T.C. YEDITEPE UNIVERSITY

INSTITUTE OF EDUCATONAL SCIENCES

Ph.D. IN ENGLISH LANGUAGE TEACHING

THE EFFECT OF EXPLICIT INSTRUCTION OF FORMULAIC SEQUENCES ON L2 ORAL FLUENCY

AYŞEGÜL NERGİS

2018

YEDİTEPE UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES DIRECTORATE

THESIS SUBMISSION and APPROVAL FORM

APPROVAL:

Doç. Dr. Hossein Farhady _____ (Advisor) _____ (Signature)

Prof Dr. Ayaes Alyel _____ (Member) _____ (Signature)

Dr. Öğr. Üyesi Evrim Evcyik-Aydın _____ (Member) _____ (Signature)

Dr. Öğr. Üyesi Sibel Tatar _____ (Member) _____ (Signature)

Dr. Öğr. Üyesi Adem SORUÇ _____ (Member) _____ (Signature)

SUBMITTED BY : Aysegül Nergis
DATE OF THESIS DEFENSE : 21.05.2018
DATE OF THESIS APPROVAL : 21.05.2018

CURRICULUM VITAE

EDUCATION:

B.A.: Marmara University, English Language Teaching, 2003

M.A: Marmara University, Primary School Teaching, 2008

Ph.D.: Yeditepe University, English Language Teaching, 2018

ACADEMIC POSITIONS

2018 – Present	Language instructor, School of Languages, Istanbul University.
2016 – Present	Sessional Lecturer, Department of English Language Teaching, Marmara University.
2016 – 2017	Language Instructor, School of Languages, Yeniyüzyıl University.
2012 – 2015	EAP instructor, School of Languages, Ozyegin University.
2008 – 2012	Research Assistant, Department of English Language Teaching, Yeditepe University.
2006- 2007	Sessional Language Instructor, Freshmen English Program, Marmara University.

TEACHING EXPERIENCE

Istanbul University

Foundation Program English (Levels: Preintermediate, Intermediate)

Higher Education Council Special Needs Group (Advanced level)

Marmara University (2016 -)

EGT3105 Research Methods

YDIO 111 Word Knowledge in English

Yeniyüzyıl University (2016 - 2017)

Foundation Program – Language Skills Instructor

ENG 151 English I

ENG 152 English II

Ozyegin University (2012 – 2015)

ENG 101 Academic Research and Writing

ENG 102 Academic Speaking

Yeditepe University (2008 – 2012)

EDEN 103 Oral Communication I – Teaching Assistant
EDEN 104 Oral Communication II – Teaching Assistant

Yeditepe University Teaching Certificate Program (2009)

School Experience

Marmara University (2006 – 2007)
English I and II (Undergraduate programs, Faculty of Education)

INTERNATIONAL PROJECTS

2007 – 2008 Teaching Assistant, Translator, Evaluator in Comenius Project: 226496-CP-AT-2005-1-COMENIUS-C21 / Research Title: "ENHANCING EARLY MULTILINGUALISM" (ENEMU).

2016 Adult Trainer in Erasmus Plus Youth Pass Project titled "Learning To Learn"

PUBLICATIONS

Nergis, A. (2015). Exploring the factors that affect reading comprehension of EAP learners. In H. Basturkmen (Ed.), *Critical concepts in linguistics: English for Academic Purposes (Volume 4),* (pp. 320-336). New York, NY: Routledge.

Nergis, A. (2013). Exploring the factors that affect reading comprehension of EAP learners. *Journal of English for Academic P*urposes, *12*,(1), 1-9. (Indexed in: Arts and Humanities Citation Index, Australian Educational Index, ERIC, Social Sciences Citation Index (*SSCI*)).

Nergis, A. (2011). Literacy culture and everchanging types of literacy. *International Journal of Online Educational Studies, 3* (3), 1133-1154. (Indexed in: Turk Egitim Indeksi, EBSCO, DOAJ)

Nergis, A. (2011). To what extent does neurolinguistics embody foreign language teaching methods?. *Procedia – Social and Behavioural Sciences 15*, 143-147. (Indexed in: SCOPUS: Elsevier)

Nergis, A. (2011). Foreign language teacher education in Turkey: A historical overview. *Procedia – Social and Behavioural Sciences 15*, 181-185. (Indexed in: SCOPUS: Elsevier)

Nergis, A. (2011). Grammar teaching beliefs of a pre-service teacher: A repertory grid analysis. *Sino-US English Teaching*, 8 (1) 1-14. (Indexed in: EBSCO, LLBA Database of Proquest, Chinese Database of CEPS, Chinese Scientific Journals Database)

REVIEWS DONE FOR INTERNATIONAL JOURNALS

System (indexed in SSCI and AHCI) (2016)

Language Awareness (indexed in SSCI and AHCI) (2015)

Asian EFL Journal (indexed in BEI) (March 2015 - present)

Studies in English Language Teaching (January 2013 – November 2016)

CONFERENCE PRESENTATIONS

Soruç, A., Nergis, A. & Genç, E. "Language Proficiency in Multimedia Learning "3rd Black Sea ELT Conference - Technology: A Bridge to Language Learning, OnDokuz Mayıs University, Samsun, 15-17 November, 2012.

Nergis, A. "A Comparison of Native and Non-Native University Instructors' Use of Discourse Markers in Lectures", 5th International Conference of English as a Lingua Franca, Bogazici University, Istanbul, 24-26 May, 2012.

Nergis, A. "A Criticism of EFL Curriculum Objectives: A Reconstructivist Approach", 3rd International Congress of Educational Research, Northern Cyprus, 4-7 May, 2011.

Nergis, A. "Investigating the Factors That Affect Academic Reading Performance", IATEFL ESP/EAP Conference, "EAP in University Settings: Teacher and Learner Competencies", Bilkent University, Ankara, 18-19 June, 2010.

UNPUBLISHED DISSERTATIONS

Nergis, A. (2018). *The effect of explicit instruction of formulaic sequences on L2 oral fluency. Unpublished doctoral dissertation*. Yeditepe University, Istanbul, Turkey (Advisor: Assoc. Prof. Hossein Farhady).

