

Proceedings of the 12th  
METU International  
ELT Convention



# Proceedings of the 12th METU International ELT Convention

Edited by

Burcin Kagitci Yildiz and Sema Turan

Cambridge  
Scholars  
Publishing



Proceedings of the 12th METU International ELT Convention

Edited by Burcin Kagitci Yildiz and Sema Turan

This book first published 2018

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2018 by Burcin Kagitci Yildiz, Sema Turan and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-0531-6

ISBN (13): 978-1-5275-0531-5

# TABLE OF CONTENTS

## **Part One: Skills Development**

The Effect of Using Concept Mapping to Improve Reading  
Comprehension in ESP ..... 2  
*Fatemeh Fazlali*

The Impact of Contextual Clues on Oral Performance of EFL Learners .. 13  
*Nazlınur Göktürk*

The Effect of SHIP on Teaching Vocabulary to EFL Students ..... 39  
*İbrahim Topal*

A Critical Analysis of Oral Interviews and Guided Role Play Activities  
in Assessment of Speaking ..... 57  
*Nuriye Karakaya Yildirim*

Enhancing EFL Learners' Literary Competence through a Holistic  
Approach ..... 76  
*Noureddine Guerroudj*

## **Part Two: Technology in English Language Education**

Digital Natives and Imaginative Materials ..... 88  
*Deren Başak Akman Yeşilel and Müfit Şenel*

Emerging Multiple Identities in Students' Social Media Interaction..... 102  
*Arzu Ekoç*

## **Part Three: Teacher Education**

Continuous Professional Development in an EFL Program:  
Meeting the Diverse Needs of EFL Instructors ..... 128  
*Yeşim Nalkesen, Zeynep Gönenc Afyon and Donald Staub*

Teacher Efficacy and Concerns in Pre-Service ELT Teachers.....	155
<i>Seyit Ahmet Çapan</i>	
Pre-Service EFL Teachers' Beliefs on Learner Autonomy based on Practicum Experiences .....	174
<i>Banu Çiçek Başaran, Fatma Gümüřok and Melike Bekereci</i>	
EFL University Teachers' Autonomy as a Prerequisite to Learners' Autonomy .....	197
<i>Fawzia Bouhass Benaissi</i>	
Shakespeare's Unruly Women .....	208
<i>Luke Prodromou</i>	
<b>Part Four: Learner Beliefs and Perceptions</b>	
Perceptions of EFL Learners towards the Impact of Verbal Teacher Immediacy Behaviors on Their Motivation to Learn English .....	222
<i>Zeynep Ünlüer and Iřıl Günseli Kaçar</i>	
Perspectives of Students as Regards Group Work: A Case Study.....	256
<i>Esrá Ataman, Elif Aksoy and řeniz Bilgi</i>	
Advanced Perception Capacity of Blind Individuals and Its Utilisation in Foreign Language Learning .....	288
<i>Nihansu Yurtseven</i>	
Grammatical Competence in Exam Writing and Take-Home Assignments .....	300
<i>Kadriye Aytaç and Nurdan Özbek-Gürbüz</i>	
Contributors.....	331

**PART ONE:**  
**SKILLS DEVELOPMENT**

# THE EFFECT OF USING CONCEPT MAPPING TO IMPROVE READING COMPREHENSION IN ESP

FATEMEH FAZLALI

## Abstract

Since concept mapping has been proposed by Joseph D. Novak in the early 1970's, many aspects of his implications have been studied. Concept mapping technique is based on cognitive structures and Ausubel's meaningful learning. Concept mapping provides logical relationship between concepts in the learner's mind by creating an active teaching context and providing learners with meaningful learning. This study was done to investigate the effect of using concept mapping on the reading comprehension of physical education students in English for Specific Purpose (ESP). In order to collect data, purposive sampling was used. Participants were 60 university students majoring in Physical Education at University of Farhangian, Nasibeh Campus in Tehran. They were placed in two groups, experimental and control groups. To teach reading, traditional teaching was used for the control group and concept mapping was taught to the experimental group. The effect of using this technique on the reading comprehension of the students was measured by administrating a pretest and a post-test. To analyze the data, SPSS 21 was used. Results showed that using concept mapping had a significant effect on the improvement of the reading comprehension of the physical education students in ESP.

**Key words:** concept mapping, concept map structure, reading comprehension.

## 1-Introduction

Reading as a critical aspect of literacy is regarded as an interaction between the reader and the text (Alderson & Urquhart, 1984). Reading comprehension is the basis for student performance in all subjects (Mohan, 2002). However, reading to learn is a struggle for many students. (RAND,



Reading study Group, 2002). Among the variety of techniques and skills of teaching reading, five of them were identified as essential by the National Reading Panel (2000) which were phonemics awareness, phonics, comprehension, fluency and vocabulary. Comprehension, in particular, plays a key role as a reading skill that EFL learners are required to master.

Reading comprehension is crucial in Iran as an EFL context as well. There is a vast range of research in this scope to find the learners' difficulties and to find some solutions to them. Gorjian, Hayati, and Sheykhiyani (2009) claimed that in the Iranian educational system, learners have difficulties in comprehending a text and there is no special policy in order to teach them some strategies for better comprehension. In this educational system, there are just some activities and questions after reading passages (post-reading activities) to check learners' comprehension; it seems strategy instruction is missing.

