Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

The effect of task type on complexity, accuracy, and fluency of Iranian EFL candidates' oral production: IELTS interview test in focus

Nafiseh SAFARI VESAL[1*], Morvarid SAFARI VESAL[2], Mansoor TAVAKOLI[3]

[1]*Islamic Azad University, Najafabad Branch, Isfahan, Iran*

[2]*Islamic Azad University, Tehran Center Branch*

[3]*Isfahan University, Isfahan, Iran.*

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. This study examined the effect of task type on dimensions of complexity, accuracy and fluency (CAF) in oral production of Iranian EFL candidates in IELTS Speaking test. Forty candidates participated in this study across two groups of intermediate and advance. They were participated in a pretest in order to check their proficiency level. After pretest the main interview conducted. There are three task types in IELTS interview each of which are conducted in a separate part namely: introduction, long monologue, and discussion. During this phase of the study, oral productions were recorded and then transcribed in the next phase in order to find changes in CAF dimensions in the above mentioned tasks. According to the results of the study, task type had a significant effect ($p < 0.05$) on complexity, accuracy and fluency of oral production of IELTS candidates' speaking test. This study has theoretical and practical implications in testing as well as teaching and learning fields.

Keywords: Task type, Oral production, Complexity, Accuracy, Fluency.

1. INTRODUCTION

Doubtlessly, second or foreign language testing especially speaking assessment has been considered as one of the most complex human activities to the difficulty in administration of the related tasks and measurement instruments. Sook (2003) adds another element as a reason of difficulty and states that "there are many difficulties involved in the construction …of any speaking assessment" (p. 2). However, researchers have continued to investigate more reliable ways of measuring the learners' internal language knowledge. The related literature, in this regard, suggests significance for the examination of oral production and performance of foreign language learners. Therefore, to investigate the foreign language learners' oral performance, the essential role of a set of influencing factors remains out of question. Among these factors is task type and with it's supposed significant role in oral production. To the best of authors' knowledge a few studies have been conducted on the above mentioned factor especially in the test context. Besides, as IELTS test has recently been increasingly employed as one of the most popular measurement tools for admission purposes in academic settings, this study tried to examine the effect of task variable on EFL candidates' oral performance on IELTS speaking test in an EFL context.

2. THEORETICAL BACKGROUND

2.1 Speaking performance

*Corresponding author. *Email address*: *nafisesafarivesal@yahoo.com*

Special Issue: The Second National Conference on Applied Research in Science and Technology

Speaking performance is, in fact, near to oral production and therefore different from speaking ability. Koizumi (2005) distinguished between speaking performance and speaking ability explaining that speaking performance can be defined as "actual instances of producing oral language in real time" (the definition adapted from Mc Namara, 1996 as cited in Koizumi, 2005, p. 36) and is observable. Speaking ability, in other side, is more abstract component and can be measured through performance. Speaking performance, as a target of many researches, has been investigated in language testing along with other fields by Wigglesworth (1998) and Wigglesworth and Elder (2010) as a few to mention among others. According to Koizumi (2005) speaking performance can be considered from various perspectives:

The first one is psycholinguistic perspective from which Levelt (1989) introduced his speech production model. Levelt (1989) in his "speech planning model"(cited in Rahimpour & Notash, 2008, p. 113) argues that L1 (first language) speech production includes three organized processes of conceptualization, formulation and articulation of the intended message. Accordingly, in the area of speech production the learners need to go through these hierarchical stages. In fact, every speaker performs these processes unconsciously during normal speech. Beyond L1 processing stages, it is believed that Levelts' model of production can contribute to an explanation of the processes in second or foreign language production as a more complex phenomenon (Finardi, 2008).

Following the second viewpoint, speaking performances can be distinguished in that whether they are monologue or dialogue. If the test taker speaks about a topic without interruption, his performance will be seen as a monologue and if there is a reciprocal interaction between speaker and listener, the performance is identified as a dialogue. One should keep in mind that both monologue and dialogue or even mixed situations exist in the real contexts. We can also talk of the integration of speaking and listening, i.e., comprehension precedes speaking (Douglas, 1997; Tavakoli, 2010).

Since listening and speaking are interwoven, a question may arise here is that whether to separate performance of these skills. Douglas (1997) believed that they should be considered as one skill test. In fact test taker performs process a set of listening and speaking tasks and receives a signal score for his/her performance. However, there is a need for more related research.

Speakers' performance can also be determined according to the preparation time given to the test taker. Spontaneous speech is regarded as speaking without planning time and prepared utterances are defined as speaking with planning time which itself can be subdivided into pre-task planning (prior to speaking) and online planning (during speaking, see Ellis, 2005 for details).

Linguistic aspects of language use can also be a useful device for categorization of speaking performance in general and for assessment in particular. Special characteristics of content, fluency, accuracy, different aspects of complexity, degree of communication strategies used in speaking performance are among those linguistic aspects influencing speaking performance. Boonkit (2010a) stated a number of important factors in an effective speaking performance among which pronunciation, vocabulary, and collocations are among those playing a crucial role in fluency enhancement. Hustijn (2011) distinguished linguistic aspects of language use as more purely or core components and less or peripheral components of language.

In the current study, we focused on the complexity, accuracy, and fluency aspects of candidates' speaking performance in an oral interview without defining the planning time. However, there are some factors influencing the speakers' real performance, in general and

under investigation or test contexts, in particular. Some of these factors will be clarified in the following sections.