Nergis, A. (2008). *İlköğretim birinci sınıf öğrencilerinin ön okuryazarlık davranışları ile annelerinin okuryazarlığa olan ilgisi arasındaki ilişki*. Unpublished Master's Thesis. Marmara University, Istanbul, Turkey. (Advisor: Assist. Prof. Murat Aşıcı)

MEMBERSHIPS TO SCIENTIFIC ORGANIZATIONS

The International Higher Education Teaching and Learning Association (HETL)

European Association for Language Testing and Assessment (EALTA) – Special Interest Group: Speaking Assessment

SCHOLARSHIPS AND AWARDS

Full Scholarship – Yeditepe University (2008-2011)

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor Dr. Hossein Farhady who guided me through the completion of this dissertation.

I give my heartful thanks to Dr. Adem Soruç and Dr. Evrim Eveyik-Aydın for being loving friends and all-supporting colleagues for long long years.

Finally, I also would like to acknowledge Dr. Ayşe Semra Akyel and Dr. Zeynep Koçoğlu for their constant support, and Dr. Jingjing Qin and Dr. Belma Haznedar for giving valuable feedback to the earlier versions of my work.

I also would like to thank my family, friends and colleagues who supported me throughout the process.

To my dear parents, Nigar and Şaban Nergis…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Abstract

The role of formulaic language use in promoting L2 oral fluency is widely recognized. However, how instruction of formulaic language affects L2 speaking is still not clear. This study examined the effect of explicit instruction of formulaic sequences on L2 oral fluency. Two groups of English for academic purposes (EAP) students were assigned as experimental group and comparison group. The experimental group (N=18) received explicit instruction of formulaic sequences (FS) and the comparison group (N=19) received explicit instruction of academic vocabulary for 5 weeks. The target items in both groups were introduced through concordancing and practiced in communicative group activities and discussions in an EAP program. Pretest, posttest and delayed posttest scores of both groups were collected and they were interviewed about their opinions on the instruction they received on posttest. L2 oral fluency scores of the participants were evaluated according to Skehan's classification of oral fluency (Skehan, 2003; Tavakoli and Skehan, 2005): *speed fluency*, *breakdown fluency* and *repair fluency*. The effect of the instruction on L2 oral fluency was also adjusted for L1 fluency and L2 working memory capacity. Quantitative data were analyzed via Repeated Measures Multivariate Analysis of Variance (MANOVA) and Repeated Measures Multivariate Analysis of Covariance (MANCOVA). Results showed that there was a significant increase in pruned speech rate and articulation rate of the experimental group compared to the comparison group after receiving explicit instruction of FS and the long term effect of the FS instruction as compared to the academic vocabulary instruction was only observed in speed fluency. However, there was no significant relationship between the L2 oral fluency scores and the number of FS used in oral elicitation tasks on posttest, indicating that the increase in speed fluency cannot be attributed to the use of FS alone. It was also found that FS instruction did not have a significant effect on L2 oral accuracy of the experimental group, but it significantly facilitated one of the experimental group's L2 oral complexity scores (Mean segmental type-token ratio). On the other hand, the comparison group made a significant gain in another L2 oral complexity score (syntactic variety) after receiving academic vocabulary instruction. Qualitative data coming from semi-structured interviews showed that both groups expressed positive learning outcomes from the instruction through concordances. The experimental group specifically emphasized

the positive effect of the FS instruction on boosting their confidence and speed in their L2 oral performance.

ÖZET

Kalıplaşmış dil kullanımının ikinci dilde konuşma akıcılığını artıran rolü geniş kabul görmektedir. Ancak, kalıplaşmış ifadeler eğitiminin ikinci dilde konuşma akıcılığını nasıl etkileceyeği henüz açık değildir. Bu araştırmanın amacı, kalıplaşmış söz dizini eğitiminin ikinci dilde konuşma akıcılığı üzerindeki etkisini incelemektir. Akademik amaçlı İngilizce öğrencisi araştırma grubu ve karşılaştırma grubuna atanmıştır. Beş hafta boyunca araştırma grubu, açık öğretim ile kalıplaşmış söz dizini eğitimi almışken, karşılaştırma grubu açık öğretim ile akademik sözcük eğitimi almıştır. Her gruptaki hedef öğeler, bir akademik amaçlı İngilizce eğitimi programında concordance adlı sorgu yöntem ile tanıtılmış ve iletişimsel grup ve tartışma etkinlikleri ile pekiştirilmiştir. Her iki gruptan öntest, sontest ve gecikmeli sonttest puanları toplanmış ve aldıkları eğitim hakkında görüşleri hakkında sontestte yarı-yapılandırılmış mülakatlar yapılmıştır. Katılımcıların ikinci dilde konuşma akıcılığı puanları, Skehan'nın konuşma akıcılığı sınıflandırılmasına göre değerlendirilmiştir (Skehan, 2003; Tavakoli and Skehan, 2005): *konuşma hızı, duraklama akıcılığı* ve *tamir akıcılığı*. Verilen eğitimin ikinci dilde konuşma akıcılığı üzerindeki etkisi ayrıca birinci dilde konuşma akıcılığı ve ikinci dilde çalışan hafıza kapasitesi açısından kontrol edilmiştir. Niceliksel veriler Tekrarlı Ölçümler için Çoklu Varyans Analizi (MANOVA) ve Tekrarlı Ölçümler için Çoklu Kovaryans Analizi (MANCOVA) ile analiz edilmiştir. Sonuçlar, kalıplaşmış söz dizini eğitimi aldıktan sonra araştırma grubunun toplam akıcılık puanında ve seslendirme hızı puanında, karşılaştırma grubuna göre anlamlı bir artış görülmüştür ve verilen eğitimin uzun süreli etkisi ise, yalnızca konuşma hızı puanında gözlemlenmiştir. Ancak, sontestteki ikinci dilde konuşma akıcılığı puanları ile konuşma örneklerinde kullanılan kalıplaşmış söz dizini sayısı ile konuşmada akıcılık puanları arasında anlamlı bir ilişki yoktur, bu da hız akıcılığındaki artışın yalnızca kalıplaşmış söz dizini kullanımı ile açıklanamayacağına işaret etmektedir. Araştırmada ayrıca, kalıplaşmış söz dizini eğitiminin araştırma grubunun ikinci dilde konuşma doğruluğu üzerinde anlamlı bir etkisinin olmadığı, ama araştırma grubunun ikinci dilde konuşma karmaşıklığı puanlarından birini (Tip-Belirteç Oranı Kesitsel Ortalaması) anlamlı şekilde arttırdığı da bulunmuştur. Diğer taraftan, karşılaştırma grubu, akademik sözcük eğitimi aldıktan sonra, ikinci dilde konuşma karmaşıklığı puanlarından birinde (sözdizimsel çeşitlilik) anlamlı bir kazanım elde etmiştir. Yarı yapılandırılmış görüşmelerden

toplanan niteliksel veriler her iki grubun da concordance adlı sorgu yöntemi ile eğitim almanın olumlu sonuçlarından bahsettiğini ortaya koymuştur. Özellikle araştırma grubu, kalıplaşmış söz dizini eğitiminin özgüvenlerini ve ikinci dilde konuşma performanslarını arttırdığını vurgulamıştır.

