or general strategy, respectively. Most of the studies conducted on HOTS drew on Bloom's (1956) taxonomy to structure classroom activities and foster thinking skills from basic knowledge recall, lower order thinking skills including comprehension, knowledge, and application all the way through to higher order thinking skills entailing synthesis, analysis, and evaluation. It is generally believed that each behaviour has to be grasped before the next one can come about (Collins, 2014), that is, lower order thinking is a requirement in the course of employing higher order thinking. 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. Similarly, whereas educational policy guidelines from all over the world emphasize the necessity of teaching higher order thinking skills (HOTS) as the pivotal and pillar component of 21st century skills in several disciplines (Pellegrino, 2012), unfortunately teachers and learners in Iranian EFL setting seem to be lagging behind (Ketabi, Zabihi, & Ghadiri, 2012). Just like the circumstances depicted earlier by Zohar (2013), in Iran also, lower-order thinking, instead of HOTS, still dominates instructional methodologies and learning outcomes. On the other hand, transfer of knowledge to other contexts/ domains as a significant determinant factor in foreign language achievement has appealed the attention of many investigators (Benander & Lightner, 2005; Graff, 2010; Perkins & Salomon, 1996). Transfer of learning as defined by Perkins and Salomon (1996) denotes the capability of conveying what one has learned in a particular situation to other new situations. One method to inspire students' prior knowledge is via questioning (Almeida, 2010). Inquiry on the prominence of questioning, as a teaching and learning instrument, is well evidenced (Almeida, 2010, 2012; Chin & Osborne, 2008). Several authors advocate the view that the improvement of the learners' questioning is likely to develop thinking skills of higher order (Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005). Higher order questioning is used as a tool for activating higher order thinking skills (Chen, 2016). However, asking higher-level questions is not a habitual practice of instructors and learners (Almeida, 2012).

By the same token, Zohar and Dori (2003) emphasized the significance of transferring HOTS across disciplines as the eventual objective of thinking pedagogy. However, transfer of thinking skills from the classroom setting to other domains proved to be challenging. One reason for such difficulty is that thinking skills are positioned in educational programs as a separate objective in parallel with typical language content (Swartz & McGuinness, 2014). Moreover, regarding the impact of higher order questions on speaking in Iran, it seems Iranian EFL classrooms are still dominantly described by knowledge transmission education through memorization and focusing on lower-order thinking skills (Vahdani Sanavi, 2014). In light of the above-mentioned discussion, this study will seek to adopt Bloom's

(1956) taxonomy as the theoretical framework in order to uncover how implementing an integrated model of HOTS instruction affects complexity and accuracy of EFL learners' oral presentation. The importance of this study first and foremost lies in its novelty to probe the influence of implementing multiple strategies for questioning including teacher's questioning, peer questioning, scaffolding, learner's questioning, and feedback as well as teaching for transfer all of which are not attempted concomitantly in any other study to date. More importantly, the outline for HOTS instruction focuses on merging thinking skills into language via higher order questions. Furthermore, transfer of learned knowledge to be applied in a new domain has been a main concern of every educational program including thinking skills instruction (Brookhart, 2010; Illeris, 2009) which has not received due attention in Iranian EFL context. Whereas there are numerous studies (Mestre, 2002) on transferability of learning, to the best knowledge of the researchers, a few, if not any, studies have been carried out on inter-domain transferability of thinking skills. Accordingly, this paper aims to extend our understanding of language learning transfer and contributes to this argument by addressing some unexplored features of the literature related to academic learning-social life interactions with a particular emphasis on language learning and transcontextual transfer of learning to novel social situations in Iranian EFL context.

Additionally, in this study, speaking is measured in terms of lexical complexity, grammatical complexity and accuracy measures. The triad of complexity, accuracy and fluency (CAF) is nowadays viewed as the qualitative dimensions of language use (Housen, Kuiken, & Vedder, 2012). Each of these notions has been operationalized in different ways.

Hence, the significant ambition of this study is to see whether or not the effects of "Inventive Intervention" transfer to the accomplishment of a novel oral task in a different domain. More specifically, responses to the subsequent questions are probed:

Q1. Does integrated approach of teaching questioning have any effect on the accuracy, lexical complexity, and grammatical complexity of intermediate EFL learners' oral production?

Q2. Do learners carry the effects of instructional treatment over to tasks in a new domain?

Method

Participants

The participants of this study were 60 (20 female and 40 male) intermediate Iranian learners enrolled in speaking courses in Shokouh institute for English as a foreign language (EFL) in Mazandaran. To ensure the homogeneity of the sample, the candidates were placed at this level depending on their scores on the Nelson English Language Tests (NELT scores 35-50). Subsequently, the applicants were randomly apportioned to one experimental group (EG) with 30 students who received an inventive integrated approach of teaching HOTS and the other 30 learners were assigned to the control group (CG) who received separate method of HOTS instruction for 13 sessions.

Instrumentation

The Nelson test of language proficiency was employed to measure the proficiency level of the participants. Nelson 300 D test was run to detect proficiency levels of learners and to verify their homogeneity. It comprised of 50 multiple-choice items, including 37 questions of structures, 7 questions of lexical and 5 questions of pronunciation. Participants were requested to answer the questions in 40 minutes. Nelson test is a common and standardized EFL proficiency test and its reliability and validity has been estimated by other researchers. Second, the instructional materials used comprised oral narrative tasks along with 5WS questioning framework hand out including taxonomy of higher order questions, with each level well-defined and exemplified. The oral tasks serving as the instructional material were categorized with characteristics of basic literature together with the real world stories and circumstances focusing on familiar topics for which the learners were asked. Additionally, three oral narrative tasks were employed for the pre-test, post-test and delayed post-test. The researcher used oral narrative tasks with various topics for pre-, post-, and delayed post-tests to control the testing effect. Additionally, to ensure that the tasks measured the same constructs, two certified IELTS teachers read the task questions and confirmed the fact that the tasks were appropriate for learners' proficiency level, measured the same constructs, and eliminated the testing effects.