2.2 Assessing Speaking

Today testing oral proficiency has been considered as one of the important issues in language testing (Hartley & Sporing, 1999 as cited in Sook, 2003). Speaking tests have always been identified to be important. Yet, due to difficulties of test construction, administration and measurement instruments in assessment of speaking skill led speaking tests not to be applicable as frequently as paper-and-pencil tests until 1970s (Fulcher, 2003). Yakhabi (2010) brought the transient nature of speech, lack of interest in oral production among linguists before the 1970s as well as problems researchers faced in devising objective criteria for tests as reasons leading to the ignorance of speaking tests.

Speaking tests can be defined as eliciting and rating oral production in the framework of an oral test (adapted from Davies, Brown, Elder, Hill, Lumley, & McNamara, 1999 as cited in Koizumi, 2005). They have one of the three types of indirect, semi-direct, and direct test formats which are identified by scholars (Clark, 1979 and O'Loughlin, 1995, 2001 as cited in Qian, 2009) as either requiring the test taker to produce a monologue or to talk with an interviewer or peer(s).

Indirect tests assess those underlying abilities that test developers normally want to assess (Houghes, 2003 as cited in Qian, 2009) in an indirect way. But there is a problem with these tests in that indirect tests do not accurately measure what we want to measure. In other words, they fail to have test validity. According to Qian (2009) semi-direct tests are those test formats in which a test taker performs based on the input s/he receives from various resources (e.g., computer program, audio/ video tape, text) and his/ her performance will be recorded to be assessed and rated later by trained raters. These tests prove useful, cost-effective and valid especially in assessments that are conducted in different locations. In spite of the advantages many candidates prefer being tested in another format regarding the lack of real interaction with examiner like typical communications (Qian, 2009).

Direct tests or, in other words, live tests are the third type of test format in which test takers perform exactly based on the skill aimed to be measured (Houghes, 2003 as cited in Qain, 2009). Since results of the tests are along with those information test developers are interested in, direct test seems to have validity. In this manner examiners would find the candidates' real behavior that is so similar to their behavior in real life situations. Meanwhile, It is also noteworthy to mention that oral production entails many features regarding the various purposes of language use (e.g. vocabulary, grammar, level of formality, relation, roles, vague words and communication strategies, etc.) in need of consideration in the course of oral language assessment (see Luoma, 2004 for details).

However, construct definitions in general are among difficulties speaking tests have usually suffered from. Speech can be evaluated from several aspects (i.e., fluency, accuracy, pronunciation, intonation, level of vocabulary used in speech, communication strategies, etc.) and can be measured by different methods. Everybody knows that all of these aspects are involved in a normal conversation and separation of some aspects to bring them into consideration may result in difficulties. Yakhabi (2010) in this regard stated that "the purpose of a speaking test is to collect evidence in a systematic way (through elicitation techniques or tasks) that will support an inference about the construct as we define it from the summary of the evidence (the 'score')" (p.44). One can conclude that through a focus on the interested aspects of speech that are target of assessment and avoidance of attributing scores to irrelevant factors, providing a valid speaking test is well possible.

Assessing speaking is achieved using one of the three main methods summarized by Ellis (2003 as cited in Koizumi, 2005) as: a) Direct assessment of outcomes of task, b) Discourse analysis, c) Rating scales or external ratings. However, in assessing speaking other methods are more broadly in use and have been used in the current study as well. First or second language speaking researchers usually seek to measure the frequency of some discourse or grammatical features or attempt to measure relative dimensions of complexity, accuracy, and frequency (CAF) of spoken language (Foster, Tonkyn & Wigglesworth, 2000). Regarding the multi-componential nature of L2 proficiency and performance many scholars believed that utilization of CAF can result in achievement of their principal dimensions (Skehan, 1998; Ellis, 2003, 2008). According to Yakhabi (2010) it can be a reason for the central role occupied by these variables in many recent applied linguistics researches. Yakhabi (2010) asserted that CAF have been used both for describing L2 speakers' performance in oral and written assessment.

2.3 Studies on IELTS Speaking Test

Seedhouse and Egbert (2004) investigated 137 transcribed oral proficiency of candidates in interviews related to IELTS Speaking Test. The study analyzed the internal organization of examiner-candidate talk as an institutional variety of interaction and the interactional structures in the three areas of trouble and repair, turn-taking and sequence, and topic development were under investigation. They also focused on how examiners apply instructions and directives in interview, and how interview elicits candidates' speech behavior. The results showed that the majority of examiners follow instructions and often candidates take advantages where the examiners do not conform to the instructions. It was also mentioned that in spite of some differences in the different test parts, the interaction was characterized as highly constrained in the overall organization. In the repair area, due to the strictly specified rules, the examiners have been briefed and there existed a number of distinctive characteristics in its organization.

In 2002, Brown aimed to verify the IELTS Speaking Test descriptors on the scales by utilization of empirical evidence for the criteria. The study analyzed band descriptors and criteria to identify relevant analytic categories for each of the fluency, grammatical range and accuracy, lexical resource, and pronunciation band scales. All interviews were in a range of proficiency levels. The findings indicated that most of the measures confirmed the validity of the criteria. For all measures the standard deviations were large and relative to the existing differences between levels. In general, while all measures of one scale aimed at assessment of that scale, no single measure could lead to rating, rather the candidate's proficiency was achieved through a range of features on performance.