As teachers can teach strategies to improve the reading comprehension of their students, Zipp, Maher, and D'Antoni, (2011), by emphasizing this role, stated that teachers are unable to teach students how to think. They believed that teachers' obligation is to help their students, and that teachers should try to trigger the cognitive processes in the students' mind with ever increasing power and clarity as they grow and learn. Thompson (2000) stated that applying reading strategies would enhance the students' rate of comprehension and change them from passive readers to active ones. As Ness (2009) argued, the instruction of reading comprehension strategies is highly beneficial in all the levels and as Javdan Faghat and Zainal (2010) restated, these strategies help the reader to interact with the text. Therefore, teachers must provide students with the correct tools in reading. Teacher strategy instruction associated with improved comprehension (NRP, 2000: RAND Reading Study Group, 2002) consists of teaching students to monitor comprehension.

As a result, based on *the idea of learning how to learn* and investigating *the process of students' learning*, the idea of concept mapping (CM) was emerged in the educational context.

### **1-1-Theoretical Framework of Concept Mapping**

Cognitive, affective and psychomotor are the distinct domains of learning which do not occur in isolation, but work together to make one whole. The cognitive domain includes aspects of thinking, knowing and problem solving (Mohan, 2002). In constructivist view, these tasks occur as a "process of personal cognitive construction, or invention, undertaken

by the individual who is trying, for whatever purpose, to make sense of her social or natural environment” (Taylor 1993, cited by Duit 2007).

Concept maps were developed in 1972 in the course of Novak's research program at Cornell University where he sought to follow and to understand changes in children's knowledge of science (Novak & Musonda, 1991). Novak (1972) said by the age of 30 months a child recognizes the concept labels by concept formation. He also believed that meaning making is a concept assimilation that never finishes in a life time. During the course of that research, it became clear that concept maps were useful to represent the change in children's understanding of a topic. At present, they can be used as an excellent tool for participating graduate students to express their understanding of their courses. Concept maps have a strong psychological and epistemological foundation, based on Ausubel's Assimilation Theory (Ausubel, 1968, 2000) and Novak's Theory of Learning, which explain that people learn new things by using their current knowledge and, more or less, by seeking ways to integrate new knowledge and related knowledge which is already known. In meaningful learning, the integration of new concepts into our cognitive knowledge structure takes place through linking new knowledge to the concepts we have already understood.

The popularity of concept mapping has spread. Recently, it is used all over the world as a means to represent a person's knowledge about a domain of knowledge, by users of all ages and in all domains of knowledge. It has subsequently been used as a tool to increase meaningful learning in the sciences and other subjects as well as to represent the expert knowledge of individuals and teams in education, government, and business.

According to Novak & Canas, (2006) “concepts are the building blocks of knowledge in all fields” and this can be triggered in individuals either for “static thinking (surface or rote thinking) or dynamic thinking (deep or meaningful thinking)”. Concept map is defined as "a graphical tool for organizing and representing relationships between concepts indicated by a connecting line linking two concepts" (Novak & Canas, 2007, p.1). In another definition, Novak (2010) described it as a new model for education. Concept maps are “visual road maps” to show pathways in a specific knowledge domain. Concept maps define the relationship between the concepts. Concept maps represent “meaningful relationships between concepts in the form of propositions” (Novak & Gowin, 1984, p.25). Novak (2006) explains “propositions as two or more concept labels linked by words in a semantic unit”. In concept maps, the concepts are seen in boxes or bubbles and there is a link line with a

connecting verb. Concept maps show hierarchy with the most inclusive concept at the top and the subtopic comes down a line which can have cross links. Concept maps have a structure and they can be in different types such as hierarchical, focus question and cross-links.

## **1-2- Characteristics of Concept Maps**

Novak (2006) specifies the characteristics of concept maps as follows:

- 1 - Concept maps include concepts usually enclosed in circles or boxes of some type, and relationships between concepts or propositions, indicated by a connecting line between two concepts. Words on the line specify the relationship between the two concepts.
- 2 - The concepts are represented in a hierarchical fashion with the most inclusive, most general concepts at the top of the map and the more specific, less general concepts arranged hierarchically below.
- 3 - Concept maps contain the cross-links. These are relationships (propositions) between concepts in different domains of the concept map. Cross-links help us to see how some domains of knowledge represented on the map are related to each other.
- 4 - Concept maps have specific examples of events or objects that help to clarify the meaning of a given concept.

## **2- Statement of the Problem**

Researchers have found that teaching reading strategies is important in developing increased student comprehension (National Reading Panel, 2005). Concurrently, researchers have identified teachers often lack a solid foundation for teaching reading comprehension strategies (National Reading Panel, 2005). Therefore, teachers need to be prepared, through professional development to teach comprehension strategies to students effectively.

## **3- Purpose of the Study**

This study aims to investigate the effect of concept mapping and examining the effectiveness of using concept mapping as a supplemental tool to improve reading comprehension of the students majoring at physical education in ESP.