Furthermore, to facilitate the participants' better performance in later novel situations, the themes of oral tasks were mainly selected from everyday life subjects to ensure interest and topic familiarity for learners. Pursuing previously conducted studies including Tavakoli and Foster (2011) and Park (2010) who formerly took on the pictures from Heaton (1975), this study employed the oral narrative task with six sets of pictures given that such a common narrative task in this line of research is supposed to assist comparison with preceding findings. All participants were asked to relate at least three sentences for each picture narrating what occurred 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 participants and would expand their linguistic properties (Ellis & Yuan, 2004).

The topic of the oral task which served as the post-test was "an ideal mate", which was selected from Chen's (2010) post-test. This pictorial-based oral task focused on directing students to use higher-order questions (analysis, synthesis and evaluation). They were requested to reply the questions concerning various criteria including physical appearance, educational degree, etc. for selecting an ideal mate.

The topic of the delayed post-treatment oral narrative task was within the domain of geopolitics focusing on some criteria including living standard, income, population, pollution, and securities to compare and contrast Iran and the neighboring countries. All learners were required to express their opinions and convince others with explanations, substantiation and instances and to reach an approval. Moreover, an audio-recorder was installed to record the participants' simultaneous oral task performance.

To score the collected data, the measures used by Ahangari and Abdi (2011) were adopted for assessing the accuracy as well as complexity of the learners' task presentation. Accuracy measurement was attained by computing the fraction of error-free clauses in the entire number of clauses created and the subsequent number was multiplied by 100. Grammatical complexity measurement was operationalized as the figure of clauses for every T-unit. In order to quantify grammatical complexity, the proportion of clauses to T-units in the contributors' narratives was calculated by dividing the figure of clauses by the number of T-units in per narrative. Lastly, lexical complexity was calculated by type-token ratio (TTR). The whole

amount of various words used (types) was divided by the entire quantity of words in the text (token) (Robinson, 1995).

HOQs Hand out

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 every 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 explain...?

3. *Application*: centered on the ability to handover knowledge from one context to another.

Qs words: What would take place if..? 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 difference between... and...?

5. *Synthesis*: concentrating on the skill to merge constituents into a comprehensible whole.

Qs words: How could you...? What would occur if...?

6. *Evaluation*: focused on the ability to decide on the significance or use of information by means of a set of values.

Qs words: Which one is superior? What is your judgment...?

Innovative Instructional Treatment

In a direct opposition to traditional approaches of teaching and challenging their trend of focusing on thinking as an isolated subject, the innovative instructional framework of the HOTS approach started with the notion of infusing thinking skills into speaking through higher-order questions. The innovative instructional treatment implemented for the EG was different from the traditional CLT approach or Audio-lingual methods

(like the one implemented for the CG) in terms of a practice of speaking. It required learners to activate their schemata and use their higher-order thinking proactively in a social context to express their thoughts and comment on others' views. This can be more cognitively demanding and it requires more efforts than the traditional methods used.

The intervention for the EG proposed a multi-stage strategy framework encompassing six principal components including teacher questioning, peer questioning, scaffolding, learner questioning, and feedback as well as teaching for transfer. Initially, teacher posed a higher-order question. For example, each participant was given a topic which mainly focused on directing learners to use higher-order questions (application, analysis, synthesis and evaluation). Students were asked to answer the higher-order questions listed on the handout and were stimulated to formulate and ask higher-order questions of their own. In addition to Bloom's work, 5Ws task was used to encourage HOQ and let students reach an agreement within a group through reasoning and argumentation. Instances of such questions were, "Why does job matter while you choose your ideal spouse?" "How does appearance affect you when you decide for your favorite mate?" Participants were required to take turns in raising HOQ; for example, one learner posed a HOQ, and the other members provided answers, commenting on others' ideas. After a fifteen- minute discussion, they needed to reach an agreement and then they talked to teacher individually.

Given the fact that higher-order questioning by itself may not provoke long and sophisticated answers (Rowe, 1986), an almost extended wait time was also added to the questioning to permit learners to hear the responses of others, understand the question, and develop and form their own ideas and answers. Next, peer questioning as a cooperative activity engaged the learners in asking one another about the targeted subject matter for which learners worked in small groups (entailing five or six participants).

The third component, the teacher's scaffolding, entailed offering learners support at the commencement of a task performance and then progressively passing on responsibility to the learners to perform the task which would be beyond their own unaided effort. For example, teacher expressed some starting parts of sentences or initial sentence patterns as "my favorite mate is...." or I like to visit Germany because.....; they completed the sentences

by expressing their own structures. As learners gradually engaged in discussions, they initiated to exercise question generation and the researcher directed learners to employ Morgan and Saxton's (2006) question stems to pose their own questions. This 5Ws task required learners to actively use their thinking skills for analysis, synthesis and evaluation. This type of task was further improved in this study by constructing a social milieu for group discussion. In order to create a circumstance for interaction, 5Ws was designed to have learners reach a consensus in a group; learners needed to reason and argue with each other.

Afterwards, the students were given feedback on their question generation to complete the feedback component of the instruction. For example, one learner said: " *I like my husband be a free job*". And teacher replied: " *oh yeah you want your future mate to be a businessman*".