1. Introduction

1.1. The Rationale of the Study

Considering the role of English as the most widely used language in the business and academic world, being able to communicate in a second or foreign language has become the primary objective for millions of language learners around the world. Today, speaking fluently is a significant goal for English as foreign language speakers due to the role of English as the medium of communication in business and academic world, tourism and information sciences. Consequently, speaking has also become one of the major factors in evaluation of L2 competence (Lennon, 1990; Riggenbach, 1991) and has been placed in the core of most ELT programs around the world (Mota, 2003). However, lack of systematic research on L2 speaking leads to poor learning, teaching, understanding and testing of L2 speaking (Mota, 2003, p. 71).

There has been a growing interest in research on L2 oral performance; however, L2 oral fluency remains as one of the most under-researched L2-related phenomena. Our knowledge on how to improve EFL programs that will enhance L2 oral performance is even more limited. In fact, there are only a limited number of instructional studies on the factors affecting L2 oral fluency. However, L2 oral fluency has a key importance for language teachers since it is often associated with L2 proficiency (Derwing, 2017; Fulcher, 2003; Stengers, Boers, Housen and Eyckman, 2011; Wood, 2012). For this reason, there is also need for more research on L2 oral fluency to expand our understanding of L2 development, from the early stages of L2 learning to the stage of effortless use of L2 in communication.

According to Derwing, Rossiter, Munro and Thomson (2004), fluency instruction can be helpful for all levels of proficiency. They also emphasized that future research can focus on to what extent listeners are influenced by other factors that contribute to the perception of fluency, such as lexical choice and the use of formulaic sequences. Other researchers also suggested that high level L2 learners can benefit from L2 fluency instruction and that fluency training could take a greater proportion of teaching time (Nation & Newton, 2009). For example, Faerch and Kasper (1983, p. 235) suggested that teaching advanced L2 learners to produce filled pauses in real-life language use and other fixed expressions can improve their

fluency. According to Guillot (1999), of all levels of L2 proficiency, it is advanced learners that are open to mastering fluency after developing formal aspects of communication, i.e. accuracy. In recent years, several researchers suggested that use of formulaic language helps ESL learners to be more proficient L2 speakers (Nesselhauf, 2003; Pawley & Syder, 1983) and enhances their L2 oral fluency (Boers, Eyckmans, Kappel, Stengers and Demecheeler, 2006; Pawley & Syder, 1983; Wood, 2004, 2009, 2012, 2016; Wray, 2000; 2002).

Native English speakers' language production is majorly formulaic (Sinclair, 1991). Pedagogical value of formulaic language is also widely recognized in that they are discussed to help L2 learners to come across as proficient speakers (Pawley & Syder, 1983) and to influence the perceptions of listeners (Boers et al., 2006). As a result of this, Lewis (1993), Willis (1990) and Wray and Perkins (2000) suggest presenting a large repertoire of formulaic language to L2 learners to increase their L2 oral and perceived fluency. In a similar vein, Ellis, Simpson-Vlach and Maynard (2008) and Wray (2000) emphasize the need to familiarize L2 learners with formulaic expressions because nonnative speakers do not have regular access to authentic language use, which in return causes dysfluences in speech. However, there are no experimental or instructional studies that investigated the effect of formulaic language instruction on fluency development (Davies, 2014; Zhang, 2017). The present study set out to fill this gap.

In this study, it is hypothesized that explicit instruction of formulaic expressions would significantly increase EAP learners' oral fluency. In accordance with the background information given above, the present study adopted a quasi-experimental design to test the effectiveness of FS instruction on L2 oral fluency. The items used in the instruction were selected from two spoken academic corpora and presented to the students in concordances. The purpose of this procedure was to present the target items in their original context to make learning of academic spoken chunks more meaningful to the students. The participants' oral production was collected on pretest, posttest and delayed posttest to examine their progress over time.

An instructional study that aims to test the effect of teaching formulaic sequences would call for an efficient way of introducing and practicing the target items in a formal context. The challenge of preparing such a program is that the target

items to be used in the instruction should be selected systematically. The instruction would need to include the most frequently used items in the given context. This challenge can be overcome by selecting the items from a relevant corpus. The next challenge would be to select the best method of presenting and practicing the target items. With advanced L2 learners, explicit instruction is discussed to be the most efficient teaching method (Roehr-Brackin, 2014), specifically for the context of formal language instruction (Nassaji & Fotos, 2007). Accordingly, the target items used in this study were systematically elicited from two spoken academic corpora.

Accuracy and complexity are the other aspects of speech in Skehan's classification of oral performance (Skehan, 2003; Tavakoli and Skehan, 2005). It is widely discussed (Ellis & Barkhuizen, 2005; Skehan, 1996; Yuan and Ellis, 2003) that these three aspects of speech tax each other during oral performance; but together they constitute the construct definition of L2 speaking (Fulcher, 2003). Considering the interactive effect of three aspects of speech on L2 oral performance, the present study will examine oral accuracy and complexity of the participants in addition to their oral fluency.

L2 oral fluency has also been discussed to be linked to L1 fluency and working memory capacity. According to Segalowitz (2010), L2 speakers cannot reach the same level of fluency in their L2 as in their L1. For this reason, there are large differences of L2 fluency between EFL speakers. Mizera (2006) asserts that what causes dysfluencies in speech are due to limited capacity of the working memory. Due to their roles in L2 oral fluency and/or dysfluency, L2 working memory capacity and L1 fluency levels of the participants were adjusted for their effect in statistical data analyses in this study.