#### **4- Research Question**

Is there any significant difference between concept mapping and traditional strategy of reading in improving reading comprehension ability of the physical education students at University of Farhangian ; Does concept mapping have any impact on improving students' reading comprehension?

#### **5- Methodology**

##### ***Participants***

The subjects of this study were selected among 60 university students from four classes of the University of Farhangian, Nasibeh Campus in Tehran. The students were majoring in Physical Education. All of them were girls and the age range was about 18 to 23 years old. Afterwards, they were divided into two groups each consisting of 30 subjects.

##### ***Materials***

###### ***Reading Comprehension Pretest***

To conduct this study, a reading comprehension test as a pretest was administered to measure the comprehension ability of the students before applying the treatment. The pretest involved five passages of comparable length (average of 500 words) with the same readability level along with a total of 30 multiple choice questions. The pretest included five passages from a course book which was "English for Specific Purposes" by SAMT Publication (1388).

###### ***Teaching Passages***

Five passages from this book were selected to be practiced in the classroom, which were the same for both groups and contained an average of 500 words for each passage.

###### ***Reading Comprehension Posttest***

A parallel reading comprehension test similar to the pretest was used as a post-test to assess the efficacy of the treatment on the students. The post-test contained 30 multiple choice questions. The passages used were selected from the same book as those used for the pretest and teaching passages.

## 6- Procedure

### *Pretest*

To determine the reading comprehension ability of the subjects, before conducting the treatment, the mentioned pretest was administered to both groups. Subjects were supposed to answer all questions in 40 minutes. The procedure which was used in administering the pretest was exactly the same for both groups.

### *Treatment*

The whole study was conducted within three months. The selected strategy for the experimental group was two versions of CM including learner-constructed CM and fill in the map CM. For the control group, a traditional reading comprehension strategy (i.e. summarizing strategy, or translation) was applied. Once the pretest was administered to both groups, a treatment procedure involving five-session instructions and practice of the particular strategy for a period of 20 minutes for each session was assigned to each group.

### *Experimental Group*

#### *Concept Map Drawing Instruction*

In this group, subjects were going to learn how to draw CMs for the five passages. The first step in CM training was to introduce to the participants with the ideas of CM and its advantages. Participants were provided with handouts that included the definition of CM and its applications, the procedure of drawing a CM and a passage with a drawn CM as a model. The strategy was taught explicitly step-by-step. Then, students were taught the reading passages and they were asked to draw a CM of their own at home and to submit them to their teacher. Every two sessions, the subjects' drawn CMs were collected and were read and some feedback was provided for them. The purpose of the evaluation of CM and provision of feedback was to help the students learn how to create a good CM progressively up to the last session. A holistic method of scoring with a scale of good, fair, and weak was applied to assess the students' CM and to provide them feedback. At the end of the course, they would receive a complete feedback report including the reviewing of the important points.

#### *Fill in the Map Concept Map Instruction*

Instead of creating a concept map, participants were provided with the drawn CMs of their passages in which some concepts or linking lines were left out and students were asked to fill in the maps based on the passages

they had been taught. Students were provided with the same handouts about CM as the first group. The difference was in the way they should do their assignments. A holistic method of scoring with a scale of good, fair and weak was applied to assess the students' process of filling CMs and to provide appropriate feedback. In each of the five sessions, students had to submit their assignments. Some useful comments and feedback, based on the students' filled maps, were provided orally.

### ***Control Group: Traditional methodology***

Participants of this group were asked to translate the passages that they had been taught before. The reading passages were taught in each session and subjects were asked to translate the passages and do the exercises in the class. Based on the participants' translations, questions and problems, some comments were prepared orally in each session.

### ***Post-test***

At the end of the treatment, a post test was administered to both groups. Based on the property of this study, there was a difference between the administration of the pretest and the post-test. The difference between the pretest and the post-test was that for the pretest, all of the groups answered reading comprehension questions without employing the strategies; whereas, on the post-test, they employed the strategies that they had learned, and then, they had to answer the reading comprehension questions in the multiple choice form.

## **7-Results**

To answer the research question a paired sample t-test was applied by using SPSS 21.

### **Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	CpTest2	22.40	30	5.184	.946
	CpTest1	17.43	30	4.614	.8

Fig.1

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 CpTest2 & CpTest1	30	.924	.000

Fig-2

**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair CpTest2 - 1 CpTest1	4.967	1.991	.364	4.223	5.710	13.663	29	.000

Fig.3

A Paired-sample t-test was conducted to evaluate the impact of the treatment on the students' scores. There was a statistically significant increase in scores from time 1 (M=17.43, SD=4.61) to time 2 (M=22.40, SD=5.18),  $T(29) = 13.66, P < .001$  (two-tailed).

The mean increase was 4.97 with a 95% confidence interval ranging from 4.22 to 5.71.