Within the last phase and in an attempt to maximize the transfer effect of teaching thinking, the teacher tried to demonstrate how certain notions acquired in the classroom could be utilized in a variety of other settings. Accordingly, drawing on previous studies dealing with the question that under what circumstances inter-domain transfer objectives will be met (Perkins & Salomon, 1996), the following strategies including *explicit generalization*, and *using an analogy* were adopted. Analogies demanded learners to apply what of questioning knowledge they gained to other new domains. For example, learners may initially comprehend better how the heart works by thinking of it as a pump. For explicit generalization learners were assigned a picture-based story task that required a special solution. Based on learners' accomplishment of that task, the researcher could find the underlying principle that learners employed to complete it. Then they presented the learners with another corresponding problem that invited a similar approach to be solved. Those learners who had gained the fullest summary of the principle from the first puzzle were among those who could successfully do the second task.

Finally, in order to offer diverse rehearsal practice for learners, extensive rehearsal of questioning was exercised in variety of subject frameworks including history, biology, geography, natural sciences and historical literature. Furthermore, to ensure transcontextual learning of thinking skills

applicable in a transdisciplinary way, the researcher adopted a holistic view of the world phenomena and teaching that is focused on the significance of the whole and the interrelatedness of its constituting parts. In fact, the adopted teaching strategies included the presentation of real world problems and issues followed by hands-on-inquiry-based experiments. An attempt was made to connect science to the learners' daily life issues and go beyond the science borders by believing that science can explain many life phenomena. Accordingly, in this way many scientific, physical and biological concepts can be related to life. For example, when the teacher talks about Acids, he can refer to the PH of our blood.

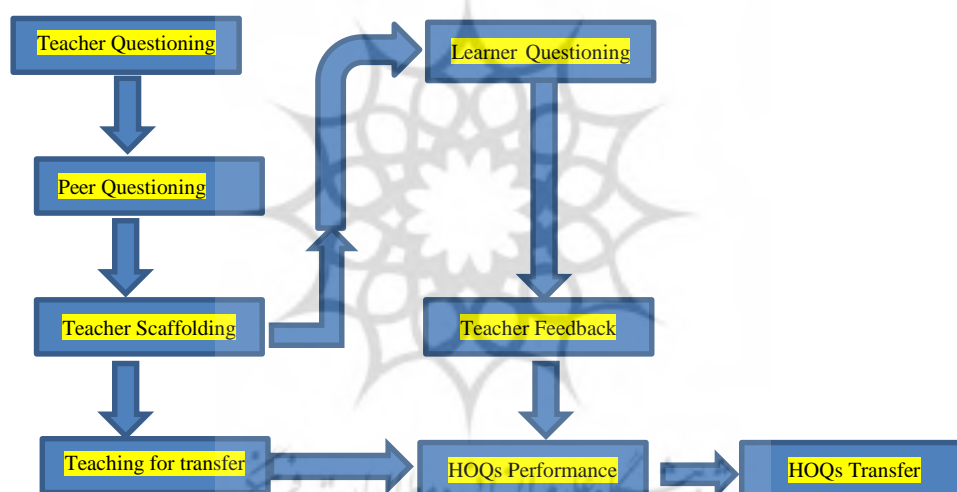


Figure 1. Six-Strategies Integrated Model of Teaching Thinking (SSIMTT)

Data Collection Procedure

The procedure for data collection entailed three phrases including a pre-test, and a post-test followed by a delayed post-test. The pretest was conducted in the first week, followed by a thirteen-session intervention. Two weeks after the intervention, the delayed posttest data were collected. The whole data were collected through oral narrative task and audio-recording from two classes. These audio-recorded data were collected only from learners' simultaneous performance on the assigned tasks. Due to some

unexpected technical problems which occurred during the recording, the total recorded run length was diverse (between 10 and 15 minutes) for participants of the two groups. Accordingly, to ensure an easier comparison, decision was made to use only the first 9 minutes of each transcript for analysis.

The learners accomplished oral performance task in their regular classroom context. The EG received the innovative method of teaching the HOTS in both a teacher-led setting and group discussion environment (see the innovative instructional treatment). The control group did not receive it and they continued to practice speaking in a teacher-led setting, which is mainly based on teacher's modeling of target language.

An example of modeling in a teacher-led milieu is described, taking the 5Ws task as an example. Initially, the teacher explained what the task was about and what learners were required to do. Learners' schemata were activated by the relevant questions concerning the topic. Then a higher-order question was posed. For an example of a discussion exercise extracted from the CG textbook, the pictures of some momentous like the pyramid of Egypt and the statue of liberty were displayed and then the learners were asked to discuss about them:

1. Discuss in which cities you can see some of these monuments or sights.
2. Which monuments have you seen or visited in real life? Which places would you like to visit? Why?
3. Which countries do you think are:
 - a. the three most interesting to visit?
 - b. the three cheapest to visit?
 - c. the three most dangerous to visit?

Finally, the two groups took the similar oral task post-test at the final phase of the course of training. Meanwhile, to gauge the transfer effect of intervention, all the participants performed on a delayed oral task following two weeks' time lapse for which they were asked to compare Iran with its neighboring countries.

Data Analysis

Having audio-recorded the participants' speech of their simultaneous performance on oral tasks, the researchers transcribed, segmented, and

analyzed the data in terms of the three production measures of accuracy, lexical complexity, and grammatical complexity. To warrant that the dissection and scoring of the transcripts were accomplished in a reliable manner, fifty percent of the data were double checked (segmented, coded and scored) by another colleague as an experienced EFL instructor with more than a decade of teaching experience. Inter-coder/inter-rater reliability coefficient amounts were beyond .93 for entire measures (with a mean of .90).