1.2. The Purpose of the Study

Considering the need for more research on teaching of FS in L2, the present study aims to investigate the effects of explicit instruction of FS on L2 oral fluency in a real classroom setting. In the last two decades, there has been a considerable increase in research regarding L2 oral fluency. Only a very small number of research studies dealt with the effects of instruction of formulaic sequences on L2 oral production. Additionally, the role of formulaic language use in enhancing L2

proficiency is widely recognized. However, the hypothesis that explicit teaching of formulaic expressions would increase L2 oral fluency has never been examined under experimental conditions.

It was hypothesized in this study that focused instruction of formulaic sequences will help to improve the participating freshmen students' L2 oral fluency. In accordance with the aims of the study mentioned above, in the present study, the following research questions were formulated:

1.What is the effect of explicit instruction of formulaic sequences on L2 oral fluency?

1. a. Does the effect of the explicit instruction of formulaic sequences change over time?

2. Is there a relationship between the number of formulaic sequences produced by the participants and their L2 oral fluency scores?

3. What is the relationship between fluency, accuracy and complexity of the participants' L2 oral performance after the treatment?

3. a. What is the effect of explicit instruction of formulaic sequences on L2 oral accuracy?

3.b. What is the effect of explicit instruction of formulaic sequences on L2 oral complexity?

4. What are the perceptions of the participants about the effectiveness of the treatment?

1.3. Significance of the Study

Although the premise that L2 oral fluency of advanced learners can be enhanced through instruction of formulaic language has gained some interest among SLA researchers, there is still no empirical evidence that could support this claim. Without concrete evidence coming from experimental studies, it is difficult to determine how to design such programs and how to select and give priority to the

target items in advanced level curricula that aim to enhance oral fluency. Previous studies on L2 oral fluency were not conducted in classroom setting with intact groups of learners; therefore, it is difficult to generalize their findings to a larger group of learners. However, many instructional implications were drawn from these studies, which highlights the need for more systematic inquiry coming from genuine classroom setting. Moreover, there are only a limited number of studies on teaching of formulaic language to EAP students. In fact, in SLA research, L2 oral performance of advanced language learners is a highly neglected topic even though it is a widely recognized and highly desired goal for EFL instruction.

Recently, Thomson, Boers and Coxhead (2017) called for replication studies that could enlighten the effect of instruction FS on L2 oral fluency. They specifically called for replication studies for the two following studies: 1) Boers et al. (2006) that investigated the effect of FS instruction on perceptions of L2 oral performance, and 2) Wood (2009) that investigated a case of one Japanese learner of English. However, neither of these two studies had a systematic methodology that specifically looked into the effect of teaching formulaic sequences on L2 oral performance. The present study aims to fill this gap. Additionally, as also argued by Lahmann, Steinkrauss and Schmid (2017), L2 fluency development of higher proficiency speakers is also neglected. The present study also aims to address this gap.

To sum up, one major contribution of this study is that it is the first systematic investigation of the effect of explicit instruction of formulaic sequences on L2 oral fluency in an experimental design with Advanced level EAP students.

Another contribution of this study will be examining the increase in quality of speech in terms of complexity and accuracy after receiving formulaic sequences instruction. There have been only a few instructional studies that has looked into this gap before.

1.4. Limitations of the Study

The present study acknowledges certain limitations resulting from the experimental nature of the research design. First of all, the sample size was not large, which affects generalizability of the findings to a larger context. The instructional

context of the study was limited to two groups of advanced EAP students in one language school, which also affects the external validity of the findings.

The present study also acknowledges the limitations of quasi-experimental research design. In the present instructional context, it was not possible to randomly assign the participants to experimental group and the comparison group. Hudson and Llosab (2015) suggest paying attention to preexisting source of variation between the experimental and comparison groups in quasi experimental studies. Therefore, a Kruskall Wallis H test was conducted on pretest scores. The results showed that the L2 oral fluency levels of the two groups were not significantly different before the instruction.

Another limitation of the study concerns the number and nature of the adjusting variables. L2 oral fluency is considered to be in association with many variables such as cultural factors and willingness to communicate in L2 (Wood, 2016), L2 working memory (Mizera, 2006) and L1 fluency (Segalowitz, 2010). In this study, L2 oral fluency scores of the participants were adjusted for working memory capacity for L2 and L1 oral fluency since these two factors were more widely and more frequently associated with L2 oral fluency related phenomena by L2 fluency researchers.

1.5. Definitions of Key Terms

Construct definitions of some of the significant terms are given below. Operational definitions are defined in detail in Methodology.

Second language (L2) fluency: “An aspect of L2 use that emerges from the complex interplay among motivation, social context of communication, neurocognitive basis of speech production and cognitive experiences” (Segalowitz, 2010, p.28)

L2 oral fluency: “the learner’s capacity to mobilize an interlanguage system to communicate meaning in real time” (Skehan, 1996, p. 46)

Accuracy: “a learner’s capacity to handle the currently attained linguistic system” (Skehan, 1996, p. 46)

Complexity: “elaboration of the underlying interlanguage system” (Skehan, 1996, p. 46).

Formulaic sequences: “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2002, p. 9).

Academic formulaic sequences: frequent recurrent patterns that have clearly defined pragmatic functions and pedagogical value for English for Academic Purposes curricula (Simpson-Vlach and Ellis, 2010).

1.6. Acronyms

AWL (Academic Word List)

EAP (English for academic purposes)

EFL (English as a foreign language)

ESL (English as a second language)

FS (Formulaic sequences)

L1 (First language)

L2 (Second language)

MI (Mutual information)

LTM (Long term memory)

NNS (Non-native English speakers)

NS (Native English speakers)

WM (Working memory)

2. Review of Literature

2.1. Introduction

It is a need in the contemporary world to be proficient in more than one language due to growing global scope of communication and economic integration. Most people in the world know more than one language; however, most speakers are not able to use a second language (L2) as skillfully and fluently as their first language (L1). According to Segalowitz (2010, p.2), researchers looking into ways to improve fluency of second language (L2) speakers need to understand why this fluency gap exists between the L1 and L2. The first problem that needs to be examined is the problem of individual differences in L1 speaking. Although a fluent speaker in L1, an L2 speaker may find it difficult to develop fluent L2 speaking performance. Another problem to deal with is the individual differences in L2 speaking. Each speaker is unique in terms of their rate of speech and their word choices (Osborne, 2011). Even after years of studying, some speakers may fail to be fluent speakers in L2. Research looking into the ways of enhancing L2 oral fluency should consider these within and between individual differences among L2 speakers (Gatbonton & Segalowitz, 2005; Segalowitz, 2010).