**Paired Sample Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 TradTest2	19.70	30	3.344	.611
TradTest1	17.83	30	3.659	.668

Fig.4

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 TradTest2 & TradTest1	30	.855	.000

Fig.5

**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 TradTest2 - TradTest1	1.867	1.907	.348	1.155	2.579	5.361	29	.000

Fig.6

A paired-sample t-test was conducted to evaluate the traditional method. There was a statistically significant increase in scores from time 1 ( $M=17.83$ ,  $SD=3.65$ ) to time 2 ( $M=19.70$ ,  $SD=3.34$ ),  $T(29) = 5.36$ ,  $P<.001$ (two-tailed). The mean increase was 1.87 with a 95% confidence interval ranging from 1.15 to 2.58. Comparing the means of both groups showed that the concept map technique was higher than the traditional method in improving the reading comprehension in ESP. Actually, the students had an improvement in their reading comprehension, but the mean score of concept map technique was significantly higher than the traditional method.

**8-Implication of the study**

The findings of the present study might offer some practical implications for the work of classroom teachers who are highly motivated to welcome challenge and novelty in their classrooms.

Using concept map techniques in improving writing is suggested for the future research.

**9-Limitations and delimitations of the study**

Owing to the practical limitation regarding doing assignments, the present study focused on a small sample of female participants. However, another study can be conducted with a larger sample and with both genders to examine whether such possibilities will provide different results. Furthermore, it should be interesting if future research focuses on the two other versions of CM which are expert constructed and cooperative CM.



## References

- Alderson, C., & Urquhart, A. H. (1984). Introduction: What is reading? In C. Alderson, & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 16-28). London & New York: Longman.
- Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston.
- . (2000). *The Acquisition and Retention of Knowledge: A Cognitive View*. Dordrecht; Boston: Kluwer Academic Publishers.
- Gorjian, B., Hayati, A., & Sheykhiyani, M. J. (2009, May). *The role of gisting and contextual guessing Reading strategies in learners' performance in multiple-choice cloze tests of reading comprehension*. Paper presented on 7th international conference by the schools of language studies and linguistics faculty of social sciences and humanities. University Kebangsaan, Putrajaya, Malaysia.
- Grabe, W. (2009). *Reading in a second language*, Cambridge: Cambridge university press.
- Hattie J.A.C. (2009). *Visible learning*, London :Routledge.
- Hedge, T. (2000). *Teaching and learning in the language classroom*, Oxford: Oxford university press.
- Javdan Faghat, L., & Zainal, Z. (2010). *Reading strategies of advance and intermediate Iranian students: A case study. edupress 2010*.University Press.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Ness, M. K. (2009). Reading comprehension strategies in secondary content area classrooms: teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons*, 49(2), 143-166.
- Novak, J. D., & Canas, A. J. (2007). The theory underlying concept maps and how to construct and use them. *Reflective education*, 3(1), 29-42. Retrieved from <http://reflectingeducation.net>
- Novak, J. D., & Gowin, D. B. (1984). *Learning How to Learn*. New York, NY: Cambridge
- Novak, J. D., & D. Musonda. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal*, 28(1), 117-153.
- Soleimani, H. (2009). *Non-parametric statistics for applied linguistics research: An introduction*. Tehran:Rahnama Press.

- Sprenger, M. (2005). *How to teach so students Remember*, USA: Association for Supervision and Curriculum Development.
- Thompson, S. (2000). *Effective content reading comprehension and retention strategies*. (Eric Document Reproduction Service No. ED 440372)
- Tholmlinson, B. (2013), *Applied Linguistics Development*, USA: Bloomsbury Academic
- Wilson, J. & Wing Jan, L. (2008). *Smart Thinking*, Australia: Routledge
- Zadina, J.N & Smilkestein , R.& Daiek, D.B.& Anter. N.N.(2014), college reading the science and strategies of Expert Readers. Boston: Wadsworth

# THE IMPACT OF CONTEXTUAL CLUES ON ORAL PERFORMANCE OF EFL LEARNERS

NAZLINUR GÖKTÜRK

## Abstract

This study reports the results of an experimental action based research conducted in an EFL classroom to investigate the impact of the presence and absence of contextual clues with speaking tasks on learners' oral performance. Specifically, it aims to explore the relative contribution of contextual clues presented with tasks to the content development and composite scores of learners. It also investigates learners' perceptions of the tasks presented with and without contextual clues. Sixty-two EFL learners divided into control and experimental groups participated in this study. A speaking test consisting of open-ended questions presented with and without contextual clues was used for the assessment of learners' performances. The test scores assigned to the learners through analytic rating were statistically analyzed through *SPSS Version 20.0*. An analysis of the test takers' discourse was undertaken to determine whether measures of content in participants' output differ according to the tasks presented with and without contextual clues. A focus group interview was conducted to obtain learners' perceptions of the tasks presented with and without contextual clues. Quantitative analysis of the data showed that learners did not perform significantly differently in the tasks presented with and without clues. A detailed analysis of the learners' performance discourse indicated that the measures of content in terms of quantity in participants' output did not significantly differ according to different task types. However, the analysis of the qualitative data revealed that learners given the tasks presented with contextual clues felt more relaxed and free to express themselves in the assessment sessions.