Weir, O'Sullivan and Horai (2003) conducted a study investigating how difficulty of the Part 2 task in the IELTS Speaking Test is affected by changes along three variables of planning time, expected response time, and scaffolded support (e.g., suggestions for content). During three phases, a set of four equivalent tasks was undertaken utilizing three mentioned variables and one stable variable. 74 candidates performed all four tasks and completed a cognitive processing questionnaire. Analysis of recorded performances indicated that unaltered version of task resulted in the highest score and there existed significant differences in responses to four tasks by three ability groups. It means that task difficulty affected differently on performance of candidates in different levels. Questionnaire analysis also confirmed the results.

Wigglesworth & Elder (2010) conducted a study exploring the relationship between planning, proficiency and task in oral modules of IELTS and that how these variables interact. Ninety candidates distributed across advanced and intermediate group took part in this study. The results did not suggest any significant difference in the performance of candidates regarding the planning time.

3. The present Study

The attempt here was made to address the following research question:

What is the effect of task type on the complexity, accuracy and fluency of EFL learners' oral production in IELTS speaking test?

Due to the rare amount of studies conducted in this respect, it is hypothesized that task type does not have any effect on candidates' complexity, accuracy and fluency of their oral performance in IELTS speaking test.

4. Method

4.1 Participants

Primarily, for the purpose of this study, 45 female candidates of IELTS test (International English Language Testing System) aged between 18 and 35 took part in the standard IELTS interview test administered in current study. All participants were Persian native speakers and M.A. English Teaching students at Islamic Azad University, Najafabad and Hamedan Branches. They also attended in IELTS Preparations classes in Esfahan and Hamedan institutes. Besides, it should be mentioned that participants had never been to any English speaking country and they were limited to the mere instructional sessions in English use. There were three criteria for choosing the participants of the study: first, concerning the proficiency level, the first interview was administered to homogenize the students' level on oral proficiency. Based on their scores in the test, a number of participants (n=3) who were in the low levels were excluded from the study. The second criterion was that they were supposed to have passed speaking courses in B.A. level. The third criterion was their participation in the second speaking interview test. So, at the end 40 students were selected to participate in the second interview. In this way, there was a guarantee in that participants were in similar proficiency levels while analyzing the data collected from the interview of the study.

4.2 Instruments

The data was designed to be gleaned through the following instruments that will be described briefly.

4.2.1 IELTS Speaking Interview

In order to determine the participants' level of oral proficiency, an IELTS test was run. As mentioned above, due to the focus of this study on the assessment of oral production, only speaking module of the test was used. IELTS interview took the form of a three-phase process adapted from Cambridge IELTS 5 (2006) that was assumed to simulate real life situations and allow the participants to show how much spoken English they can produce (Cameron & Todd, 2003).

4.2.2 IELTS Speaking Interview with different task

In the second IELTS interview with those participants whose level of oral proficiency allowed them to take part in second phase, there were three monologue tasks adapted from a study conducted by Wigglesworth and Elder (2010). In the second phase of IELTS interview all subjects were asked to perform one of the three tasks randomly including description and explanation of: 1- a subject they have studied 2- a book or movie they found interesting and 3-

an event with great influence on them. As claimed by this study, these tasks were parallel and in accordance with standards of IELTS test. Furthermore, it was claimed that the design of the study was controlled for differences between the tasks enhancing internal validity.

4.2.3 Rating Checklist of IELTS

To rate oral production of participants, the researchers and three certified IELTS examiners made use of a standard, detailed checklist in the administration of interviews and after completion of the test and recording speech productions. This checklist carefully elaborated and operationalized nine band scales based on which the rater can determine the result of speaking module of the test. The use of the check list helps IELTS raters to judge objectively.

4.2.4 Accuracy Coding Checklist

In order to count error-free T-units in the spoken production of participants a detailed checklist considering all possible grammatical errors was used to determine grammatical errors of speech and then error-free T-units or correct forms of speech were calculated. This calculation method made accuracy estimation more reliable.

4.3 Procedure

To carry out this study, these procedures were followed:

Primarily, 45 female M.A. English Teaching students at Islamic Azad University, Najafabad and Hamedan Branches took part in the study. The first interview was conducted having their oral proficiency determined by three certified examiners. Those participants whose scores allowed them to remain in the study, after two weeks in a predetermined day, took part in the second interview. Finally, post-performance rating of 40 participants was done by raters.

4.3.1 Test administration

The speaking proficiency test, as a subtest of the IELTS test, was given out through some of the standard interview sessions. The test specially prepared for the speakers of other languages. An interview was conducted in about 11 to 14 minutes and each candidate was required to go through 3 parts. Part1 lasted between 4 and 5 minutes, wherein both the interviewers and candidate introduced them. Then, the candidate talked about some general topics like her/his family/home, her/ his job/studies and interests as well as some other familiar topics. In Part 2, the candidate was required to talk about a particular topic which was description of a friend in this study. There was one minute of preparation time to make some notes if she wished. Then, speaking took place for between 1 and 2 minutes before the examiner provide the candidate with one or two rounding-off questions. In Part 3, for around 4 to 5 minutes, the interviewers and the candidate engaged in a discussion which was thematically linked to what was followed in Part 2 but with more abstract issues. All interviews were administered and recorded by one of the researchers so that more detailed points not detected by the interviewer or any breach either by interviewer or candidate could be recognized and then three certified examiners judged records for more reliability. The judgments of the researchers were not interpreted in data analysis given the point that ratings might be contaminated by the researchers' (as one of them, herself was interviewer) attitudes, as Wigglesworth and Elder (2010) did in their study. Therefore, 40 subjects were selected as the focused group. In order to prevent subjects' oral performance from the effects of task familiarity and their active memory, two weeks after administration of the first IELTS interview, subjects were asked to take part in the second interview with one of the three different tasks in part2. Subjects in the second interview went