Given this high inter-rater reliability, the collected data were examined by the researcher. The raters followed the rubrics below while judging the lexical complexity, accuracy, and grammatical complexity of the participants' oral performance: Accuracy was calculated by computing the fraction of error-free clauses in the entire number of clauses created and the subsequent number was multiplied by 100 (Ahangari & Abdi, 2011). Grammatical complexity was measured by computing the average T-unit length by calculating the number of T-units and dividing them by the total number of words produced. Lastly, lexical complexity was calculated by type-token ratio (Robinson, 1995). Next, series of independent samples t-test were run to answer the research questions of the study. Additionally, for interpreting the effect size, the criterion proposed by Cohen (1988) was used (.01= small effect, .06= medium effect, and .14= large).

Results

Testing Assumptions

According to Field (2009), a number of assumptions including interval data, independence of subjects, normality, and homogeneity of variances should be fulfilled before utilizing parametric tests. The principal assumption is achieved in that the extant data are measured based on an interval scale.

The notion of *independence* of participants achieved as the presentation of any given subject is not dependent on that of other participants (Bachman, 2005). The normality of pre-test scores assumption was tested by the Kolmogorov-Smirnov test, the results of which are presented in Table 1. The pre-test scores for three measures of performance in the experimental and the control group have normal distribution given that , as reported in

Table 1, the *p-values* of the Kolmogorov-Smirnov test for the scores were above the critical value (.05), representing the existence of normality of distribution.

Table 1
Kolmogorov-Smirnov Test of Normality on the Pre-Treatment Data

Measure	Group	Kolmogorovsmirnov			ShapiroWilk		
		Statistics	df	Sig.	Statistics	df	p
Accuracy	Experimental	.144	30	.114	.961	30	.324
	Control	.142	30	.125	.958	30	.280
L.Complexity	Experimental	.118	30	.200	.953	30	.200
	Control	.166	30	.034	.946	30	.136
G.Complexity	Experimental	.171	30	.026	.918	30	.024
	Control	.185	30	.010	.920	30	.028

The last assumption concerning the homogeneity of variances should also be noticed when recording the outcomes of the inferential statistics. Levene's Test was employed for this aim. A close look at Table 2 discloses that we have not ignored the homogeneity of variance notion for the two groups' performance scores as the *p-value*, for Levene's test for three measures of performance (accuracy, lexical complexity, and grammatical complexity) are .064, .742, .372; respectively, which are greater than .05.

Table 2
Test of Homogeneity of Variances for Performance Scores in Three Groups

Measure	Levene Statistic	df	p
Accuracy	2.368	58	.064
L.Complexity	.549	58	.742
G.Complexity	3.188	58	.372

The first research question asked whether integrated approach of HOTS instruction has any effects on the accuracy, lexical complexity, and grammatical complexity of intermediate EFL learners' spoken performance. Table 3 indicates the descriptive statistics and the results of the Independent Samples t-Test for the EG and CG groups in the pretest.

Table3

Descriptive Statistics and Independent t-test for Three Measures of Performance of Pretest

Measure	Group	Mean	N	SD	T	Df	Sig(2tailed)
Accuracy	Experimental	13.60	30	1.77	.683	58	.497
	Control	13.90	30	1.62			
Lexical Complexity	Experimental	13.93	30	1.63	.161	58	.873
	Control	13.86	30	1.56			
GrammaticalComplexity	Experimental	13.82	30	1.11	.376	58	.708
	Control	13.94	30	1.28			

To measure the experimental and control groups' spoken production regarding accuracy, lexical complexity, and grammatical complexity, the collected data were submitted to statistical analysis of independent samples t-tests. As concerns R.Q.1, the subsequent results were gained from the presentation of the control and the experimental groups in the pretest.

As shown in Table 3, the data gained from the pretest pointed out a difference between the mean scores of the two groups in terms of accuracy measurement. The control group scored higher than the experimental group. The statistical analysis of the pretest results along with a comparison of groups' means indicated t observed to be .683, with the *p-value* of $0.49 > 0.05$

Hence, the variance between the two groups was not substantial at $p > .05$. Accordingly, it confirms that the two groups were not meaningfully different in terms of accuracy measure at the commencement of the study.

Table 3 represents the mean scores and standard deviations of lexical complexity measure for the experimental and the control group. The data gained from the pretest demonstrated a difference between the means of the control and the experimental group. However, the observed t was found to be .161 with the probability value of $p > .05$ which is less than the t critical (which is 2). Consequently, it is concluded that the difference between the two groups is not significant at $p < .05$.

Concerning the grammatical complexity, the mean score (see Table 3) for the learners in the experimental group is 13.82 and for the control group it is 13.94. Based on the results of the t-test, the t-observed .708 is less than the t-critical, suggesting that the difference between the two groups is not statistically significant in terms of grammatical complexity measurement. In

short, it reveals that the two groups were not considerably different in terms of three measures of oral performance at the onset of the study. Table 4 indicates the descriptive statistics and the results of the Independent Samples t-test for the EG and CG groups in the posttest.

Table4
Descriptive Statistics and Independent t-test for Three Measures of Performance of Posttest

Measure	Group	Mean	N	SD	T	Df	Sig(2tailed)
Accuracy	Experimental	15.97	30	.49	18.82	58	.000
	Control	14.00	30	.28			
Lexical Complexity	Experimental	15.89	30	.44	1.55	58	.126
	Control	15.71	30	.42			
GrammaticalComplexity	Experimental	16.40	30	.60	21.29	58	.000
	Control	14.02	30	.09			

As Table 4 demonstrates, an obvious difference is found between the posttest mean scores of the experimental and the control groups in terms of accuracy measure. The statistical analysis of the posttest results along with a comparison of the groups' means revealed the t observed to be 18.82, with the probability level of $.000 < .05$, which is much greater than t critical. Additionally, the estimation of the effect size indicated a large difference between the two groups as the effect size was .85. This indicates that there is a meaningful difference between the control and the experimental group in terms of accuracy measure.