Another issue related to L2 speaking that has not been efficiently examined is the linguistic items that maximize oral fluency. In fact, research investigating L2 oral fluency is quite rare and there is need for empirical evidence that could suggest what kind of fluency maximizing activities and linguistic items should be implemented in language instruction programs to improve L2 oral fluency. In order to extract the linguistic structures to focus on in such fluency-enhancement programs, the nature of language processing by native and nonnative speakers should be examined. There has been considerable debate about the effect of use of formulaic expressions in speech on fluency of speech (Lewis, 1993; Nattinger & Decarrico, 1992; Wray, 2002); however, there is no concrete evidence coming from instructional studies that explore whether teaching formulaic language can improve speaking performance in L2.

In fact, a considerable proportion of language is known to be formulaic (Conklin & Schmitt, 2008; Jones & Haywood, 2004), so the use of formulaic language should lead to enhanced fluency (Wray, 2002). Due to recent realization of the significant implications of this idea, several researchers in the field of applied linguistics and language teaching have focused on the importance of lexical bundles, collocations and other idiomatic expressions in communication in a second language and examined the use of formulaic language from a myriad of perspectives. While some researchers have focused on the use of such word combinations, i.e. formulaic sequences in written products (Nesselhauf, 2003; 2005), some other studies have focused on the role of formulaic sequences (henceforth, FS) in reading (Hsu, 2010; Jones and Haywood, 2004) and some others have looked into their role in spoken language (Boers, et al., 2006; Qi & Ding, 2011; Stengers et al., 2011). Biber et al. (1999) showed that English lexical bundles can be found in 28 % of English spoken academic discourse and 20 % of English written academic discourse. In another study, Foster (2001) found that 32.3 % of unplanned native speech was made up of formulaic sequences. Erman and Warren (2000) also found that 52.3 % of spoken English discourse and 52.3 % of written English discourse was made up of formulaic language. In a recent study, Gholami, Karimi and Atai (2017) found that around 38% of focus on form episodes (i.e. deliberate attention on form during speech) in academic communication in a university classroom is formulaic and they are made up of collocations, lexical bundles and idioms. Although the aims of these studies remarkably varied, they all contribute to our understanding of the significance of formulaic sequences in discourse. For this reason, researchers agree that formulaic sequences have an important role in L2 teaching as well (Lewis, 1993; Nesselhauf, 2003).

2.2. Formulaic Sequences

The definition of formulaic expressions is problematic, because there is no consensus on how to label them (Wray & Perkins, 2000). A plethora of different labels have been used in the literature to define fixed expressions depending on the ways of arriving at them: lexical bundles, formulaic sequences (Conklin & Schmitt, 2008), conventional expressions (Bardovi-Harlig, 2009, Bardovi-Harlig & Vellenga, 2012), chunks (Sinclair, 1991), semantic formulas (Bardovi-Harlig & Griffin, 2005), lexical phrases (Nattinger & DeCarrico, 1992).

Researchers and theoreticians find it difficult to define FS, since FS contain a wide variety of linguistic items, such as collocations (e.g., *heavy air*; *tell a secret*), idioms (e.g., *kick the bucket*; *break a leg*), similes (e.g., *busy as a bee*; *free as a bird*), and discourse organizers (e.g., *on the other hand*; *on the contrary*). While some FS are completely fixed expressions (e.g., *in my opinion*; *in my experience*), some may have possible varieties, showing syntagmatic expansion (e.g. *in a* highly *similar manner*) or paradigmatic substitution (e.g. *carry out/conduct experiment*; *getting off the topic/subject*) (Stengers, et al., 2011). In all of these variations, some FS are truly fixed in the morphological sense even though not grammatical in form (e.g., *on the other hand*), some FS allow variances in accordance with grammar rules (e.g., *conduct, conducts/conducted (an) investigation/investigations*).

According to Weinert (1995), researchers may label formulaic expressions in a different ways, but, in fact they have the same concept in their minds. In recognition of the complexity of this problem, Wray and Perkins (2000) listed the terms used in the literature to describe formulaic sequences and formulaicity (see Table 1 below). The length of the list demonstrates the degree of the problem of terminology surrounding formulaic language.

Table 1
Terminology of Formulaic Langauge

Amalgams	Gambits	Praxons
Automatic	Gestalt	Preassembled speech
Chunks	Holistic	Prefabricated routines and patterns
Clichés	Holophrases	Ready-made expressions
Co-ordinate constructions	Idiomatic	Ready-made utterances
Collocations	Idioms	Rote
Composites	Irregular	Routine formulae
Conventionalized forms	Lexical(ized) phrases	Schemata
Fixed Expressions	Lexicalized sentence stems	Semi-preconstructed phrases that constitute single choices
Formulaic language	Multiword units	Sentence builders
Formulaic speech	Non-compositional	Stable and familiar expressions with specialized subsenses
Formulas/formulae	Non-computational	Synthetic
Fossilized forms	Non-productive	Unanalyzed chunks of speech
Frozen phrases	Petrification	

Adapted from: Wray, A. & Perkins, M. R. (2000, p. 3).

A comprehensive definition of FS was done by Wray (2000) as follows:

> "A sequence, continuous or discontinuous of words or other meaning elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from the memory at the time of us, rather than being subject to generation or analysis by the language grammar" (Wray, 2000: 465).

According to Wray's definition, FS are prefabricated, so they do not allow for creativity, or generating variances. However, Nattinger and DeCarrico (1992) disagree with this idea and assert that FS are not always conventional units, they allow for variations. They offered two principles that determine the degree to which lexical phrases can be flexible or conventionalized: 1) syntagmatic simplicity, and 2) paradigmatic flexibility (Nattinger & DeCarico, 1992, p. 17). With the help of these principles, it is easy to understand which forms are mostly preferred in pragmatic acts: speakers usually prefer to use syntactically simple, but lexically flexible composites. For example, 'Modal + you + VP' is a syntactically simple composite; however, it allows several lexical variations such as '*could you please pass the salt?*', *'could you hand me that pencil?', 'would you lend me a dollar?'* in that the modal and the verb can be filled with several different linguistic items (also given in Nattinger & DeCarrico, 1992, p. 18). Their model of defining FS allows for a great variation, i.e. creativity on the speaker's part, as can be seen in this example. Addition of different Modals (*would*, *can*) and adverbials (such as *please*, *kindly*) also affects the number of FS that can possibly be created from a simple composite.