through 3 parts of test but in part2 they randomly performed one of the equalized tasks (i.e. description of a subject, book or movie, and an event, adapted from Wigglesworth and Elder, 2010). As mentioned earlier, the time administered in part 2 is similar to the time administered in the second part of previous interview administered for determining the subjects' level of proficiency. In a way similar to other studies data collection procedure, oral L2 production of subjects in the second interview was also tape-recorded.

4.3.2 Scoring scheme

In the first phase of the study, based on the IELTS band scales, proficiency level of participants was determined by three IELTS certified examiners. The inter-rater reliability index was calculated. In the second phase of the data collection, three certified IELTS examiners rated the spoken English produced by the subjects based on the taped-records which were then transcribed. The rating scores were on a scale of 0-9 as it was in the first phase of the study. Rating scores were considered in data analysis and their inter reliability was determined. To measure oral production of subjects in three types of tasks as introduction, monologue, and discussion in terms of complexity, accuracy and fluency, the researchers and their colleague as raters made use of a defined measurement criterion more elaborated on in data analysis section. Inter-coder and inter- rater reliability indices were also determined.

4.4 Measurement of variables

To analyze the collected data, the participants' tape-recorded L2 oral productions were transcribed and coded in terms of complexity, accuracy and fluency and finally scored. The inter-coder and inter-rater reliability indices were also determined to examine the consistency of scores between the raters. Let us now turn our attention to definitions of CAF measurements for more clarifications in this part.

Accuracy is the ability to produce grammatically appropriate language in communication, regarding current level of the language speaker. Accuracy, hence, was measured through the number of error-free T-units divided by the total number of T-units (Gilabert, 2004). According to Mehrang and Rahimpour (2010), "T-units contain main clauses as well as subordinate clauses attached to or embedded in them and error-free T-units are defined as only those correct T-units in terms of grammar, syntax, vocabulary and spelling" (p.12). Meanwhile, those speakers produce language with less hesitation markers and excessive breaks or repetitions during normal speaking can be judge as fluent speakers. While fluency was calculated as a variable of the number of repetitions (as used by Kawauchi, 2000; Elder & Iwashita, 2005), complexity was measured by means of the total number of different grammatical verb forms used in the tasks (Ellis & Yuan, 2003 as cited in Ellis, 2005). Complexity could be identified as the skillful usage of language and in fact the speakers' willingness to use unfamiliar forms during communication.

The earlier mentioned procedures and instruments provided us with these types of data:

a) Three separated sets of scores each representing the complexity scores of participants' oral productions over three performed tasks. b) Three separated sets of scores each representing the accuracy score of oral performances. c) Three separated sets of scores representing the fluency level of spoken language for each participant. d) Three separate scores of total performance for each participant on complexity, accuracy, and fluency. ANOVA (analysis of variance) was run to have a statistical assessment of the gathered data.

5. RESULTS

The study focused on the influence of certain task types on the CAF measures in the EFL learners' oral productions. There were three task types: task 1, introduction (candidate and examiner introduced themselves), task 2, monologue (an individual long turn speaking) and task 3, discussion (a two-way discussion).

A brief look at table 1, clearly accounts for the task based differences in three measures of complexity, accuracy and fluency. Concerning complexity, the first and the third tasks equally contributed to the highest records (M=0.04), while the second task had the lowest mean record (M=0.03).

As to the accuracy, the first (M=0.31), third (M=0.29) and the second (M=0.15) tasks resulted in the high to lowest mean records, respectively.

In relation to the fluency, task 3 (M=0.08) led to the highest mean record. Tasks 1 (M=0.05) and 2 (M=0.04) were the next ordered mean scores in order.

To assess the above suggested effect of task, separate one-way ANOVAs were conducted with task as the factor and each of complexity, accuracy and fluency scores as dependent variables to deal with the extent of differences among the earlier mean scores.

As for the complexity, there was a statistically significant difference at the $p < 0.05$ level in complexity records over the three tasks [$F (2, 118) = 4.45$, $p=0.01$].

In addition to reaching statistical significance, the actual difference in mean scores over the three tasks was quite large. The effect size, calculated using eta squared, was 0.14. Post-hoc comparisons using Tukey HSD test indicated that the mean complexity score under task 1 (M=0.04, SD=0.18) was significantly different from the complexity record under task 2 (M=0.03, SD=0.014). Besides, there was a statistically significant difference in the mean complexity scores under task 2 and task 3 (M=0.04, SD=0.013). Meanwhile, the difference in the complexity scores over task 1 and task 3 did not reach significance. Graph 1 clearly represents the complexity records over the three different tasks.

Concerning the accuracy, another one-way between-groups analysis of variance was conducted to explore the impact of task type on the accuracy scores. There was a statistically significant difference at the $p < 0.05$ in the accuracy scores over the three implemented tasks [$F (2, 118) =21.49$, $p=0.000$].