As Table 4 shows, regarding the lexical complexity measure, the experimental group scored higher than the control group. But the analysis of the posttest results and a comparison of the groups' means indicated t observed to be 1.55, with probability value of $0.12 > 0.05$. It is quite obvious that the value of t observed does not exceed the value of t critical. Accordingly, the difference between the two groups was not substantial at $p < .05$. In fact, no meaningful difference is evident between the experimental and the control groups in terms of lexical complexity measure.

Concerning the grammatical complexity measure of oral presentation of the two groups as displayed in Table 4, we find that the participants in the experimental group performed better than the control group (CG) learners

since the mean score of EG is 16.40 but the mean score of the CG is 14.02. The statistical analysis of t-test revealed the *t* observed to be 21.29 at a probability level of $0.000 < 0.05$, which is much greater than *t* critical. The effect size was also calculated and found to be .88, which is a large effect with respect to Cohen's guideline for interpreting the effect size. Therefore, it is confirmed that there is a meaningful difference between the experimental and control groups in terms of grammatical complexity measure.

To summarize, the results of the analysis gained from the independent samples t-tests revealed that the experimental group performed better than the control group regarding the accuracy and grammatical complexity measures of spoken performance. However, regarding the lexical complexity of the learners' oral production, even though the experimental group's lexical complexity was better than that of the control group, this variance was not statistically substantial. In line with the results of the study, the first and the third components of the first research question are answered positively; however, the second one is answered negatively. The second research question probed whether EFL learners are capable of transferring their knowledge of HOQs to their later task performance in a new domain. To answer this question, an independent samples t-test was run to measure all learners' performance on a delayed oral posttest. Table 5 below demonstrates the descriptive statistics and the results of independent samples t-test for the scores gained from the participants' performance on the delayed post treatment oral narrative test.

Table5

Descriptive Statistics and Independent Samples t-test for Three Measures of Performance

Measure	Group	Mean	N	SD	<i>T</i>	Df	Sig(2tailed)
Accuracy	Experimental	14.31	30	.97	1.84	58	.070
	Control	13.98	30	.80			
L.Complexity	Experimental	15.45	30	.44	3.24	58	.002
	Control	14.65	30	.42			
G.Complexity	Experimental	16.10	30	.60	15.82	58	.000
	Control	13.86	30	.09			

As shown in Table 5, regarding the accuracy measurement, the data gained from the delayed posttest showed a difference between the means of the experimental and the control group. Nevertheless, the t-test analysis of the means of the two groups indicated the t observed to be 1.84, with probability value of $0.07 > 0.05$, that is less than t critical. Thus, drawing on this data analysis, the variance between two groups is not meaningful at $p < .05$.

As for the lexical complexity measure of the participants' performance, the statistical analysis of t-test indicated the t observed to be 3.24 at a probability level of $0.002 < 0.05$, which is greater than t critical. The effect size estimation (.15) also revealed a large effect. Therefore, it is confirmed that there is a substantial difference between the experimental and the control group.

Regarding the grammatical complexity measure, the analysis of the results of the delayed posttest along with a comparison of the groups' means indicated the t observed to be 15.82, with the probability level of $0.000 < .05$, which is much greater than t critical. The effect size was also calculated and found to be .81 which is a large effect. It denotes that there is a substantial variance between the experimental and the control groups.

In short, the results of the inferential statistics demonstrated in Table 5 indicate statistically meaningful differences between the experimental and the control groups concerning measures of lexical complexity and grammatical complexity; however, the variances between the two groups concerning accuracy measure were not statistically substantial. Consequently, as the obtained outcomes propose, as far as lexical and grammatical complexities are regarded, the benefits of innovative integrated HOQs instruction are carried to the presentation of a novel task in a different domain. However, this is not the case for accuracy measure. Accordingly, the outcomes of the independent samples t-test displayed in Table 5 confirm the second hypothesis conjecturing that Iranian EFL learners can positively transfer their thinking skills knowledge to their task performance in other domains on later occasion.

Discussion

The current study attempted to examine the way manipulating the inventive integrated model of teaching HOTS affects L2 oral performance and whether the effects of instructional treatment carry over to tasks in a new domain or not. Regarding the first research question posed, probing the impact of experimental treatment on accuracy and complexity of speech, it was found that the inventive group outperformed the control group learners in a number of ways. The most obvious variance between the control and the experimental groups was regarding the quantity of error-free clauses produced as well as a wider range of various clauses and t-units employed to convey the intended meaning which assisted them to produce more grammatically complex and accurate utterances than those of the control group.

Based on the findings of the present study, one plausible interpretation is that the inventive instruction may cover the accuracy concern in the foreign language oral performance as it entailed a satisfactory post-question wait-time which offers learners an additional chance to express their ideas and revise their output as well. This interpretation concurs with Brown's (2001) belief. He comments that effective speaking instruction entails giving the learners practice with fluency and accuracy as well as providing opportunities for learners to talk. Furthermore, the outcomes of this study concerning the beneficial impact of HOQs on lexical complexity of learner' spoken output are consistent with those of Farooq (2007) and Shomoossi (2004) maintaining that higher-order questions resulted in more words as well as word types. With respect to grammatical complexity, the current results maintain the presence of an advantageous effect of integrated HOQs training on the complexity of participants' spoken presentation placing them in line with a number of previous studies (Nunan, 1990; Xu, 2014). The evidence reported in this study confirms the results of Chen (2016) suggesting that HOQs not only fostered the length of speakers' utterance but also the content of their utterance was much more complicated leading them to produce grammatically complex utterances.

