Relationship between Second Language Deep Vocabulary Knowledge and Speaking Performance: Mediation of Task Type*

Davoud Amini (Corresponding author) *
Department of English Language and Literature, Azarbaijan Shahid Madani University.

Rana Ansari**
MA in EFL, Azarbaijan Shahid Madani University.

Bahram Behin***
Department of English Language and Literature, Azarbaijan Shahid Madani University.

Abstract

Most of the models accounting for L2 oral production have deemed a significant role for vocabulary knowledge in this process. Previous studies have demonstrated a positive relationship between different aspects of lexical knowledge and performance or proficiency of second language skills including the speaking performance. Meanwhile, the findings have suggested a determining role for the task type used for measuring speaking performance when one or more aspects of lexical knowledge are in focus. This study was conducted to investigate the relationship between the EFL Learners' deep vocabulary knowledge (DVK) and speaking performance by scrutinizing the mediating role of task type. To this end, 102 bachelor ELT students were given Word Associate Test to measure their DVK, and a planned presentation task and unplanned tasks of description, narration and reasoning to elicit speaking performance. The elicited samples of speaking performance were transcribed and analyzed in terms of fluency, accuracy, lexical complexity and grammatical complexity. Structural equation modeling indicated a lack of causal relationship between DVK and aspects of speaking performance as measured with both planned and unplanned tasks. However, mixed results were obtained in the case of the correlations of fluency, accuracy, grammatical complexity and lexical complexity with DVK across different tasks. Although the findings do not provide evidence for a strong relationship between DVK and speaking performance when DVK is analyzed in isolation from other aspects of vocabulary knowledge, the variation witnessed in findings provide further proof for the importance of task effectiveness in the study of lexical access.

Key words: Deep vocabulary knowledge, Speaking performance, Task type, planned task, Narration task, Description task, Reasoning task.

Introduction

Speaking is an outstanding performance dimension for the majority of second/foreign language (L2) learners when their success in learning L2 is evaluated (Trent, 2009). Due to the importance attached to L2 speaking, there have been some attempts to formulate overarching models of L2 speaking where the role of such language components as lexicon, grammar and formulaic expressions is elaborated. Most of these models have attached a high level of importance to learners' lexical knowledge as a key factor in speech production (Levelt, 1989, Kormos, 2006). As a consequence, a substantial amount of studies have been done to show the developmental efficacy of vocabulary knowledge and almost all of them have demonstrated the crucial effect of vocabulary knowledge on second language speaking performance (Funato & Ito, 2008; Hilton, 2008; Segalowitz & Freed, 2004; Uchihara & Saito, 2016).

Along with various dimensions of lexical knowledge, researchers have recognized the depth of vocabulary knowledge (DVK) as a plausible candidate for influencing speaking performance (Koisume, 2005; Koisume & In'nami, 2013). Stahr (2008) claimed that depth of vocabulary knowledge has an essential role in helping learners to access and activate words for both receptive and productive processing. Language users need access to many aspects of meaning while speaking. For instance, they have to make associations between concepts on mind and mental lexicon available to them. Typically, EFL learners resort to developing a passive knowledge of vocabulary which is dominantly confined to semantic knowledge of individual words rather than their syntagmatic and paradigmatic relations with other words in actual language use.

In tracking the effective role of deep vocabulary knowledge on speaking performance, the type of task used to measure speaking proficiency/performance plays a significant role (Koisumi, 2005; Safari Vesal, Safari Vesal, & Tavakoli, 2015; Teng, 2007). Skehan and Foster (1999) asserted that an area of future interest for language L2 assessment is to probe into the relevance of task features associated with

task types to the language performance at different levels and dimensions. Hence, the current study intended to probe more deeply into the relationship between deep vocabulary knowledge and speaking performance of L2 learners with due regard to planned and unplanned tasks utilized to elicit speaking performance from the L2 learners. The type of rhetorical mode is another aspect of the speaking task performance to be examined here. The speaking performance of Iranian EFL learners upon three modes of descriptive, narrative, and reasoning tasks has been scrutinized. Therefore, the present study explored the following questions:

1 Is there any relationship between deep vocabulary knowledge and speaking performance of Iranian advanced EFL learners?

2 Are there any differences between four aspects of speaking performance (fluency, accuracy, lexical complexity, and grammatical complexity) in their relationship with deep vocabulary knowledge?

3. Does the type of task used to elicit and measure spoken performance (planned vs. unplanned) make up a source of variation in the relationship between deep vocabulary knowledge and speaking performance?

4. Does the type of the rhetorical mode in unplanned oral production tasks (narrative, descriptive, and reasoning) make up a source of variation in the relationship between deep vocabulary knowledge and speaking performance?

Review of Literature

Assessing speaking performance

Testing oral proficiency has a long history of more than a century ranging from a focus on assessing phonological aspects of speech through reproductive oral skills to outcome-oriented task-tests and spontaneous performance skills (Fulcher, 2003). According to Ellis (2003), three methods used for measuring oral performance are assessing the outcome of tasks, assessing discoursal features of oral production and assessing performance with rating scales (p. 296). In the

first method, the purpose of the task is evaluated as fulfilled or unfulfilled based on "the outcome of the task" (Ellis, 2003, p. 296). The advantage of this method is that it affords an objective measurement, involving no judgment on the part of the assessor, and also it's easy and quick. Discourse analytic assessment includes the measures of accuracy, complexity and fluency. External ratings, like the direct assessment of tasks, involve the assessor observing a performance of a task and making judgment though the judgment might be relatively subjective.