Wray and Perkins (2000) also discussed the definition of FS to be along a continuum of specificity and flexibility. On this continuum, at the one end, tightly idiomatic expressions exist (such as *by and large, by way of, on the other hand*); and, at the other end, more flexible formulas exist (for eg., *'so/very' sorry to +keep+'TENSE' + you/him/her+ waiting*).

As can be seen from these discussions about what constitutes FS, it is a challenge to prepare a list of FS to make use of in speaking classrooms. With this challenge in mind, the present study will draw on both corpus-based frequencies and native and non-native English speaking language instructors' judgment to elicit a list of FS that can be taught through explicit instruction in an academic setting. The following procedures were followed to elicit the target FS: First, seven popular course

books focused on L2 speaking were scanned to elicit lists of FS to be used in the focused instruction. Second, each FS in this list was checked on Michigan Corpus of Academic Spoken English (MICASE) and British Academic Spoken English (BASE) for their actual frequency in academic spoken discourse. Finally, the FS that were found in both corpora were judged by two native and four non-native English speaking language instructors to explore whether they are worth teaching or not. The rationale behind using both corpus-based decisions and teacher judgment is that using only corpus-based decisions would yield dysfunctional lists of FS to be implemented in an instructional context and using only teacher judgment would cause this experimental study to orient around assumptions rather than systematic methods.

In terms of teaching FS, there are a few suggested methods in the literature (for eg., Lewis' Lexical Approach). In terms of second language teaching, researchers are interested in formulaic language use, because it is a significant part of proficient L2 speakers' competence (Bahns, 1993; Nesselhauf, 2003). According to Wray (2000), FS help manipulating information, help buying time for processing, enhancing processing and organize discourse. FS can also be produced without pauses, contributing to the "smoothness" of speech (Wood, 2006), so they lead to less pauses in speech and longer fluent runs (Wray, 2002). FS are also believed to help the second language (L2) learner to be more accurate in L2 as well as more fluent (Wray, 2002).

However, there are not so many experimental studies to confirm these assumptions; especially instructional studies dealing with this phenomenon are quite rare. In a 2009 study, Wood investigated the effects of focused instruction of formulaic sequences and fluency on the performance of one Japanese learner of English in spontaneous narratives in English. The fluency workshop was implemented for six weeks. The Japanese student in the study listened to native speaker recordings, worked on the transcriptions of the recordings, recording and listening to his own speech and joining free talk circles. In the end, Wood observed fluency gain in the participant's speech sample along with more extensive and complex use of FS. One major limitation of this study was that it did not implement systematic instruction of formulaic language; the instruction focused merely on awareness raising and noticing activities. Another limitation was that it reported the results of a case study. Results coming from a larger group of learners may enlighten our inquiry of whether explicit

teaching of FS can help advanced level students to speak more fluently. In an experimental study, Boers et al. (2006) implemented awareness raising activities with a group of sophomore students studying modern languages. The experimental group received a focused instruction on FS, whereas the comparison group received traditional grammar-lexis activities. However, this study also had a major limitation related to the instruction of FS (=experimental group). The instruction was only focused on phrase-noticing exercises in which the students were instructed to pay attention to syntagmatic dimension of vocabulary. The results showed that FS helped the experimental group students to be perceived more proficient L2 speakers by raters. The limitation of this study was lacking an extensive FS instruction and a clear description of procedures for selecting FS is vague and non-systematic.

2.2.1. The Use of Formulaic Sequences in L1 Speech

Gaining proficiency in L2 is a difficult task, since it requires much effort and time to become sensitive to the majority of the target language forms. Even proficient L2 learners may have difficulties keeping up with the arbitrary regulations of L2 (Nesselhauf, 2003). It seems like an impossible task to be ready to learn, process and use all possible variations of FS in L2 even for very advanced learners. Thus, in order to deeply comprehend what to expect from L2 speakers in terms of the use of FS in L2, we should look into how native speakers cope up with FS in their L1.

As described above, prefabricated chunks are "stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 2002, p. 9). The description of lexical chunking indicates that there are processing advantages of using chunks, and the ability to rely on them is one factor that allows native speakers to be fluent (Schmitt, 2000; Siyanova-Chanturia and Martinez, 2015).

Indeed, research suggests that native speakers possess intuitive power to recognize and store FS in their L1. Ellis et al. (2008) conducted a series of experiments with small groups (N= 6-18) of native English speaker (NS) and non-native English speaker (NNS) participants. Both groups were instructed to read 54 written and 54 spoken academic formulas. In the first experiment, the participants judged whether the presented words were likely to be a formulaic expression in

English. The experiment 2, the participants were instructed to read aloud formulaic strings. The pause between the onset of the written string and the beginning of the spoken response was recorded for each participant. In the third experiment, the participants were asked to read aloud the final word of the each given string as quickly as possible. The results of the experiments showed that native speakers and English as a second language (ESL) learners were sensitive to different metrics (Ellis et al., 2008, p. 386). For NS what they perceived a formulaic expression depended on the Mutual Information (MI) of the given string of words. MI refers to the association between the words in a string. Higher MI items have a greater coherence than is expected by chance regardless of their frequency and literal meaning. For NNS, on the other hand, frequency of FS in the input was more useful for articulation speed.

These findings suggest that different parameters are important for NS and NNS: Input is paramount for NNS. In order to encode linguistic units as formulaic expressions, they need to see and hear them in different contexts. However, it is different for NS: They are inherently "attuned to these constructions as packaged wholes" (Ellis et al., 2008, p. 391). So they intuitively recognize high MI items as formulas. According to Ellis et al., these results suggest that, even after over 10 years of experience in L2 learning, there is a great gap between NS and NNS L2 learners in terms of FS storage which might explain why NNS are generally perceived as non-fluent speakers compared to NS. This may indicate that not only the amount of input but also how FS are taught has a significant role in detecting how L2 learners use FS in their speech.