**Keywords:** Speaking testing, content, task type

## Introduction

With the increasing value attached to speaking as part of one's language competence within the communicative language teaching paradigm, the assessment of the speaking skill in second language learning has become a burgeoning area of research over the past two decades. However, assessing the speaking skill is essential, yet challenging in nature since there are so many factors influencing our impression of how well someone can speak a language (Luoma 2004). As pointed out by Bachman and Palmer (1996), the measurement of oral performance continues to be "one of the areas of most persistent difficulty in language testing" (p.154). Several studies have been conducted to investigate the validity of scores and rating scales (Meredith 1990, Halleck 1992, Reed 1992), the rater's behavior and its impact on the learner's performance (Barnwell 1989, Thompson 1995), the interviewer's variation (Brown 2003), and the interviewee's individual differences, e.g. proficiency level (Young 1995), the ability to self-adjust (Ross 1992, Cafarella 1994) and differences in linguistic backgrounds (Young and Halleck 1998). While there is now a considerable amount of research on the discourse produced within oral language interviews, more than half of the analyses pertain to characterizing interviewer and candidate discourse in general (Brown 2003).

Recently, there has been a keen interest in the detailed analysis of discourse features of test-taker's oral performance and various studies have been carried out in both language pedagogy and experimental contexts (e.g., Skehan and Foster 1997, Bygate 1999, Foster, Tonkyn, and Wigglesworth 2000, Iwashita et al. 2008). Skehan (1998), among numerous others (Foster and Skehan 1996, 1999, Porter and O'Sullivan 1994, 1999, Wigglesworth 1997), has attempted to manipulate psycholinguistic aspects of tasks in order to modify or predict task difficulty.

Foster and Skehan (1996) investigated the effects of planning time and three different tasks (i.e., personal information exchange, narrative, and decision-making) on the variables of fluency, complexity, and accuracy. They found that planning time had more influence on narrative and decision-making tasks than on personal information exchange tasks. The researchers argue that personal information exchange task is the easiest one as the information is familiar to all learners, and is well-structured. In a similar vein, Foster and Skehan (1999) explored the effects of inherent task structure and processing load of a narrative retelling task. They suggested that more structured tasks elicited more fluent language, and complexity of language was influenced by processing load. Likewise,

Tavakoli (2009) investigated the impacts of storyline complexity and task structure of narrative tasks on second language oral performance. The researcher found that structured tasks elicited more accurate and fluent language than less-structured tasks, and storyline complexity enhanced syntactic complexity of L2 oral performance; however, no clear impact on the lexical diversity of the language produced was found, which is congruent with the findings of the study conducted by Skehan and Foster (1999). Yet, Iwashita, McNamara and Elder (2001) appear to cast doubt on the possibility of categorizing tasks in terms of difficulty following this approach. The researchers conducted a study (following Skehan 1998) to explore if cognitive demands of tasks influence L2 oral performance. They tried to manipulate task difficulty by increasing the cognitive demands of tasks through task design and performance conditions, and analyzed the oral performance data through both specifically constructed scales and detailed discourse measurements. They surprisingly found that task performance conditions failed to influence task difficulty and task performance.

Of relevance to the present study is also the research on the effects of task types and structures on oral performance, in particular on the content development or elaboration of test-taker. Content that examinees attempt to convey is not always included in the construct definition of speaking tests and analytic scales as a feature of performance, though the linguistic components of oral performance (e.g., pronunciation, grammatical accuracy and vocabulary) are commonly assessed. McNamara (1996) points out that most general-purpose proficiency tests assess speaking with a great focus on the linguistic quality of test-takers' oral performances, rather than assessing their performances based on real-world criteria or task fulfillment. As such, test-takers' oral performances are simply evaluated on the basis of their linguistic ability, rather than their ability to communicate effectively.

Regarding the inclusion of speech content in rating scales, Sato (2012) explored the relative contribution of linguistic criteria (i.e., grammatical accuracy, fluency, vocabulary range, and pronunciation) and the elaboration of speech content, that is to say, how well the examinees developed ideas on given topics, to scores on a test of speaking proficiency. He asked nine raters to rate 30 students' monologic performances on three different topics. The findings indicated that content development substantially contributed to the intuitive judgments of raters and composite score of test-takers. He suggested that the quality of ideas that examinees attempt to convey should be involved as a criterion in oral assessments on the grounds that relying solely on linguistic features may

lead to incorrect inferences with regard to L2 learners' oral communication skills.

One of the most often cited studies on the effect of task types on the quality and quantity of content, as well as the other features (e.g., fluency, grammatical accuracy) of test-takers' oral performances is that of Brown, Iwashita and McNamara (2005). They used a discourse-analysis based methodology to investigate the relationship between test scores and actual performance discourse across two integrated speaking tasks of the TOEFL IBT, reading-speaking and listening-speaking, along with the independent speaking task. They found that test takers' performances did not significantly differ across task types, except for the listening-speaking integrated task, which proved to be significantly more difficult than both the integrated reading-speaking task and the independent task. As mentioned earlier, they examined content as a feature of performance, measured it in terms of quantity and schematic structure. While they found that most performance features examined were not task specific, a notable exception was the finding that learners produced more ideas in the integrated tasks than in the independent tasks, which leads them to suggest that content category tended to be task specific.