Next, regarding the second major aim of this study, the obtained results revealed that multi-staged integrated teaching of HOTS via questioning led to better accomplishment of a novel task in terms of lexical and grammatical

complexity. Accuracy measure, in contrast, did not give the impression to be responsive to this treatment. As the learners' performance revealed, their inability to produce more accurate utterances while carrying out the task in a new domain was partly due to their restricted scope of vocabulary knowledge, specifically those terminologies relevant to novel domain. This finding aligns with the conclusions of Goh and Burns (2012) who point out that students' lack of words can hinder them from expressing themselves accurately.

All in all, drawing on the outcomes of the study, the experimental group learners' outperformance concerning learning transfer can be vindicated in terms of beneficial effects of six components of integrated questioning model. Questioning as the core component of integrated teaching whether modeled by teacher or generated by learners was shown to be effective. Noticeably, during the course of instruction, the instructor's questioning plays a key role in the teaching environment. It is an essential activity as grounds classroom dialogue to be commenced. In congruence with the results of Crowl, Kaminsky, and Podell (1997), getting learners generate their own questions during task performance has been shown to advance promotion of thinking skills. In view of that, questioning strategy is regarded as the vital skill to encourage learners to think critically (Etemadzadeh, Seifi, & Roohbakhsh Far, 2013). Similarly, peer questioning was also found to be advantageous given that it expedites knowledge construction, and stimulates thinking.

Accordingly, consistent with Belland (2014), it is proposed that peer questioning as a strand of group-based learning can facilitate students' learning and its significance can be augmented when implemented in a team-based atmosphere advocating a learner-focused instruction. In line with Wigle (1999), this goal was mainly achieved via involving all learners in collaborative small group discussion within a favorable classroom atmosphere. The advantage of collaborative working on transfer of learning is also endorsed by findings of other investigations maintaining that inter-correlation of learning as a collaborative social activity proposes a practice guide to operative transfer for any program stimulating the transfer of knowledge and skills from instruction to practice in social contexts (Tuomi-

Gröhn, 2007). However, assuming that small group discussion solely did not suffice to activate and improve thinking skills among learners (Dwee, Anthony, Salleh, Kamarulzaman, & Kadir, 2016), facilitation/assistance strategies including scaffolding and feedback were also found to be necessary. Accordingly, scaffolding constituent of inventive model proved to be helpful. Hence, as proposed by Brown and Duguid (1993), one of the greatest and effective ways to support learners to become accustomed to the burdens of new circumstances is to offer them apposite awareness and expertise that are essential to encounter and resolve multifaceted real-world problems. One such actual way to apply this notion is to plan an educational milieu that can offer a realistic environment with lifelike accomplishments and scaffolding (Oliver & Herrington, 2000).

Finally, the findings related to feedback module of integrated approach maintain feedback provision in a dialogic way can contribute to the augmentation of thinking skills. Furthermore, in their account of features of a positive transfer climate, Rouiller and Goldstein (1993) find feedback significant and believe that, together with various situational cues, it influences the transfer process. Likewise, a bulk of literature advocates that it is feasible to teach effectively for transfer. Some outlines that appeared in current transfer studies comprise such ideas as modeling practice, offering feedback and employing co-operative learning groups (Mikulecky, Albers, & Peers, 1994). Hence, whereas the extremist stance of thinking transfer holds the view that thinking skills are not transferable and advocate the domain specificity of thinking skills (Wall, 2015), drawing on the results of this study it is evidenced that thinking skills can be transferred to other domains; however, it entails particular modifications in curriculum design and pedagogic practice. In short, drawing on the teaching paradigm implemented in this study and in accordance with Miri, David, and Uri (2007), our findings advocate the view that one tactic to promote transcontextual learning of thinking skills applicable in a transdisciplinary way is to include general non-disciplinary-bounded questions, strategies, and dialogues within the instructional curriculum. Questioning technique is assumed as an appropriate way to involve the learners and pause the monotonous rhythm of only teachers speak in the class. Meanwhile, whereas transference of thinking skills via stand-alone courses proved to be

ineffective (Swartz & McGuinness, 2014), the integrated method of HOTS instruction for which teaching thinking skills and curricular content are attempted concurrently may be the best panacea to break the bond, facilitate and maximize transcontextual transfer. Lending support to the insightful claim advanced by Perkins and Salomon (2012) and pointing out that transfer of thinking skill knowledge may not happen automatically, the integrative approach has adopted some extra phases and strategies including diverse practice, analogy,.. as essential conditions for paving the road of learning transfer.

In short, research has evidenced that the enhancement of thinking skills aids students academically (Stroupe, 2006). Thinking skills are decisive to the futures of learners and the societies in which they live, and given the important nature of thinking skills, educators including instructors of second languages must stimulate explicit learning of thinking skills (Jacobs, Helke, & Renandya, 2018). Accordingly, it seems that the inclusion of HOTS in educational curriculum of institutes majoring other than language is fruitful for promotion of various knowledge and academic skills (Dilekli & Tezci, 2020). Similarly, in line with Ghanizadeh, Al-Hoorie and Jahedizadeh's (2020) suggestion, EFL instructors and teacher trainers are recommended to recognize the importance and basis of HOTS and become aware of the diverse dimensions of HOTS before they can successfully nurture these skills in their pupils.