Wray (2000) listed the main benefits of FS for speakers' productions as:

1. Manipulates information
2. Buys time for processing
3. Creates shorter processing route
4. Organizes the discourse

According to Wray (2000) the benefits of FS for listener comprehension are as follows:

1. Gets the hearer to do things

2. Indicates the speaker's individual identity
3. Indicates the speaker's group identity
4. Organizes the discourse

As can be seen, according to Wray, FS serves well both to the speaker's production and the hearer's comprehension. Indeed, FS specifically discourse markers makes it "easier to sequence the ideas fluently, and, simultaneously, signal to the hearer where it would be most appropriate, and inappropriate, to begin a turn and what the overall character of the speaker's message is" (ibid, p. 478).

SLA researchers have demonstrated considerable amount of focus on the importance of teaching FS in L2 classrooms (Wood, 2002, 2004, 2012; Wray, 2000; Wray and Perkins, 2000). The following chapter will look into this point in detail.

2.2.2. Instructional Studies on Formulaic Sequences

In earlier studies, there was a debate about whether to teach FS. According to Krashen and Scarcella (1978), chunk learning plays only a very small role in language acquisition because language use in real world conditions requires speakers to use the target language in creative constructions. However, this is not so true for some other researchers. For example, Nattinger and DeCarrico (1992) disagree with Krashen and Scarcella, arguing that the language used by people in every day is quite formulaic and predictable, and also they usually encounter the target language forms in routine situations. As also pointed out by Nattinger and DeCarrico (1992, p. 27), not only linguistic behavior, but also all types of human behavior and learning are predictable and ritualized. Wray (2013) also advocates the need for teaching FS considering their role in communication, but emphasizes that there is not consensus on how to teach FS effectively and which FS should be prioritized in instruction. According to McCarthy (2010), FS has received little attention in research on L2 vocabulary learning. Similarly, Siyanova-Chanturia (2017) emphasizes the role of L2 vocabulary in L2 pedagogy and the central role of multi-word units (i.e. formulaic sequences) in L2 vocabulary instruction. She also remarks that although there has been a growing interest in research on knowledge and processing of FS; however, there is still need for more pedagogical research on FS.

Overall, most studies dealing with L2 pedagogy and formulaic language are not instructional studies, but studies that aim to explore FS used in spoken academic context (Sanchez Hernandez, 2013) and demonstrate how FS are used for discourse signaling across different stages of university lectures (Csomay, 2013; Nesi and Basturkmen, 2006) and how spoken and written FS in academic context differ from each other (Biber, Conrad and Cortes, 2004), indicating the need for purely instructional studies on FS.

Today, SLA researchers agree that knowledge and use of FS is highly important for L2 competence (Fan, 2009; Nation & Newton, 2009; Schmitt, 2000; Spöttl and McCarthy, 2004; Wood, 2002, 2012, 2016; Wray, 2000; Wray and Perkins, 2000). Educational researchers also recognize the significance of FS for L2 speakers and suggest new ways to integrate them into L2 curricula (Lewis, 2000; Nattinger & DeCarrico, 1992; Wood, 2009, 2010). A myriad of studies compared FS use by NS and NNS speakers (Bardovi-Harlig, 2009; Boers et al., 2006; Conklin & Schmitt, 2008; Ellis et al., 2008; Jiang & Nekrasova, 2007; Wullf, 2008). In general, the results of these studies indicate that NS outperform NNS in terms of FS use. So, if the purpose of L2 instruction is to help learners to gain L2 competency, then it is imperative for L2 instructors to integrate FS instruction into their syllabuses (Nattinger & DeCarrico, 1992). According to Wood (2002), today, FS play an important role in EFL curriculum. He bases his ideas on the following premise:

> "if formulaic sequences are a key element of natural language production, it would seem that a large amount of exposure to natural, native-like discourse, be it oral or written, would be an important part of a pedagogy designed to promote their acquisition" (p. 9).

FS also have an important role in pragmatic use of L2. Widdowson (1989) assigned an important role on 'formulaic chunks' in his discussion of pragmatic competence. He suggested that communicative competence was not merely knowledge of grammar rules and applying them; it was rather knowledge of a stock of linguistic patterns and knowing how to apply these patterns in accordance with situational demands. So, just like Nattinger and DeCarrico (1992), his understanding

of formulaic language use allowed for adaptation into new contexts as it is for native speakers in the course of their daily language use.

Teaching FS is also important for advanced L2 learners to develop their pragmatic skills. For example, Bardovi-Harlig and Griffin (2005) found that high intermediate learners already possess a good level of pragmatic skills. The participants of this study were 43 ESL learners from different language background. They worked in pairs to watch videos of different scenarios in which the speakers could not solve pragmatic problems. Following that, the students were instructed to remedy the communication problems they just watched through role plays. The results of the study showed that these awareness raising activities involving recognition of appropriate semantic formulas were useful for them to become more competent speakers of L2.

In a 2012 study, Bardovi-Harlig and Vellenga investigated the effects of instruction type on L2 oral production of FS. In this study, ESL learners ranging from beginner to upper-intermediate received focused instruction on formal (=experimental group) and informal expressions (=comparison group). The results of the study showed that learners benefited from "all-purpose" instruction more than specific ones. The authors suggested that, in order to improve their students' repertoire of fixed expressions, teachers should involve more activities that require production on their students' part. As can be seen from the implications of the mentioned studies above, there is growing evidence that indicates the role of focused instruction of FS in determining pragmatic skills of L2 learners. Hence, if ESL teachers wish their students to develop in pragmatic use of L2, they should involve awareness raising activities and oral production in classroom as much as possible. In a similar vein, Ding (2007) discussed the emphasis on FS by the Chinese foreign language teaching circle because repetition and memorization is highly valued in their educational system and praised as a method of accurate learning.

Although there is a growing recognition of the significant role of FS in L2 pedagogy, some researchers believe that FS are difficult for L2 learners to master due to "poverty of learner experience" (Wray, 2000, p. 468), which means that L2 learners usually do not receive enough input to acquire them properly and they are often not

taught so well. Even though a difficult procedure, various approaches to how to teach FS have been offered in the literature.

In a 2000 article, Wray introduces three approaches of how to teach FS. The first one of these approaches was proposed by Nattinger and DeCarrico (1992). In their teaching model, the focus is on the usefulness of FS in teaching conversation, so their approach is mostly pragmatic. Their model starts teaching chunks as fixed routines, and after that, introducing possible variations, which allows the learner to analyze the patterns further. The second model was formulated by Lewis (1993). According to Lewis, it is beneficial for learners to be familiarized with a very large amount of particularly institutionalized linguistic items. The last model of teaching FS was postulated by Willis (1990). Willis proposed that teachers should provide the learners the most useful patterns of the target language, so they should start with teaching the most useful words in the language. Millar (2011) also confirmed that native speakers felt disturbance while reading the malformed expressions produced by non-native speakers, which indicates the importance of encouraging EFL students to produce the correct forms of expressions for the sake of intelligibility.