According to Sandlund, Sandqvest and Nyroos (2016), oral proficiency interviews (OPI) have been the most common setups for measuring speaking performance which provide a favorable condition for evaluating accuracy, complexity and fluency of learners' speech. Fluency has been commonly defined in terms of whether production rate or freedom from dysfluency markers such as false starts, self-initiated repetitions, modifications and repairs. A common method for assessing speech fluency is rate of speech where the number of raw syllables excluding dysfluency markers is an indicator of oral fluency.

Speaking accuracy is a discoursal feature of speech production which has been used to indicate freedom from grammatically erroneous language while getting engaged in real-time communications (Ellis & Barkhuizen, 2005). Complexity likely denotes the lexical and structural variation in speaking which are referred to as lexical complexity and grammatical complexity, respectively. In many task-based studies, the 'standard' for measuring grammatical complexity is the degree of subordination in AS units. The common measure of lexical complexity which has also been adopted in this study is the type-token ratio of words (Ellis & Barkhuizen, 2005)

Relationship between DVK and speaking performance
Vocabulary has an undeniable role in learning language skills (Schmidt, 2010; Stahr, 2009). Therefore, a common research method to investigate the effect of lexical knowledge and use on aspects of second language proficiency/ performance has been correlational studies looking for the relationship between knowledge of the words and L2

skills. Despite the apparent absence of consensus regarding the definition of vocabulary knowledge (Schmidt, 2010), it has been generally conceived of in terms of breadth (size) and/or depth (quality) criteria (Read, 2000) in these studies. For example, Tahmasebi, Ghaedrahmat and Haqverdi (2013) reported a significant relationship between Iranian EFL learners' vocabulary depth and breadth and their language proficiency.

Alderson (2000) reiterated that lexical knowledge is "a strong indicator of listening, speaking and writing proficiency" (p. 35). Accordingly, breadth and depth of vocabulary has been reported as a significant predictor of reading performance (Grabe, 1991; Hunt & Beglar, 2005; Qian, 2002; Qian & Schedle, 2004; Rashidi & Khosravi, 2010). Most of these studies show that the relevance of breadth to reading comprehension is more conspicuous than depth of vocabulary knowledge (Kaivanpanah and Zandi, 2009; Li and Kirby, 2014). Besides, studies indicate a positive correlation between vocabulary depth and writing performance (e.g., Atai and Dabbagh, 2010; Baba, 2009; Dabbagh and Janebi, 2017). The enhancing effect of breadth and depth of lexical knowledge on listening comprehension of EFL learners has similarly been approved by research (e.g., Afshari, & Tavakoli, 2016; Baleghizade, & Khaledian, 2016; Stahr, 2009)

Speaking has not been an exception regarding the relationship between lexical knowledge and L2 performance. A set of studies indicate a positive correlation between breadth and depth of lexical knowledge and speaking performance. Koisume and In'nami (2013) managed to indicate that vocabulary knowledge accounts for about 84 percent of speaking proficiency including 63 % by vocabulary size and 60 % by vocabulary depth. Similarly, Uchihara and Saito (2016) demonstrated that productive vocabulary knowledge of EFL learners correlated significantly with their fluency.

Psycholinguistic models of speech production such as Levelt (1989) or Kormos (2006) testify to the significant position of accessing appropriate words in the mental lexicon in real-time processing of speech. In Levelt's three-stage delineation of oral production process,

lemmas as lexical chores make up an indispensable aspect of conceptualization of the intended message (de Bot, 1992). The facility with which the lemma of an intended word is accessed depends to a large extent on the strength of its relations with the cognitive networks of mental lexicon (Skehan, 1998). Since depth of vocabulary knowledge by definition pertains to the synagmatic and paradigmatic relations of the target words with other related words (Read, 2000), it is reasonable to conceive that deep vocabulary knowledge facilitates oral production of speech. Therefore, DVK has to be taken in to account with learners' network knowledge of words. Word Associate Test (WAT) has been the most commonly-used measure of DVK (Read, 1998).

The relevance of task type to speaking performance
Using test tasks is a common procedure for assessing actual performance of L2 learners in real-life oral communications. In Fulcher's (2003) model of speaking assessment, task characteristics play a central role in regulating what the test measures. The quality of learners' speaking performance, as a result, depends to a large extent on characteristic conditions of doing the task (Skehan, 1998; Skehan & Foster, 1996). That is why Koisume (2005) asserts that criterial measures of speaking such as fluency, accuracy and complexity cannot be considered as reflecting general capability of L2 learners as these accomplishments might be deeply affected by characteristics of the assessment task. The reason for task effectiveness is mainly related to the cognitive processes involved in each type of task performance (Skehan, 1998; Robinson, 2001).

According to Tarone (2005), there are four aspects of tasks that can determine task accomplishments. These dimensions of task performance include function, interaction, planning time and linguistic processing involved in the fulfillment of the task. One aspect of task effectiveness in oral production is presence or absence of planning time in the design of the task. Regarding planning time, there are two forms of participation in speaking performance. Learners might be allowed to

prepare for the task whether as planning prior to or as online preparation during task performance (Ellis, 2003).