The supplience of stimulus materials with speaking tasks is theoretically intended to provide the examinee with a stimulant to trigger the intended speech. Butler et al. (2000) highlighted that, "... speech cannot exist in a vacuum; the examinee must receive input in some form, either spoken or written, in order to know what task he/she is being required to perform" (p.16). Concerning this matter, Akhondi, Malayeri and Samad (2010) conducted a study to examine the impact of the length of the input given prior to the interview on the overall speaking performance of test-takers, including their use of target vocabulary and grammatical structures in speech. They compared the interviewees' performances in two experimental groups receiving different lengths of the same prompt with the performances of a control group who had an impromptu interaction with the interviewers in the absence of any prompt. They found that providing interviewees with prompts noticeably affected the test takers' performances, which leads the researchers to claim that longer prompts provide more "fruitful floor for the participants to ground their discussions and responses on".

Although a number of studies have focused on task design and the impact of task types on oral performance of examinees, few attempts, if any, have been taken on researching alternatives to open ended question types which are currently used to assess speaking in most EFL classroom based assessment contexts. Besides, little empirical research has been

conducted to see the relative contribution of the presence of contextual clues with speaking tasks to both the overall oral performance and content development of low proficiency L2 learners. Thus, the present study attempts to explore the impact of the presence and absence of contextual clues with two monologue types of open ended speaking tasks (i.e., narrative and descriptive) on test-takers' oral performance. In particular, it aims to find out how the suppliance of contextual clues in narrative and descriptive tasks affect the content development/elaboration of learners. By relying on in-depth analyses of the actual performance discourse and scores assigned through analytic rating, this study will seek to contribute to our understanding of L2 speech performance, and further offer implications for designing speaking tests in EFL classroom based assessment contexts.

### **Research Questions**

The research questions addressed in this study are:

1. Do L2 learners' oral performance scores on narrative and descriptive tasks with and without contextual clues differ?
  - Does providing contextual clues influence the composite scores, in particular the content development of the learners?
2. What are the learners' perceptions regarding the task demands of narrative and descriptive oral assessment tasks with and without contextual clues?

### **Hypotheses**

In accordance with the previous studies conducted on the impact of different task types on oral performance (Skehan and Foster 1996, 1999, Tavakoli 2009), the following hypotheses were formulated.

1. Both speaking tasks (i.e., narrative and descriptive) presented with clues will facilitate better development of content in test-taker's oral performance as compared to the ones presented without clues.
2. Learners are expected to perform significantly differently on narrative and descriptive tasks presented with and without contextual clues.
3. Learners are expected to favor the tasks with contextual clues over the tasks presented without contextual clues.

## **Methodology**

### **Context of the Study**

The study was carried out in the School of Foreign Languages at Hacettepe University where each student enrolled is required to have a certain level English proficiency to be eligible for the freshman year. In the curriculum, four hours a week is devoted for speaking which is integrated with listening in all levels. Although the assessment of speaking is not included in general proficiency tests administered in the school, it is assessed as a part of achievement tests throughout the quarters. The contribution of speaking scores of learners to their overall grades is %10. All the speaking exams conducted at the preparatory school include similar tasks. The exam consists of two parts. In the first part, the learners are asked a few warm-up questions. In the second part, they are asked to talk about the topic that they pick up from a box for 1-2 minutes on their own. The questions given to the learners are open-ended questions eliciting different discourse types (e.g., descriptive, narrative, explanatory and predictive). For instance, the learners are asked to describe something that is familiar to them, such as their home or favorite friend. The learners' performances are evaluated using the public version of IELTS speaking scale, which is an analytic scale.

### **Participants**

62 Turkish-speaking adult L2 learners of English enrolled in preparatory classes at Hacettepe University participated in this study. 36 of them are female while 26 of them are male with an age range of 19 to 25. The proficiency level of the learners is A2 according to Common European Framework of Reference for Languages (2001). The learners were divided into two groups as experimental and control groups based on their average scores on the level achievement test conducted by preparatory school to ensure that both groups have similar language abilities. The first group has a mean score of 78.17 (out of 100) whereas the second group has a mean score of 77.96 (out of 100). When the two mean scores were compared through an independent samples *t*-test, it was found that the significance of the difference between them is  $p=.91$  (The alpha level was set as  $p=.05$ ). According to this result, it can be concluded that the difference between the means scores of both groups is not statistically significant, and the groups are very identical in terms of their language abilities.



## Instruments

A speaking test consisting of open-ended questions was used for the assessment of learners' oral performances (Appendix-A). The open-ended questions presented with and without contextual clues were designed to elicit two different monologue types of discourse (i.e., narrative and descriptive) from learners on a par with their proficiency level and regular activities in class. For each task type, narrative and descriptive, a set of five prompts of equal difficulty were developed by the researcher who was also the teacher of both groups in question as she was quite familiar with the teaching and learning process in class. This task format was chosen because it is a common format used in classroom based assessment contexts, and test-takers need to demonstrate their ability to use language effectively at some length (Underhill 1987). To further ensure the content validity of the test, a group of expert judges, namely the head of the testing unit of the department and two coordinators, were asked to evaluate the speaking test. Thanks to the researcher's and experts' judgments, the content validity of the test was tried to be secured. With regard to the reliability of the test, Underhill (1987) suggests that "the classical measures of test reliability have little relevance for oral tests because they are designed for rigid, pre-planned tests consisting of a fixed number of individual questions" (p.106). As such, more useful information could be gathered by comparing each marker's scores with her/his own scores, or with the scores of other markers. Based on this, the inter-rater reliability of the scores assigned to the learners by two raters was examined in this study.