In a way similar to complexity record, the actual difference in the accuracy mean scores over the tasks was quite large. The effect size based on the calculated eta squared was 0.37. Post-hoc comparisons using Tukey HSD test highlighted the significance of difference in the mean accuracy scores over tasks 1 (M=0.31, SD= 0.08) and 2 (M=0.16, SD=0.07) as well as 2 and 3 (M=0.30, SD=0.09). However, the difference in the mean accuracy scores between tasks 1 and 3 was not statistically significant. The above highlighted difference patterns in the mean accuracy records over the three tasks are depicted in figure 2.

As for the final fluency measure, another one-way analysis of variance was employed to account for the role of task in yielding certain fluency records. Again, there was a statistically significant difference in the fluency mean records across the three tasks [$F (2, 118) =2.18$, $p=0.000$].

Despite reaching statistical significance, the actual difference in the task based fluency records was moderate. The effect size based on the eta squared was 0.04. Post-hoc comparisons

using the Tamhane test accounted for a significant difference in the mean fluency records between tasks 1(M=0.05, SD=0.06) and 2 (M=0.04, SD=0.04), 2 and 3 (M=0.08, SD=0.07) as well as 1 and 3. All the binary differences reached statistical significance in this fluency measure, a graphical record of which is suggested in figure 3.

6. DISCUSSION

The corresponding null hypothesis as the first prediction about the research question which aimed at determining if task type had any effect on the complexity, accuracy and fluency of IELTS test candidates' oral production in the EFL context was significantly rejected. The findings revealed that task type has a significant effect on the complexity, accuracy, and fluency over three tasks of introduction, description and discussion. Based on the results, there was a significant difference in the subjects' performance under tasks 1 and 3 and their performance under task 2 with regard to complexity and accuracy. Concerning fluency, unlike the other two measures (i.e., complexity and accuracy), all three tasks were significantly different from each other and task 2 (descriptive task) involved the lowest record. Given the procedure of frequency calculation (the number of repetitions), the higher the score, the less fluent oral production would be. In this line, descriptive task involved less number of repetitions and therefore, more fluent and comprehensible speech than the other two tasks.

Regarding high complexity and accuracy and lower fluency in task 1 (introduction) and task 3 (discussion), as stated by Skehan (2009c) based on the Levelt model (1989) conceptualization stage of oral speech provides the pre-verbal message for the Formulation stage. In this stage, pre-verbal message for non-native speakers is related to small amount of lexicon words in mind in addition to their significantly less organization. Therefore, there is not appropriate condition for providing smoother parallel processing. Non-native speakers often face more repairs and replacement of words in Formulation stage and it seems that they require more effort to pass this stage. The more difficult lexical items required for pre-verbal message, the higher demands on mental lexicon would be. As a result, the automatic process of Formulation stage may face some difficulties with some degrees of effect on accuracy and fluency.

According to Skehan (2009c) despite the fact that some studies were conducted in contexts different from testing conditions, research studies such as those reviewed by Skehan (2001) and others provided generalizations such as the following which can be meaningful in testing contexts and to some extend can be observed in obtained data too:

- "Tasks with exchange of interpersonal information or tasks on concrete topics usually lead to more accurate and fluent language.
- Interactive tasks usually lead to higher accuracy and complexity.
- Tasks with clear structure enhance accuracy and fluency." (p. 511)

Given the obtained results for oral performance of candidates in task 1 (introduction) and task 3 (discussion), they were involved in an exchange of personal information and speaking on familiar topics in task 1 and interaction and discussion in task 3 improving their accuracy and complexity test scores. According to Skehan (2009c), a dialogic task has a degree of influence on performance. In an interaction while the interlocutor has the floor, another one has more time to think about and process what he will produce very soon. In other words, he has time of regrouping, preplanning and preparation of the appropriate ground for the next message. This is in fact lemma retrieval process which is given more time to work (see Levelt, 1989). Skehan (2009c) brought a note from Skehan (2009b) in this regard that one's interlocutor can provide him with useful scaffolding opportunities to perform his own task much easier.

Skehan (2009c) provided another view point on the influence of interactive task on performance, stating that having interaction with an interlocutor might make it important for the speaker to produce more precise language and avoid making errors. In other words, presence of an interlocutor in addition to requiring a performance with more precise information may result in more accuracy.

Therefore, accuracy and complexity of speech were higher in the above mentioned tasks in line with previous studies. But the reason for high number of repetitions might lie in the issue that candidates were clearly aware that they were being assessed and so in a stressful situation, they might have produced language with higher accuracy in expense of less fluency.

Regarding the second task in the current study, it was performed more fluently. Given the point that candidates in this phase of IELTS test interview were equipped with written questions and one minute for preparation, it is supposed that the candidates were sure that there was no interference with their speaking in the task. Meanwhile, they had a preparation time for the subsequent task. This time might be a cause leading them to feel relax when performing the task; hence, the decrease in the number of repetitions in their productions.

It is worth noting that as mentioned by Elder and Iwashita (2005) at this point just few of the well-known oral proficiency tests (e.g., EST, IELTS, and TOEFL) give the candidates an amount of planning time to take advantage before they perform the speaking task. This planning time is allowed just in the interest of fairness. It means that in this given time, it is expected that the candidates use their chance of absorbing presented information and be able to use them in their oral production. IELTS test as one of those well-known tests allows the candidates one minute of preparation in the second task in order to digest written input for subsequent performance. However, there is not any time allowed for other tasks. Therefore, one cannot suppose that in the second task of our study results take a form of those with planning tasks.