In addition, the discoursal mode of speech plays an important role both in terms of content and structure of the speaking task. Brown and Yule (1983) proposed four main types of expository talk including instruction-giving, description, narration and opinion-giving as a crucial factor in determining the challenges involved in task fulfillment. Bygate (1987) also presented a similar and finer distinction between types of speaking tasks in two different categories: factual and evaluative. The first one consists of narration, description, instruction and comparison, and the second one comprises explanation, justification, prediction and decision making (Luoma, 2004).

Characteristics and types of speaking task interact with each other in a dynamic manner to ordain ultimately-assessed speaking performance. Foster & Skehan (1996) examined the mediating role of planning time in performance on three types of tasks each representing a discoursal mode, i.e., narration, decision making personal information exchange. According to their findings, the interactional relationship between planning time and discoursal mode was a source of difference in the ultimate speaking performance. Narration and decision-making tasks exhibited better enhancement as a result of planning. In a similar study, Skehan & Foster (1999) demonstrated that task structure and processing load involved in performing the task affected fluency and complexity of performance in a narrative retelling task.

The comparative performance in different types of tasks from discoursal mode point of view has been the subject of a few studies. Safari Vesal, Safari Vesal and Tavakoli (2015) probed into the relevance of task type in CAF dimensions of speaking. They indicated that learners demonstrated a higher level of accuracy and grammatical complexity but lower level of fluency in the introduction and discussion tasks when compared to description tasks. Teng (2007) reported varying FAC performances on three tasks of answering questions, presentation and picture description. Learners displayed higher

complexity and fluency for answering the questions task when compared to picture description task.

The current study intended to examine combined effects of task type and planning time in the assessment of speaking performance and the way these moderating variables might overshadow how L2 learners' vocabulary knowledge can be reflected in their speaking performance. The unplanned speaking task in the current study was aimed to elicit speech from participants of an interview in response to a set of pre-planned questions. There were sets of questions each targeting one of the three rhetorical (or discoursal) modes in speaking, i.e., narration, description and reasoning (Brown & Yule, 1983). The three kinds of "descriptive, narrative and opinion tasks imply three aspects of statics, dynamic, and abstract relationship" (Brown & Yule, 1983, p.109).

Method

Design

The current research project was a correlational study to examine any possible go-togetherness of L2 learners' deep vocabulary knowledge on the one hand and their L2 speaking performance on the other while controlling for any possible mediation of task type in eliciting oral speech from participants of the study. It is a correlational study amid developing a model through structural equation modeling (SEM) for explaining the relationship between DVK and speaking performance by mediation of task types.

Participants

The participants of this study were 102 male and female fresh students studying BA in English Language and Literature at Azarbaijan Shahid Madani University, Iran. They ranged in age from 19 to 22. The students at this level of education are considered to be at higher-intermediate to advanced level in their general English proficiency. The participants were chosen from three intact classes of an oral listening-speaking course. The participants in the three classes attended a 15-week regular listening-speaking course intended to improve their public speaking skills. The students in the three classes went through identical instructional procedures, and all the classes were taught by the same

instructor. The oral performance data required for the research was collected as part of the formal mid-term and final-exam based on the descriptions in the course syllabus.

Instruments

In this study, a paper and pencil test for measuring the depth of vocabulary knowledge alongside two types of speech elicitation tasks, an unplanned speaking task and a planned speaking task at three genre levels of narration, description and reasoning were used to collect data.

Word Associate Test

John Read's Word Associate Test (WAT) (Read, 1998) was used for measuring deep vocabulary knowledge. WAT is composed of 40 items that are intended to evaluate learners' capability in distinguishing lexical relations including collocational, synonymous and hyponymical relations between the stimulus word and eight given options. According to Zhang and Koda (2017), WAT is scored in three main scoring systems of 40, 160 and 320 points (Appendix A). In this study the 160-point system was adopted for data analysis. That is, neither the selection of distracters was penalized nor the non-selection of them was awarded. Each associate received one point.

Planned Speaking Task

The first task used to elicit speaking performance from the participants was a controlled presentation task which was administered as part of the requirements for the mid-term exam in the middle of the semester. The task involved rehearsed public speech presentation based on the instructions and model presentations the students had received during seven weeks of instruction in public speech skills. A set of TED talks had been selected by the instructor to model public speeches and familiarize learners with a set of skills and strategies involved in a successful public speech. TED talks ranged in their length from 6 to 12 minutes. Every time the classes met, a selected TED talk was played to the students for listening purposes followed by an analysis of skills and strategies involved in a high-quality public speech.

A set of instructions were delineated by the teacher to avoid mirror-copy presentations. Students were asked to use their own personalized perspectives while sticking to the same topic, but they were allowed to use parts of the content information they received in the model presentations. They were also kept committed to demonstrate as many of the public skills and strategies as they could to increase the chances of a higher score. The topics of presentations were randomly selected by the instructor for each of the learners. The participants' performances on the speaking task were audio-recorded and transcribed later for the statistical analysis.

Unplanned Speaking Task
The second task which was aimed to elicit samples of learners' oral production performance in free speech was an unplanned speaking task administered to the learners as an oral interview. The interviewees were supposed to answer orally a set of questions meant to elicit free speech and their performance was recorded for later analysis. The interview was conducted as part of the requirements for the final exam in the speaking-listening class. On the day of their final exam, the participants were interviewed by the researchers individually in private sessions. Each individual interview session lasted for about 6 minutes.