As for the rating scale used for the assessment of learners' oral performances, the public version of IELTS speaking scale was adapted. As the main focus of the study was to investigate the impact of the presence and absence of contextual clues with the tasks on learners' content development/elaboration, the criterion of (a) *Fluency and Coherence* in the IELTS scale was divided into two separate criteria as (a) *Fluency* and (b) *Content Development*. A concise definition of *Content Development* as a criterion was given as "the degree to which the test-taker is conveying relevant and well-elaborated/developed ideas on given topics" (Sato 2012, 226). Besides, brief descriptors for each level were developed to achieve high inter-rater reliability (Appendix-B).

In addition to the speaking tasks and adapted scale, a focus group interview was conducted to be able to investigate the learners' perceptions regarding the task demands of narrative and descriptive oral assessment tasks presented with and without contextual clues. Focus group interview was chosen as a method of data collection for it elicits rich qualitative data

efficiently (Dörnyei 2007). Homogenous sampling technique was adopted to select the participants for "... the dynamics of the focus group works better with homogenous samples" (Dörnyei 2007, 144). As the researcher's main aim was to create a supportive atmosphere in which discussion was promoted by giving chance to the participants to explain their points of views, she asked more general questions such as "What do you think about the oral assessment type you were involved?", "Do you think you could be more successful if you were in the other group?", "Do you find it useful to have contextual clues?". According to Marshall and Rossman (1999), focus group interviewing generally includes 7 to 10 participants, but the size of the group can change. In this study, 10 learners participated in the focus group interview.

### **Procedure**

The purposes of the assessment of speaking at Hacettepe University Preparatory School were to evaluate the learners' oral performance in English, and to provide learners with useful feedback in the process of learning. The learners were required to answer the given prompts presented with or without contextual clues by narrating or describing them. They were given one minute planning time before they started to talk about the topics given. They were randomly given the tasks, and thus they fell into the groups of descriptive and narrative tasks presented with and without clues in the study. The time allotted for each learner was 5 minutes, including the warm-up session. The conversations were recorded on a tablet for transcription and analysis.

In order to ensure valid and reliable scoring, a second rater with two years of experience in both teaching and assessing the speaking ability was involved in the assessment procedure together with the researcher. The second rater also worked as an instructor in the same school with the researcher, and both raters had the formal training on the use of IELTS speaking scale. Prior to the oral assessment, the second rater was informed about the adapted version of scale to be used in the assessment. Two weeks after the oral assessment, 10 learners were interviewed together by using focus group method, through which the researcher aimed to get detailed information on which task type the learners favored and why.

### **Data Analysis**

The test scores assigned to the learners by two raters using an analytic scale were statistically analyzed through *SPSS Version 20.0*. First, Pearson

correlation coefficient was computed between two sets of *composite* scores assigned to the learners by two raters to confirm inter-rater reliability. Next, to see if there was a significant difference between experimental and control group with regard to their composite scores and scores from the content development/elaboration criterion in the scale, an independent samples *t*-test was run on the data. This analysis was intended to explore the relative contribution of providing contextual clues with tasks on learners' oral performance and content development. A one-way ANOVA was also performed to investigate the effects of descriptive and narrative tasks presented with and without clues on the learners' composite scores.

In addition the statistical tests employed in the study, an analysis of the test takers' discourse was undertaken to determine whether measures of content in participants' output differ according to the tasks presented with and without contextual clues. After transcribing the recorded performances, T-units and clauses were counted as measures of content in terms of quantity in the participants' oral performance data. Segmentation of learners' utterances into T-units and clauses was done according to guidelines applied by Brown, Iwashita, and McNamara (2005) for a study of a detailed analysis of test-takers' discourse. According to these guidelines, a T-unit is defined as an independent clause and all of its dependent clauses (Hunt 1970); a clause is a unified predicate containing a finite verb, a predicate adjective, or a non target-like predication in which the verb or part of the verb phrase is missing (Berman and Slobin 1994, cited in Brown, Iwashita, and McNamara 2005). On the basis of this approach, coordinated clauses are treated as separate T-units while subordinated clauses together with the main clause are counted as a single T-unit. Besides, coordinated verb phrases with the same subject are treated as part of a single unit (Foster, Tonkyn and Wigglesworth 2000). To see whether there was a significant difference between groups with regard to the number of the T-units and clauses, an independent samples *t*-test was run on the data. Besides, Pearson correlation coefficient was computed to assess the relationship between the number of clauses as a measure of content in terms of quantity and composite scores assigned to the learners by two raters.

Finally, data collected through a focus group interview regarding the learners' perceptions of the tasks presented with and without contextual clues was categorized qualitatively.

## Results

### Inter-Rater Reliability

Since this is a test of production where raters' judgments affect the decision to be made about the performances of learners, the reliability of the test scores assigned to the learners by two raters was examined by calculating the correlation coefficient of the scores.