Regarding planning time in the interview, it was proved in a study conducted by Elder and Wigglesworth (2010) as mentioned earlier. It was concluded that differences in oral performance were not due to the presence of such variables like planning time and also planning time has little advantage on performance in Part 2 of oral module of IELTS test.

Another reason for findings related to the second task, might be the structure of the task. It means that giving written input to candidates along with instructions could make it clear what was actually required from candidates to perform, and hence, enhance fluency and accuracy to some extent.

Much attention might be paid to decreasing the number of repetitions in performing task 2 wherein in comparison with the previous task performed by candidates led their fluency of oral production take priority over other measures. Therefore, fluency of speech production became higher in expense of lower accuracy. Regardless of the degree of significance of such changes in each of the above mentioned tasks, the null hypothesis denying the effect of task type on the complexity, accuracy, and fluency of the subjects' oral performance was rejected.

7. CONCLUSION

The present study was primarily an attempt to investigate the effect of task type on complexity, accuracy, and fluency of EFL candidates' oral production and focused on IELTS interview test. The major contribution of this study is related to the existing literature, given the issue that based on the final results, task type had a significant effect on the complexity, accuracy and fluency in oral language of IELTS candidates in speaking module over three tasks

of introduction, monologue description and discussion. Meanwhile, results revealed that candidates performed introduction task and discussion task with greater variety of verb forms (complexity), fewer number of errors used per T-unit (accuracy) and less fluency (more number of repetitions). Concerning task 2, oral production records of candidates were lower in terms of complexity and accuracy but higher in the number of repetition of words. It seems that candidates faced more difficulties in performance of second task. Therefore, results obtained in this investigation may shed light on the issues related to characteristic responses of candidates' oral production and benefit those involved in development of IELTS test programs.

Clearly, there have been some limitations in need of recognition by researchers and those involved in speaking testing. Making students speak English may be influenced by affective filters (e.g. stress, shyness, etc.) or, if not well justified, may bore them. As a result, data may, to some degree, express the learners' unwillingness to keep the conversation. So there is a need for further studies in related fields prior to any generalization in the results.

REFERENCES

[1] Ahmadian, M.J., & Tavakoli, M.(2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity and fluency in EFL learners' oral production. Language Teaching Research, 15(1), 35-59.
[2] Boonkit, K.(2010). Enhancing the development of speaking skills for non-native speakers of English. Procedia Social and Behavioral Sciences, 2, 1305-1309. Retrieved September 25, 2011, from http://www.Sciencedirect.com
[3] Brown, A. (2002). Candidate discourse in the revised IELTS Speaking Test. IELTS Research Reports, 6(3), 1-19.
[4] Cambridge ESOL.(2006). Cambridge IELTS 5: Examination papers from University of Cambridge ESOL Examinations: English for Speakers of Other Languages. Cambridge: Cambridge University Press.
[5] Cameron, P., & Todd, V.(2003). The New Preper for IELTS: General Training Modules. Sydney: University of Technology.
[6] Deville, M.C., & Fulcher, G.(2003).The oral proficiency Interview: A Research Agenda, 36(4), 498-506.
[7] Douglas, D.(1997). Testing Speaking Ability in Academic Contexts: Theoretical Consideration. New Jersey: Educational Testing Service.
[8] Elder, C. & Iwashita, N.(2005).Planning for test performance; Does it make a make a difference? In Ellis, R.(E.),Planning and Task Performance in a Second Language (pp.219-238),John Benjamins, Amsterdam.
[9] Ellis, R.(2003). Task-based language learning and teaching. Oxford: Oxford University Press.
[10] Ellis, R.(Ed.).(2005).Planning and Task Performance in a Second Language. John Benjamins, Amsterdam.
[11] Ellis, R.(2008). Principles of instructed second language acquisition. CAL (Center of Applied Linguistics), Retrieved September 25, 2009, from http: // www.cal.org/ resources/digest/instructed 2nd language.html.
[12] Finardi, K.R.(2008).Effects of task repetition on L2 oral performance. Applied Linguistics, 47(1), 1-14.
[13] Foster, P., Tonkyn, A., & Wigglesworth, G.(2000). Measuring spoken language: a unit for all reasons. Applied Linguistics, 21(3), 354-375.
[14] Fulcher, G.(2003). Testing Second Language Speaking. TESL-EJ(Teaching Engligh as a Second or Foreign Language), 8(1). Retrieved Jun, 2011 from http://www.Teslej.org.ej29/r5.htm.