The interviewer explained to the students that they had 1 minute to prepare their answer to each question. Some of the interviewees used their one minute preparation time but the others answered questions immediately. Some of them misunderstood the questions but the interviewer had to redirect them to answer the questions based on the expected rhetorical mode. For example, when an interviewee expressed a personal opinion while a narration or description was expected, the interviewer had to stop and redirect the speaker to the expected mode.

Elicitation Questions
The oral interview was intended to elicit speaking in three rhetorical modes of description, narration and opinion (also known as discourse modes corresponding to exposition, narration and argumentation, respectively). Description directed participants to describe an individual or a place. Narration questions asked learners to talk about

an event experienced by themselves or someone else.   Opinion questions required the participants to share their beliefs or judgments about a person, thing or fact.

The questions to elicit free speech in three rhetorical modes of speaking were adopted from a set of preparation texts for free expression. First, 120 questions were made by researchers, based on two practice books: 1) Taking Sides (Amini, 2014) which is a textbook for argument-making and debating skills in free speaking 2) The Speaking Test of IELTS (Ramezanee & Hakimi, 2004) which is a preparation text for the speaking section of IELTS as elicitation for the speaking performance in three rhetorical modes of description, narration and opinion expression (40 questions for each mode). To ensure the validity of designed questions in eliciting the three intended modes of speech, 120 designed questions were given to 3 experts and they were encouraged to select 6 out of each set of 40 questions that would be best ones in their opinion, in eliciting one of the three modes of oral production of speech. Finally, a set of 18 questions were selected to be used in this study.

Due to some logistic restrictions, it did not become possible to interview and record all of the participants simultaneously. As a result two conspicuous concerns threatened the internal validity of the questions designed for measuring aspects of students' free speaking ability. First, if all interviewees were asked the same questions, their performance might be affected by practice effect since the waiting students could have consulted the finished ones about the content information required for answering the interview questions which would have privileged them compared to earlier takers of the test. Second, having totally different questions for each interviewee by having as many questions as the participants could pose another type of threat to the internal validity since it is practically hard, and somehow impossible, to ascertain a large number of different questions as being parallel and measuring the same construct. Since these two threats acted in opposite directions on the internal validity, a midway position seemed to be a reasonable choice to account for both threats. Therefore,

a set of 18 questions were coded and arranged into three baskets of 6-question sets. A coded scheme with a cyclical arrangement of questions was designed in which the 18 questions (Appendix B) were distributed in sets of 3 questions (targeting the three modes of speaking) in a way that each question was given to 6 participants, but each participant answered a different combination of the three questions. Therefore, a rotatory distribution of questions (Appendix A) was used in which an equal proportion of participants were exposed to the same question while protecting the measurement against the test effect and content validity.

Measures of Speaking Performance
The performances of the students on planned and unplanned speaking tasks were recorded and carefully transcribed. To quantify some quality aspects of participants' speech production, a set of measures common in literature for measuring accuracy, fluency, grammatical complexity and lexical complexity were utilized. The rate of errors per AS-units (Skehan & Foster, 1997), type-token ratio (Robinson, 1995), amount of subordination (Foster & Skehan, 1996) and the number of pruned syllables per minute were used as indexes of accuracy, lexical complexity, grammatical complexity and fluency, respectively.

Procedure
First of all, 102 BA students were chosen from three Listening-Speaking classes. For the planned task, students selected and prepared a TED talk speech which had been practiced throughout the oral listening-speaking course. For scrutinizing the relationship between DVK and speaking performance in unplanned task, an oral interview test inspired by IELTS speaking test was designed. All their presentations and answers to interview questions were recorded and transcribed. To achieve the inter-rater reliability, all of the recordings were transcribed by a second rater and all of the gained scores by the main and other rater were collected and the average scores for each aspect of speaking performance (fluency, accuracy, lexical complexity, and grammatical complexity) were scored by two raters, including the course instructor and another senior English teacher.

Data Analysis

In this study the correlational analysis was used for handling the data. According to these correlational relations one variable may predict another. However, this correlation does not imply causation. Therefore, a structural equation modeling (SEM) was used to study any paths on the strengths of correlations between the components of speaking performance as measured through different types of oral tasks and depth of vocabulary knowledge.

Results

Relationship between DVK and speaking performance in planned task

An assessment of the normality of data is a prerequisite for parametric statistical tests. The Kolmogorov-Smirnov test (KS) was used for this purpose. Table 1 demonstrates the descriptive data and their normality.

Table 1. One-Sample Kolmogorov- Smirnov Test for planned task

	DVK	Average Fluency	Average Accuracy	Average Gram. Comp	Average L. Comp
N	102	102	102	102	102
Normal Parameters	104.76	114.8	.1829	11.36	58.82
Mean	18.18	37.56	.1742	3.91	13.29
Std. D	.123	.77	.212	.095	.65
Most Extreme	.61	.77	.212	.95	.058
Differences	- .123	- .41	- .161	- .067	-.065
Absolute	1.241	.777	1.145	.963	.659
Positive	**.092**	**.582**	**.060**	**.311**	**.777**
Negative					
Kolmogorov-Simirnov					
Asymp. Sig. (two-tailed)					