**Table-1. The Correlation Coefficients of Raters' Ratings**

		First Rater	Second Rater
First Rater	Pearson Correlation	1	,951**
	Sig. (2-tailed)		,000
	N	62	62
Second Rater	Pearson Correlation	,951**	1
	Sig. (2-tailed)	,000	
	N	62	62

\*\* . Correlation is significant at the 0.01 level (2-tailed).

As presented in Table 1, the correlation coefficient obtained for two raters is .951, indicating a quite high inter-rater reliability. Besides, it was found that the correlation coefficient for first and second rater was statistically significant with a *p*-value smaller than .05, which means that test scores assigned to the learners by two raters are reliable.

### Comparison of Tasks and Task Types

The main intent of the present study is to empirically investigate the impact of the presence and absence of contextual clues with narrative and descriptive tasks on learners' performances in a classroom based assessment context. My first hypothesis was that both speaking tasks (i.e., narrative and descriptive) presented with clues will facilitate better development of content in test-taker's oral performance as compared to the ones presented without clues.

**Table 2. The Descriptive Statistics of Learners' Performances**

	N	Composite Score		Content Development	
		Mean	SD	Mean	SD
Descriptive (-)	15	29,00	5,78	6,26	1,53
Narrative (-)	16	29,37	5,27	6,37	1,20
Descriptive (+)	15	27,33	4,45	6,06	1,22
Narrative (+)	16	26,68	5,36	5,81	1,47

Table 2 illustrates the descriptive statistics of the scores assigned to the learners by two raters using the adapted version of IELTS speaking scale. In terms of the composite scores given by raters, learners got the highest mean scores ( $M=29,37$ ;  $M=29$ ) (out of 45) in the narrative tasks and descriptive tasks presented without contextual clues respectively, followed by descriptive and narrative tasks presented with contextual clues. With regard to the scores obtained from content development criterion in the scale, learners got the highest mean scores ( $M=6,37$ ;  $M=6,26$ ) (out of 9) in the narrative and descriptive tasks presented without contextual clues respectively, followed by descriptive and narrative tasks presented with clues.

To compare the scores from the content development criterion in the scale and the composite scores between experimental and control groups, an independent samples *t*-test was conducted. The results revealed that learners in experimental ( $M=27$ ,  $SD=4.87$ ) and control ( $M=29.19$ ,  $SD=5.43$ ) groups did not differ significantly in their composite scores,  $t(60)=1.67$ ,  $p>0.05$ . The results also showed that there was no significant difference between experimental ( $M=5.93$ ,  $SD=1.34$ ) and control ( $M=6.32$ ,  $SD=1.35$ ) groups in their scores from the content development criterion in the scale,  $t(60)=1.13$ ,  $p>0.05$ . These results show that learners in the experimental group who were given the test tasks with contextual clues obtained scores which were quite similar to those learners in the control group who were given the test tasks without contextual clues, which provides little evidence to support my first hypothesis.

My second hypothesis was that learners are expected to perform significantly differently on narrative and descriptive tasks presented with and without contextual clues. To determine if there were any significant differences in the learners' composite scores with regard to the effect of different task types, a one-way ANOVA was performed. According to the results of the test, the learners' composite scores did not differ significantly in the narrative and descriptive tasks with  $F(3,58)=.957$ ,

$p > 0.05$ . In addition, a *MANOVA* was employed to see if the learners' scores from each criterion such as "fluency", "lexical resource", "grammatical range and accuracy" and finally "content development" differ in terms of the narrative and descriptive tasks presented with and without contextual clues. The results showed that task types did not seem to affect the learners' "fluency" scores with  $F(3,58) = .037$ ,  $p > 0.05$ ; "lexical resource" scores with  $F(3,58) = .077$ ,  $p > 0.05$ ; "grammatical range and accuracy" scores with  $F(3,58) = .172$ ,  $p > 0.05$  and "content development" scores with  $F(3,58) = .159$ ,  $p > 0.05$ . Taken together, these results suggest that different task types do not significantly influence the learners' oral performance in a classroom based assessment context. Thus, this data fails to provide evidence for my second hypothesis.

### The Analysis of Learners' Performance Discourse

The learners' actual performance discourse was further analyzed to determine whether measures of content in terms of quantity in participants' output differ according to the tasks presented with and without contextual clues. As measures of content in terms of quantity, T-units and clauses were counted in the participants' oral performance data. Table-3 demonstrates the descriptive statistics of the numbers of T-units and clauses in the learners' oral performance.

**Table 3. The Descriptive Statistics of Learners' Performances**

	N	T-Unit		Clause	
		Mean	SD	Mean	SD
Control Group	31	11,83	4,25	14,93	5,80
Experimental Group	31	12,22	5,05	14,67	5,47

To compare the means of T-units and clauses between experimental and control groups, an independent samples *t*-test was conducted. The results of the test indicated that there was no significant difference between experimental and control groups with regard to the means of T-units and clauses with  $p > 0.05$ . Thus, it can be concluded that the presence and absence of contextual clues with the tasks did not affect the quantity of content conveyed by the learners.

Besides, a Pearson correlation coefficient was computed to assess the relationship between the number of clauses as a measure of content in terms of quantity and composite scores assigned to the learners by two raters. There was a positive correlation between the two variables,  $r = .556$ ,