[15] Gilabert,R.(2004).Task complexity and L2 Narrative Oral Production. Unpublished Doctoral Dissertation. University de Barcelona, Barcelona.
[16] Hulstijn, J.H.(2011). Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment. Language Assessment Quarterly, 8, 229-249.
[17] Itkonen, T.(2010). Spoken Language Proficiency Assessment: Assessing Speaking, or Evaluation Acting?.Unpublished M.A.Thesis. English Philology Department of Modern Language, University of Helsinki.
[18] Koizumi, R.(2005). Relationship Between Productive Vocabulary Knowledge and Speaking Performance of Japanese Learners of English at the Novice Level. Unpublished Ph.D. Thesis. University of Tsukuba, Japan.
[19] Kouicem, Kh.(2010). The effect of Classroom Interaction on Developing the Learners' Speaking Skill: The case of third year LMD students of English at Constantine University. Unpublished M.A. Thesis. Mentouri University, Constantine.
[20] Luoma, S.(2004). Assessing Speaking (J.C. Alderson & L. F. Bachman, Eds.). Cambridge: Cambridge University Press.
[21] Mehrang, Z, & Rahimpour, M.(2010).The impact of task structure and planning conditions on oral performance of EFL learners. Procedia Social and Behavioral Sciences, 2, 3678-3686.
[22] Qian, D.(2009). Comparing Direct and Semi-Direct Modes for Speaking Assessment: Affective Effects on Test Takers. Language Assessment Quarterly, 6, 113-125.
[23] Rahimpour, M. & Notash, M.Y.(2008).Investigating gender and topic in influence in learners' Task –prompted oral performance. Journal of the Faculty of Letters and Humanities, University of Isfahan, 2(52), 111-123.
[24] Seedhouse, P. & Egbert, M.(2004). The interactional organization of the IELTS speaking test. IELTS Research Reports, 6(4), 1-45. Retrieved September 25, 2011, from http://www.ielts.org
[25] Skehan, P.(2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. Applied Linguistics, 30(4), 510-532.
[26] Sook, K.H.(2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. Retrieved Dec. 3, 2011, from http://www.asian-efl-journal/dec_03_sub.gl.htm(10f30).
[27] Tavakoli, M. Ahmadian, M.J. (2012). The effect of different types of plannings on oral production of morphosyntactic features. Journal of TEFL (Teaching English Language and Literature Society of Iran, 3(10), 101-134.
[28] Weir, C., O'Sullivan, B., & Horai, T.(2003). Exploring difficulty in Speaking tasks: an intra-task perspective. IELTS Research Reports, 6(5), 1-42.
[29] Wigglesworth, G.(1997). An investigation of planning time and proficiency level on oral test discourse. Language Testing, 14, 85-106.
[30] Wigglesworth, G, & Elder, C.(2010).An Investigation of the Effectiveness and Validity of Planning Time in Speaking Test Tasks. Language Assessment Quarterly, 7(1), 1-24.
[31] Yakhabi, M.(2010).Effect of Comprehensible Input and Output on the Accuracy and Complexity of Iranian EFL Learners' Oral Speech. Unpublished M.A. Thesis Islamic Azad University, Najafabad Branch.

TABLES

Table 1. Complexity, Accuracy and Fluency Mean Scores over Tasks

	Tasks	N	Minimum	Maximum	Mean	Std. Deviation
Complexity	Task 1	40	.02	.09	.04	.018
	Task 2	40	.02	.06	.03	.014
	Task 3	40	.02	.07	.04	.013
Accuracy	Task 1	40	.12	.45	.31	.084
	Task 2	40	.06	.36	.15	.072
	Task 3	40	.15	.50	.29	.089
Fluency	Task 1	40	.00	.20	.05	.062
	Task 2	40	.00	.15	.04	.048
	Task 3	40	.00	.30	.08	.076

Table 2. One-way ANOVA Results for Task Effects on Complexity

		Sum of Squares	Df	Mean Square	F	Sig.
Task-based Complexity	Between Groups	.002	2	.001	4.45	.016
	Within Groups	.014	118	.000		
	Total	.016	120			

Table 3. One-way ANOVA Results for Task Effects on Accuracy

		Sum of Squares	Df	Mean Square	F	Sig.
Task based Accuracy	Between Groups	.292	2	.146	21.49	.000
	Within Groups	.387	118	.007		
	Total	.679	120			

Table 4. One-way ANOVA Results for Task Effects on Fluency

		Sum of Squares	Df	Mean Square	F	Sig.
Task based Fluency	Between Groups	.018	2	.009	2.18	.000
	Within Groups	.230	118	.004		
	Total	.247	120			