The correlation between DVK on the one hand and fluency, accuracy, grammatical complexity and lexical complexity on the other were .108 (Sig=.282), -.094 (Sig=.346), .208 (Sig=.036) and .065 (Sig=.517) respectively. Because of the measurement used to calculate

accuracy (number of errors per A-S units) was in fact an indicator of inaccuracy, the correlation level between DVK and accuracy turned out to be negative. Out of four calculated correlations, only the correlation between DVK and grammatical complexity reached the significance level (.036) with 95 percent probability. The structural equation modeling of the relationships in planned presentation task has been presented in Figure 1.
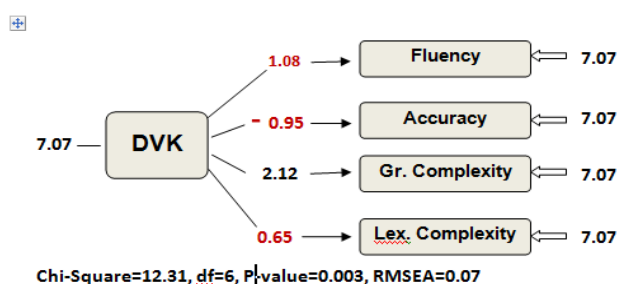

Figure 1. SEM modeling of DVK and speaking performance in planned task

The RMSEA which should be less than (≤ 0.05) is more than the significance level ($0.07 \geq 0.05$) and the P-value which should be more than (≥ 0.05) is less than the significance level ($0.003 \leq 0.05$). Therefore, there is not any causal relationship between DVK and speaking performance in general.

Relationship between DVK and performance in unplanned tasks
The result of descriptive statistics obtained through the unplanned tasks is shown in the Table 2.

Table 2. Summary scores of DVK and speaking in unplanned tasks

	Minimum	Maximum	Mean	Std. deviation
DVK	48.00	134.00	104.76	18.11
Descriptive Fluency	30.00	215.84	81.99	31.05
Descriptive Accuracy	0.00	2.14	.6465	.42
Descriptive Gram. Comp.	2.59	26.65	10.48	3.92

Descriptive Lex. Comp	39.00	135.50	63.36	12.23
Narrative Fluency	41.30	170.57	82.99	30.55
Narrative Accuracy	.00	3.97	.82	.67
Narrative Gram. Comp.	4.55	28.45	10.40	3.81
Narrative Lex. Comp.	36.00	91.50	63.77	11.37
Reasoning Fluency	38.80	186.05	87.77	32.22
Reasoning Accuracy	.00	4.65	.75	.63
Reasoning Gram. Comp	4.69	29.50	10.20	3.64
Reasoning Lex. Comp.	32.50	84.00	60.50	11.27

The test of Kolmogrov-Smirnov was used to test the normality for the scores at unplanned tasks in three kinds of descriptive, narrative and reasoning genres. As evident in Table 3, the results in all of the four tasks are greater than the significance level which is 0.05, so the normality of scores was approved.

Table 3. KS normality check for speaking performance on unplanned tasks

	Fluency	Accuracy	Gram. Comp.	Lex. Comp.
N [a. b.]	102	102	102	102
KS Z (Descriptive Task)	.875	1.032	.835	.947
Assymp. Sig. (Descriptive Task)	.429	1.032	.835	.947
KS Z (Narrative Task)	1.113	.1726	.963	.631
Assymp. Sig. (Narrative Task)	.168	.152	.311	.821
KS Z (Reasoning Task)	.777	1.145	.963	.659
Assymp. Sig. (Reasoning Task)	.582	.060	.311	.777

a. Test distribution is normal   b. Calculated from data

The correlation coefficients between DVK scores and each of the four components of the speaking performance elicited through the unplannned tasks were calculated in seeking for any statistical correlatedness.

Table 4. Correlations between DVK and speaking performance on unplanned task

		DVK
Deep Vocabulary Knowledge	Pearson Correlation	1
	Sig. (2 tailed)	
Fluency (Description Task)	Pearson Correlation	.017
	Sig. (2 tailed)	.866
Accuracy (Description Task)	Pearson Correlation	-.026
	Sig. (2 tailed)	.794
Gram. Complexity (Description Task)	Pearson Correlation	.166
	Sig. (2 tailed)	.095
Lex. Complexity (Description Task)	Pearson Correlation	.104
	Sig. (2 tailed)	.298
Fluency (Narration Task)	Pearson Correlation	.114
	Sig. (2 tailed)	.158
Accuracy (Narration Task)	Pearson Correlation	-.259**
	Sig. (2 tailed)	.009
Gram. Complexity (Narration Task)	Pearson Correlation	.224*
	Sig. (2 tailed)	.023
Lex. Complexity (Narration Task)	Pearson Correlation	.086
	Sig. (2 tailed)	.389
Fluency (Reasoning Task)	Pearson Correlation	.211*
	Sig. (2 tailed)	.034
Accuracy (Reasoning Task)	Pearson Correlation	-.222*
	Sig. (2 tailed)	.025
Gram. Complexity (Reasoning Task)	Pearson Correlation	.153
	Sig. (2 tailed)	.126
Lex. Complexity (Reasoning Task)	Pearson Correlation	.123
	Sig. (2 tailed)	.220

      * Correlation is significant at .05

      * Correlation is significant at .01

As indicated in Table 4, the correlations between DVK and fluency, accuracy, grammatical complexity and lexical complexity were .886, .794, .95, .298 respectively. Out of the four calculated correlations between DVK and aspects of perfomance in the descriptive task, none of the aspects reached the significance level and turned out to be statistically insignificant. Out of the four calculated correlations in the narration task, accuracy and grammatical complexity reached significance at 0.01 and 0.05 levels respectively. Finally, out of the four correlations between DVK and aspects of performance on the reasoning task the correlation between DVK and fluency and accuracy reached the significance level (.034) with 95 percent probability.