As for the implication of this study, it seems that pedagogy for L2 speaking in Iran needs a major rethink and a real modification. As reviewed in this study, the current teaching methods do not actually provide optimal learning settings for learners to practice speaking. Learners need opportunities for authentic language use to be capable of dominating the talk and converse as in everyday life circumstances. It is the HOTS approach which can offer such a real change to develop the learning. This study has revealed how to integrate thinking and L2 speaking by using the HOTS approach.

Declaration of interest: none

References

- Ahangari, S., & Abdi, M. (2011). The effect of pre-task planning on the accuracy and complexity of Iranian EFL learners' oral performance. *Procedia - Social and Behavioral Sciences*, 29, 1950-1959. doi:<https://doi.org/10.1016/j.sbspro.2011.11.445>
- Almeida, P. A. (2010). Classroom questioning: teachers' perceptions and practices. *Procedia - Social and Behavioral Sciences*, 2(2), 305-309. doi:<https://doi.org/10.1016/j.sbspro.2010.03.015>
- Almeida, P. A. (2012). Can I ask a question? the importance of classroom questioning. *Procedia - Social and Behavioral Sciences*, 31, 634-638. doi:<https://doi.org/10.1016/j.sbspro.2011.12.116>
- Bachman, L. F. (2005). *Statistical analysis for language assessment* (2nd edn.). NY: Cambridge University Press.
- Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology*. New York, NY: Springer.
- Benander, R., & Lightner, R. (2005). Promoting transfer of learning: Connecting general education courses. *The Journal of General Education*, 54(3), 199-208.
- Bloom, B. S. (1956). *Taxonomy of education objectives: The classification of educational goals*. New York: David McKay.
- Boa, E. A., Wattanatorn, A., & Tagong, K. (2018). The development and validation of the Blended Socratic Method of Teaching (BSMT): An instructional model to enhance critical thinking skills of undergraduate business students. *Kasetsart Journal of Social Sciences*, 39(1), 81-89. doi:<https://doi.org/10.1016/j.kjss.2018.01.001>
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*: ASCD.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd edn.). White Plains, NY: Longman.
- Brown, J. S., & Duguid, P. (1993). Stolen knowledge. *Educational Technology*, 33(3), 10-15.
- Chen, M. H. (2010). *Infusing thinking skills into an L2 classroom: a case study of an innovation in a Taiwanese university*. (Doctoral dissertation). School of education, communication and language sciences, Newcastle University.
- Chen, M. H. (2016). Theoretical framework for integrating higher-order thinking into L2 speaking. *Theory and Practice in Language Studies*, 6(2), 217-226. doi: <http://dx.doi.org/10.17507/tpls.0602.01>

- Chin, C., & Osborne, J. (2008). Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, 44(1), 1-39. doi:10.1080/03057260701828101
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (Vol. 2nd). Hillsdale, NJ: Erlbaum.
- Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking. *Curriculum & Leadership Journal*, 12(14).
- Crowl, T. K., Kaminsky, S., & Podell, D. M. (1997). *Educational psychology: Windows on teaching*: Brown & Benchmark Publishers.
- Diaab, S. (2016). Role of faulty instructional methods in Libyan EFL learners' speaking difficulties. *Procedia - Social and Behavioral Sciences*, 232, 338-345. doi:https://doi.org/10.1016/j.sbspro.2016.10.032
- Dilekli, Y., & Tezci, E. (2020). A cross-cultural study: Teachers' self-efficacy beliefs for teaching thinking skills. *Thinking Skills and Creativity*, 35, 100624. doi:https://doi.org/10.1016/j.tsc.2019.100624
- Dipper, L. T., Black, M., & Bryan, K. L. (2005). Thinking for speaking and thinking for listening: The interaction of thought and language in typical and non-fluent comprehension and production. *Language and Cognitive Processes*, 20(3), 417-441. doi:10.1080/01690960444000089
- Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual review of applied linguistics*, 21(1), 43.
- Dwee, C. Y., Anthony, E. M., Salleh, B. M., Kamarulzaman, R., & Kadir, Z. A. (2016). Creating thinking classrooms: Perceptions and teaching practices of ESP practitioners. *Procedia - Social and Behavioral Sciences*, 232, 631-639. doi:https://doi.org/10.1016/j.sbspro.2016.10.087
- 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.
- Etemadzadeh, A., Seifi, S., & Roohbakhsh Far, H. (2013). The role of questioning technique in developing thinking skills: The ongoing effect on writing skill. *Procedia - Social and Behavioral Sciences*, 70, 1024-1031. doi:https://doi.org/10.1016/j.sbspro.2013.01.154
- Farooq, M. (2007). Exploring the effectiveness of spoken English classes of Japanese EFL learners. *The Journal of Liberal Arts*, 3, 35-57.
- Field, A. (2009). *Discovering statistics using spss* (3rd edn.). London: SAGE.
- Ghanizadeh, A., Al-Hoorie, A. H., & Jahedizadeh, S. (2020). *Higher order thinking skills in the language classroom: A concise guide*. Switzerland: Springer.