FIGURES

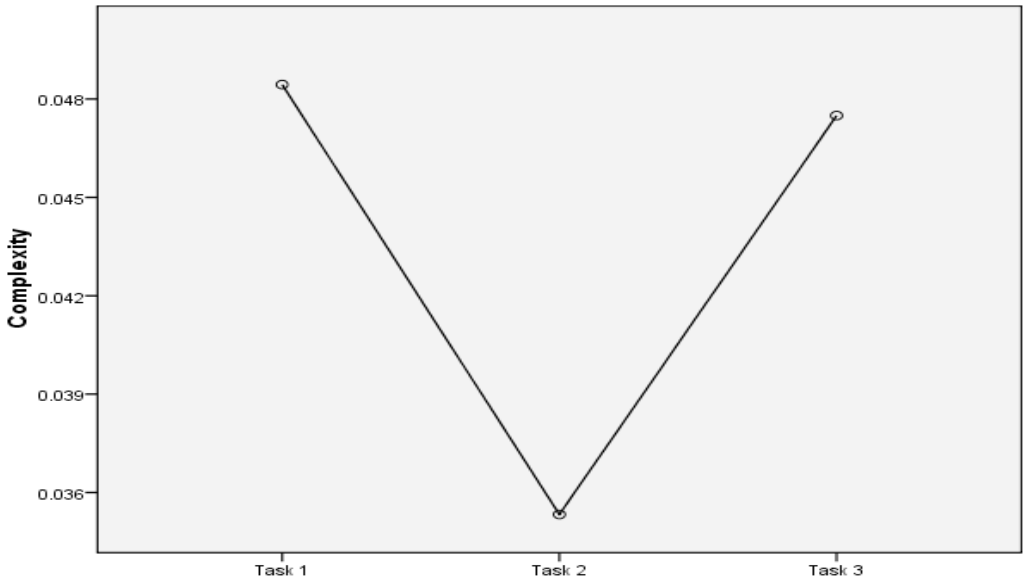

Figure 1. Task based Records of Complexity

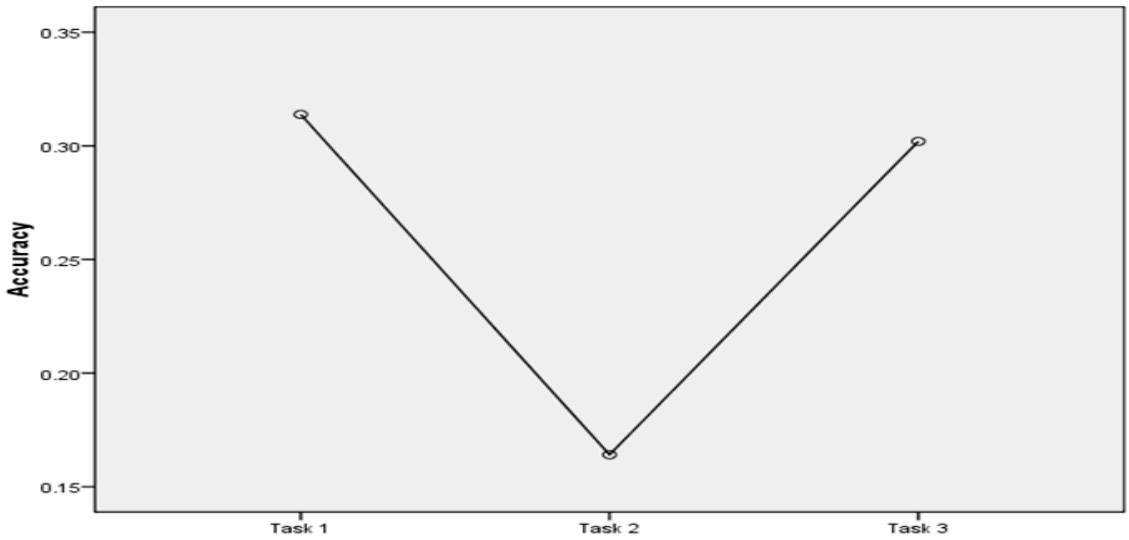

Figure 2. Task based Records of Accuracy

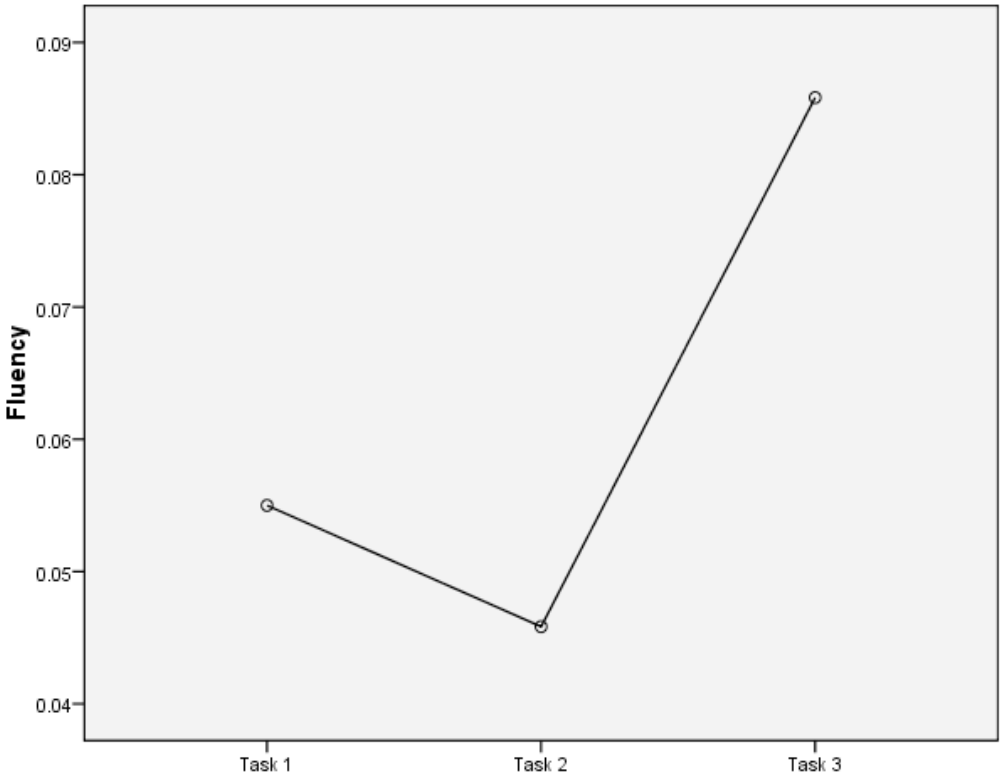

Figure 3. Task based Records of Fluency