The structural equation models of the relationships between DVK and speaking performance on total performance on three modes of speech in unplanned tasks, description task, narrative task and reasoning task have been indicated in Figure 2.
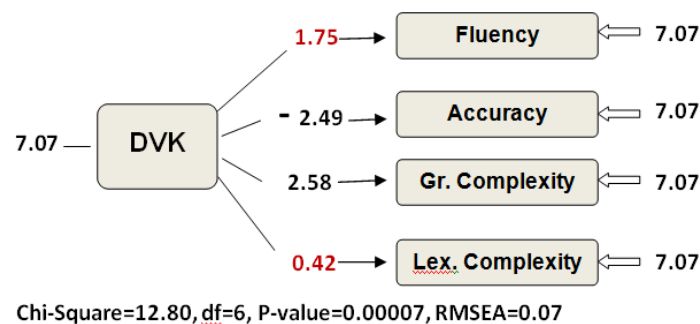

Chi-Square=12.80, df=6, P-value=0.00007, RMSEA=0.07

Figure 2. SEM modeling of DVK and total speaking performance in Planned Tasks

As demonstrated in Figure 2, the Chi-square is in balanced degree (12.80) and the P-value ($0.00007 \le 0.05$) is far from the significance level. RAMSEA also like the other measurements index is more than significance level ($0.07 \le 0.05$). Therefore, no causal relationship between DVK and speaking performance on unplanned interview tasks was established. The relationship between DVK and thethree tasks in each of the three rhetorical modes of description, narration and reasoning have been presented in Figures 3-5.
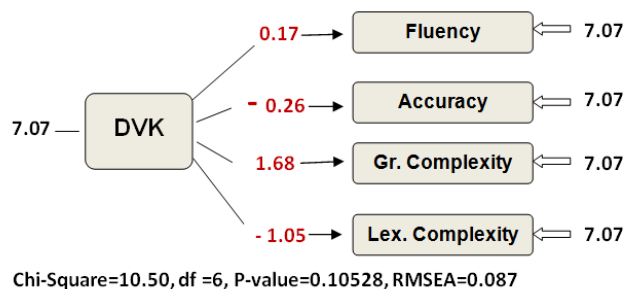
Figure 3. SEM of the relationship between DVK and performance on Description Task

As seen in the figure 3, the result of Chi- sauqare is 10.50 which shows the fitness of the model. On the other hand, the result of P-value is (0.10528 ≤0.05) and the RMSEA is (0.087≥ 0.05) that show the badness of fit. Therefore, no causal relationship is confirmed.
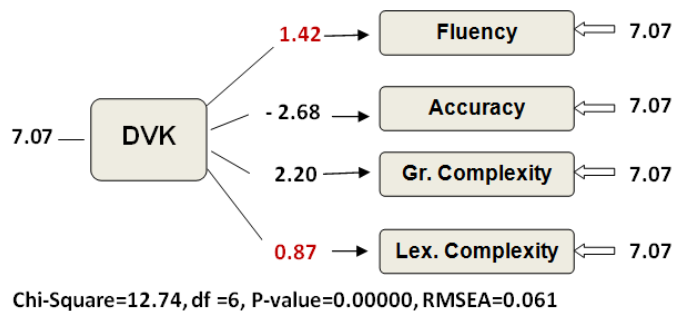

Figure 4. SEM of the relationship between DVK and performance on Narration Task

Figure 4, which carries the SEM for narrative unplanned task shows the result of Chi-square (12.74) which confirms the fitness of the model though the P-value in this model (0.0000≤ 0.05) and RMSEA (0.61≥ 0.05) are not in accordance with their index point.

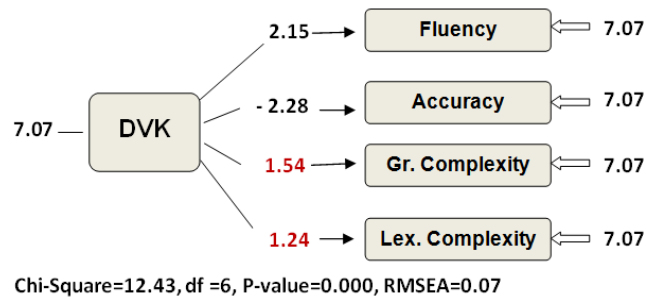Chi-Square=12.43, df =6, P-value=0.000, RMSEA=0.07

Figure 5. SEM of the relationship between DVK and performance on
Reasoning Task

In the last SEM (Figure 5), which demonstrates the reasoning unplanned task, like the previous SEMs, Chi- squrare indicates the fitness of model (12.43) while the P-value and RMSEA (0.000≤ 0.05), (0.07≥ 0.05) are far away from their determined index.

To sum up, the present study was a probe into the relationship between depth of lexical knowledge and speaking performance by mediation of task types. Since the task types had an important role in this investigation, the main results can be summarized as below:

1. Although the correlation between grammatical complexity and DVK in planned presentation task reached significance, the causal relationship between DVK and speaking performance on this task was not confirmed.

2. No correlations or causal relationship between the components of performance on unplanned descriptive task was approved.

3. Despite the observed correlations between DVK and accuracy and grammatical complexity in unplanned narration task, no causal relationship between DVK and speaking performance as elicited by narrative interview task was established.