- Goh, C., & Burns, A. (2012). *Teaching speaking: A holistic Approach*. New York: Cambridge University Press.
- Graff, N. (2010). Teaching rhetorical analysis to promote transfer of learning. *Journal of Adolescent & Adult Literacy*, 53(5), 376-385. doi:<https://doi.org/10.1598/JAAL.53.5.3>
- Heaton, J. B. (1975). *Writing English language tests: a practical guide for teachers of English as a second or foreign language*. London: 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. doi:<https://doi.org/10.1002/tea.20072>
- Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32): John Benjamins Publishing.
- Illeris, K. (2009). Transfer of learning in the learning society: How can the barriers between different learning spaces be surmounted, and how can the gap between learning inside and outside schools be bridged? *International Journal of Lifelong Education*, 28(2), 137-148. doi:10.1080/02601370902756986
- Jacobs, G. M., Helke, T., & Renandya, W. A. (2018). Explicit inclusion of thinking skills in the learning of second languages. *LEARN Journal: Language Education and Acquisition Research Network*, 11(1), 15-25.
- Jaganathan, P., & Subramaniam, I. (2016). Incorporating higher order thinking skills in task-based learning for Malaysian undergraduates. *International Journal of Contemporary Applied Sciences*, 3(2), 274-288.
- 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.
- Mestre, J. P. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Journal of Applied Developmental Psychology*, 23(1), 9-50. doi:[https://doi.org/10.1016/S0193-3973\(01\)00101-0](https://doi.org/10.1016/S0193-3973(01)00101-0)
- Mikulecky, L., Albers, P., & Peers, M. (1994). *Literacy transfer: A review of the literature* (Technical Report TR 94-05). Retrieved from Philadelphia:
- Miri, B., David, B.-C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, 37(4), 353-369. doi:10.1007/s11165-006-9029-2

- Morgan, N., & Saxton, J. (2006). *Asking better questions*: Pembroke Publishers.
- Nunan, D. (1990). The questions teachers ask. *JALT Journal*, 12(2), 187-202.
- Oliver, R., & Herrington, J. (2000). Using situated learning as a design strategy for web-based learning. In B. Abbey (Ed.), *Instructional and cognitive impacts of web-based education* (pp. 178-191): Idea Group Inc.
- Park, S. (2010). The influence of pretask instructions and pretask planning on focus on form during Korean EFL task-based interaction. *Language Teaching Research*, 14(1), 9-26. doi:10.1177/1362168809346491
- Pellegrino, J. W., & Hilton, M. L. . (2012). *Education for life and work: developing transferable knowledge and skills in the 21st century. Committee on defining deeper learning and 21st century skills*. Washington DC: Center for Education; Division on Behavioral and Social Sciences and Education; National Research Council, National Academic Press.
- Perkins, D. N., & Salomon, G. (1996). Learning transfer. . In A. Tuijnman (Ed.), *International encyclopedia of adult education and training*. Kidlington, Oxford: Pergamon Press.
- Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist*, 47(3), 248-258. doi:10.1080/00461520.2012.693354
- Richland, L. E., & Begolli, K. N. (2016). Analogy and higher order thinking: Learning mathematics as an example. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 160-168. doi:10.1177/2372732216629795
- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99-140. doi:https://doi.org/10.1111/j.1467-1770.1995.tb00964.x
- Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly*, 4(4), 377-390. doi:https://doi.org/10.1002/hrdq.3920040408
- Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37(1), 43-50. doi:10.1177/002248718603700110
- Shomoossi, N. (2004). The effect of teachers' questioning behavior on EFL classroom interaction: A classroom research study. *The Reading Matrix*, 4(2), 96-104.

- Stroupe, R. R. (2006). Integrating critical thinking throughout ESL curricula. *TESL Reporter*, 39, 19-19.
- Swartz, R., & McGuinness, C. (2014). Developing and assessing thinking skills. *The International Baccalaureate Project*.
- Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 61(s1), 37-72. doi:<https://doi.org/10.1111/j.1467-9922.2011.00642.x>
- Tsui, L. (2002). Fostering critical thinking through effective pedagogy. *The Journal of Higher Education*, 73(6), 740-763. doi:[10.1080/00221546.2002.11777179](https://doi.org/10.1080/00221546.2002.11777179)
- Tuomi-Gröhn, T. (2007). Developmental transfer as a goal of collaboration between school and work: A case study in the training of daycare interns. *Actio: An International Journal of Human Activity Theory*, 1, 41-62.
- Vahdani Sanavi, R., & Tarighat, S. (2014). Critical thinking and speaking proficiency: A mixed-method study. *Theory and Practice in Language Studies*, 4(1), 79-87.
- Wall, T. F. (2015). The transferability of higher order cognitive skills. *Procedia - Social and Behavioral Sciences*, 174, 233-238. doi:<https://doi.org/10.1016/j.sbspro.2015.01.652>
- Wigle, S. E. (1999). High quality questioning. *The Education Digest*, 65(4), 62-63.
- Xu, S. (2014). Infusing higher order question into speaking activity: A case study of L2 learner's speaking proficiency and higher order thinking skill. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, 6(3), 96-108.
- Zohar, A. (2013). Challenges in wide scale implementation efforts to foster higher order thinking (HOT) in science education across a whole school system. *Thinking Skills and Creativity*, 10, 233-249. doi:<https://doi.org/10.1016/j.tsc.2013.06.002>
- Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, 12(2), 145-181. doi:[10.1207/S15327809JLS1202_1](https://doi.org/10.1207/S15327809JLS1202_1)
- Zohar, A., & Schwartz, N. (2005). Assessing teachers' pedagogical knowledge in the context of teaching higher order thinking. *International Journal of Science Education*, 27(13), 1595-1620. doi:[10.1080/09500690500186592](https://doi.org/10.1080/09500690500186592)

Biodata

Davood Mashhadi Heidar is an assistant professor of TEFL in the Dept. of ELT at Islamic Azad University - Tonekabon Branch in Iran. He has taught courses on Critical discourse analysis, FLA, SLA and ESP in ELT. His current research focus includes Psychology of language, Critical thinking and EFL speaking.

Mohammad Reza Mollahosseiny got his PhD in TEFL from Islamic Azad University, Tonekabon Branch. He has been engaged in teaching for ten years at Islamic Azad and SAMA Universities. He has taught courses on FLA and ESP in ELT. His research interests include speaking, critical thinking, and error feedback to second language learners.