4. Despite the observed correlations between DVK and fluency and accuracy in unplanned reasoning task, the study failed to affirm a causal relationship between DVK and speaking performance in the reasoning task.

Discussion

The results demonstrated that there was no significant causal relationship between deep vocabulary knowledge and speaking performance, and also the four aspects of speaking performance (fluency, accuracy, lexical complexity and grammatical complexity) are not equally related to deep vocabulary knowledge.

This study isolated only one aspect of L2 vocabulary knowledge, i.e., DVK, to study its effect on L2 speaking performance while previous studies have mainly investigated this relationship in terms of a combined mixture of lexical knowledge dimensions. So far, a high level of go-togetherness between breadth and depth of lexical knowledge and speaking proficiency/performance has been reported (e.g., 63% explained by size and 60 percent explained by depth in Koisume and In'nami's 2013 study). Koisume and In'nami (2013) asserted that the order of entering size and depth of vocabulary knowledge in studies based on regression analyses plays a role in relative effectiveness of the two in determining oral conduct in L2 speaking. They suggested that the effect of size and depth of vocabulary knowledge be investigated separately. However, the current study did not provide support for their research hint.

On the other hand, the mixed findings in previous research are commonly attributed to the type of task tests used to elicit speech performance. This line of research is one approach to deal with the complexity of processes involved in speech production. Koisume (2005) had already reported some task differences in the relationship between vocabulary knowledge and speaking performance. The study reported here presumed different cognitive-psychological processes involved in speech production for tasks informed by pre-planning differences and discoursal modes of description, narration and reasoning which has already been introduced in some other studies (Koisume, 2005; Safari Vesal, et al, 2015; Teng, 2007).

This conceptualization is worthy of further research attention to account for the nuances of speech production. For example, according to Skehan and Foster (1999), the processing load of task determines the

complexity of speaking performance, and some tasks might demonstrate a varying degree of lexical processing as a result of a higher level of processing load. In the present study, although a general relationship between DVK and speaking performance was not established, the variation witnessed for this relationship in terms of different aspects of speaking accomplishment examined here (i.e., FAC) across different task types in terms of the rhetorical mode of expression provide an important clue for investigating the cognitive processes involve in speech production as elicited by a variety of task types.

Conclusion

The sporadic variation witnessed in the relationship between DVK and FAC aspects of speaking performance in this study provides further proof for task effectiveness in this regard. Once such a load of task effectiveness is established, the area of assessing speaking will be immensely affected since each task will be able to stimulate certain aspects of cognitive processes including lexical access which may lead, in turn, to the belief that the knowledge and skills in using L2 lexicon need to be treated in a more systematic manner so that it is meticulously accommodated into second language teaching programs.

Taking task effectiveness into account in measuring speaking performance prevails an opportunity to adopt an assessment approach which is more compatible with the dynamic systems theory of language acquisition and use. We need further research before we admit the concept of task effectiveness in the process of lexical access as a source of variation in speech production capacity of L2 learners.

References

Afshari, S., & Tavakoli, M. (2016). The relationship between depth and breadth of vocabulary knowledge and Iranian EFL learners' listening comprehension. *International Journal of Research Studies in Language Learning 6*, 3, 13-24.

Alderson, J. C. (2000). *Assessing reading.* Cambridge: CUP.

Amini, D. (2014). *Taking sides: A textbook for topic-based conversation classes.* Tabriz: Forouzesh Publications.

Atai, M. R., & Dabbagh, A. (2010). Exploring the role of vocabulary depth and semantic set in EFL learners' vocabulary use in writing. *TELL 4*, 2, 27 – 49.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*, 191–208.

Baleghizade, S., & Khaledian, P. (2016). Investigating the relationship between depth and breadth of vocabulary knowledge and listening comprehension across different proficiency levels among Iranian EFL learners. *International Journal for 21st Century Education,* 3 (Special Issue 'Language Learning and Teaching'), 63-75.

Brown, G., & Yule, G. (1983). *Teaching the spoken language*: *An approach based on the analysis of conversational English.* Cambridge: Cambridge University Press.

Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.

Dabbagh, A., Janebi, M. (2017). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal,47*(5), 575-590 .

de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13*, 1-24.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Amsterdam, the Netherlands: John Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language.* Oxford: OUP.

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

Fulcher, G. (2003). *Testing second language speaking.* Essex, U.K.: Pearson Education Limited.

Funato, S., & Ito, H. (2008). An empirical study on basic requirements for Japanese EFL learners to achieve oral fluency in English. *Annual Review of English Language Education in Japan, 19*, 41-50.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, *25*, 375-406.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal, 36,* 153–166.

Hunt, A. & Beglar, D. (2005). *A framework for developing EFL reading vocabulary*. Reading in a Foreign Language, 17(1), 23-59.

Kaivanpanah, S., & Zandi, H. (2009). The Role of Depth of Vocabulary Knowledge in Reading Comprehension in EFL Contexts. *Journal of Applied Sciences*, *9*, 698-706.

Koizumi, R. (2005). Attainment levels of speaking ability of Japanese junior high school students. *JLTA (Japan Language Testing Association) Journal*, 7, 31-50.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language teaching and Research, 4*(5), 900-913.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. New York: Lawrence Erlbaum Associates.

Levelt, W. J. M. (1989). *Speaking*: *From intention to articulation.* MA: MIT Press.

Levelt, W. J. M. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: Walter de Gruyter.

Li, M., & Kirby, J. R. (2014). The effects of vocabulary Breadth and Depth on English Reading. *Applied Linguistics*, 36(5), 1-25.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

McNamara, T. (1996). *Measuring second language performance.* Essex, U.K.: Addison Wesley Longman Limited.

Qian, D. & Schedle, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance, *Language Testing, 21*(1), 28-52.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Ramezanee, A., & Hakimi, K. (2004). *The speaking test of IELTS with ample samples and examples.* Tehran: Rahnama Publishers.

Rashidi, N., & Khosravi, N. (2010). Assessing the Role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics, 14* (1), 81-108.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Sandlund, E., Sundqvist, P., Nyroos, L. (2016). Testing L2 Talk: A review of empirical studies on second language oral proficiency testing. *Language and Linguistics Compass, 10* (1), 14-29.

Schmitt, N. (2010). *Researching Vocabulary. A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26*,173–199.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93-120.

Stahr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139-152.

Stahr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577- 607.

Tahmasebi, G.-A., Ghaedrahmat, M., &Haqverdi, H. (2013). The relationship between language proficiency and Iranian EFL learners' knowledge of vocabulary depth versus vocabulary breadth. *Latin American Journal of Content and Language Integrated Learning*, *6*(2), 96-111.

Tarone, E. (2005). Speaking in a second language. In E. Hinkel (Ed*.), Handbook of research in second language teaching and learning* (pp. 485-502). Mahwah, NJ: Lawrence Erlbaum.

Teng, H. C. (2007). A study of task type for L2 speaking assessment. Paper presented at the Annual Meeting of the International Society for Language Studies (ISLS), Honolulu, HI, 2007.

Trent, J. (2009). Enhancing oral participation across curriculum: Some lessons from the EAP classroom. *Asian EFL Journal, 11*(1).

Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability, *Language Learning Journal, 47*(1), 64-75.

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: Issues, findings and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(1), 1-30.

Appendix A

Sample of the oral interview questions distribution scheme and three scoring systems of WAT

Participant	Questions Code	WAT Score One-point (160)	WAT Score Correct-wrong (320)	WAT Score All or Nothing (40)
P1	D1 N1 O1	77	202	6
P2	D2 N2 O2	126	253	12
P3	D3 N3 O3	99	220	8
P4	D4 N4 O4	117	234	12
P5	D5 N5 O5	85	179	2
P6	D6 N6 O6	97	210	4
P7	D1 N2 O6	115	236	11
P8	D2 N3 O1	106	241	11
P9	D3 N4 O2	123	247	11
P10	D4 N5 O3	107	219	4
P11	D5 N6 O4	107	113	6
P12	D6 N1 O5	109	235	11
P13	D1 N3 O5	80	220	4
P14	D2 N4 O6	118	235	9
P15	D3 N5 O1	108	217	9
P16	D4 N6 O2	93	232	10
P17	D5 N1 O3	105	245	11
P18	D6 N2 O4	102	232	10
P19	D1 N4 O4	119	239	11
P20	D2 N5 O5	110	234	8
P21	D3 N6 O6	101	226	10
P22	D4 N1 O1	117	232	8
P23	D5 N2 O2	100	230	8
P24	D6 N3 O3	119	240	11

P25	D1 N5 O3	106	221	6
P26	D2 N6 O4	112	231	5
P27	D3 N1 O5	103	235	12
P28	D4 N2 O6	110	243	9
P29	D5 N3 O1	84	223	7
P30	D6 N4 O2	97	226	7
P31	D1 N6 O2	123	254	15
P32	D2 N1 O3	65	29	2
P33	D3 N2 O4	118	235	9

Appendix B
Elicitation Questions of Free Speaking Task

A.   Description Task

D1: How do you describe the quality of education in your country, in comparison to some other countries you know?

D2: what kind of life does a housewife lead? Describe a housewife's day.

D3: Describe your ideal man (or woman) as your future husband (or wife).

D4: What's happiness to you? How do you define it? Illustrate your definition with some descriptive examples.

D5: Is it easy to find a job in your country? Describe the procedure for winning a job here.

D6: Who is the person you admire very much? Why do you admire him/her? Describe those characteristics of him/her that appeal to you.

B.   Narration Task

N1: Tell us in detail about a time when you really felt that your freedom was unfairly restricted. What happened exactly? What did you do?

N2: Have you ever seen someone being harmed by someone who had an evil eye? What happened exactly?

N3: Tell us about a true dream, that is, a night dream that you or someone else had and it came true the other day.

N4: Tell us a story in your life when you were punished, but this punishment ended up well because it had a constructive lesson for you.

N5: Tell us about a time when you met someone or learned something new that changed you in some way.

N6: Judging people by appearance can sometimes be deceiving. Tell about the experience of a situation that you or someone else judged a person, later realizing that it was wrong.

C.   Reasoning Task

R1: Brain drain has been a serious threat to our country's human resources. Do you think the blame for brain drain goes to our government and authorities or the emigrants themselves must be blamed?

R2: Do you agree with the statement?

"Women have been created to be mothers and wives. They can be more helpful at home."

R3: Do you think scientists should be allowed to clone animals and human organs?

R4: Are you in favor of single-sex classes or co-ed ones? What are your specific reasons for it?

R5: Do you think joining the globalization process by becoming a member of WTO will make our country richer or it will make the poverty conditions worse?

R6: In your opinion, what is the appropriate age for boys' and girls' marriage?