These three models of teaching FS were reviewed by Wray (2000); more than a decade has passed since then. With recent technological advancements, corpus linguistics has gained popularity and gained concrete ground in vocabulary research as well. According to Fan (2009), the study of fixed expressions such as collocations has become a statistical matter; in addition to this, concordances can allow the L2 learner to learn in authentic and multiple contexts. In a manner of speaking, the direction and means of research in FS have radically changed for the last 20 years or so with new educational and research technologies. Research dealing with vocabulary today primarily involves the use of corpus-based data and concordance programs.

Although corpora research has developed rapidly in the last few decades, there is still need for studies that will help identify essential formulaic academic vocabulary (Simpson-Vlach and Ellis, 2010; Martinez, 2013; Martinez and Schmitt, 2012), specifically spoken academic vocabulary (Simpson et al. 2002; Gardner and Davies, 2014).

There are only a limited number of studies that investigated the list of formulas to teach in academic contexts. For example, Ellis et al. (2008) reported a series of three consecutive corpus-based experiments in order to define Academic Formulas List (AFL). AFL is a pedagogically useful list of FS that are most frequently used in spoken academic discourse. In the first experiment, 11 NS and 11 NNS participants were asked to judge whether visually presented word strings were likely to be found in English or not. In the second experiment, 6 NS and 6 NNS were instructed to read aloud given FS. The participants' response time (the pause between the onset of the written string and the beginning of the spoken response) and articulation time were calculated. In the third experiment, 18 NS and 16 NNS participants were instructed to sit in front of a screen and to read aloud the final word of each string that appeared there as soon as possible. The results of the experiment 1 and 2 showed that for NNS the frequency of the string was what determined their voice onset time. However, for NS, a different element was a more determiner of voice onset time: Mutual Information (MI) value of the strings. MI refers to how much linguistic items co-occur when used in a string. A lower MI score means that the string of words co-occurs together more likely due to chance, while a higher MI score indicates a stronger association between the words in the string. Overall results of the study suggest that NS possess the intuitive ability to extract co-occurrence information that shapes formulaic expressions. In the mentioned study, Ellis et al. (2008) also attempted to formulate *formula teaching worth* (FTW) measure. FTW describes what experienced English for academic purposes (EAP) instructors think about given formulaic expressions: are they worth teaching at all? This question is important since frequency-based list of formulas elicited with the help of corpus analyses do not always include items that would make sense to NS. In their 2008 study, Ellis et al. selected 54 academic formulas from spoken corpus and 54 from written corpus. They asked experienced EAP instructors and language testers to rate these formulas in terms of their potential for teachability on a scale of 1 (disagree) to 5 (agree). The EAP instructors and language testers were asked whether they perceive these 108 items as formulaic sequences, whether they thought the phrase had a cohesive meaning or function, as a phrase and whether they thought the phrase was worth teaching. The participants had the highest intercoder reliability coefficient (Cohen's k) in the last question (k= 0.83). It means they were mostly in agreement about the pedagogical value of the given list of FS. Ellis et al. (2008) remarked that this FTW index could

help EAP curriculum developers to come up with a better classification of FS that should be included or discarded in EAP curricula.

To summarize the implications of all the research findings cited above, the reasons for teaching FS in EFL classrooms can be listed as follows:

1. The good repertoire of FS help ESL learners to be more proficient L2 speakers (Nesselhauf, 2003; Pawley & Syder, 1983)
2. Productive activities involving FS help students to develop in L2 pragmatics (Bardovi-Harlig & Griffin, 2005; Bardovi-Harlig & Vellenga, 2012).
3. Every day language use is formulaic, so ESL instruction should include teaching of FS if the aim of instruction is authentic language use (Lewis, 1993; Nattinger & DeCarrico, 1992).
4. A good repertoire of FS helps L2 learners to be more fluent speakers (Boers et al., 2006; Bygate, 1987; Wood, 2002, 2004, 2007, 2009).

2.2.3. Explicit Instruction of L2 Vocabulary and Formulaic Sequences

The roots of types of language instruction are based on the type of learning they encourage. Of the two most prevalent types of learning discussed in L2 pedagogy research, explicit learning refers to learners' deliberate and focused attempt to solve a problem (Dörnyei, 2009; Roehr-Brackin, 2014). Thus, explicit instruction is based on the idea that learning occurs more effectively when it depends on explicit knowledge and conscious attention. Implicit learning, in contrast, occurs due to unconscious and effortless processing of knowledge through long exposure of input (Roehr-Brackin, 2014).

Roehr-Brackin (2014) asserted that cognitively mature L2 learners benefit well from explicit instruction (p.776). This idea has its roots in earlier research on explicit instruction (for instance see Alderson, Clapham and Steel, 1997; DeKeyser, 2003; Ellis, N. 2011; Ellis, R., 2005a; 2005b, 2006; Hu, 2011; Scheffler and Cincia , 2011). Nassaji and Fotos (2007) discuss that explicit learning may be more effective than implicit learning of target structures especially in formal instruction contexts. Nassaji and Fotos further claim that the extent of explicit instruction "can lead to the development of implicit knowledge underlying spontaneous and naturalistic L2 use"

(p. 10). What they refer to as spontaneous language use can also be expressed as fluent use of L2 knowledge, which is the focus of the present study.

In a seminal paper, Norris and Ortega (2000) carried out a meta-analysis of 49 form-focused instruction studies published between 1980 and 1998. They had two aims for writing this review:

1. to present a summary of empirical findings about types of language instruction (basically explicit/implicit instruction, focused on form and focused on meaning)
2. to discuss the role of specific research methods in the reported findings.

Upon examining over 250 research papers indexed in the Educational Resources Information Center (ERIC) electronic database, Norris and Ortega (2000) decided to include only quasi-experimental or experimental studies that include an instructional treatment, and targeted forms or functions in morphology, syntax, or pragmatics in order to prepare the final data for a meta-analysis. After this, 77 studies remained in the meta-analysis data bank. After reviewing each publication, Norris and Ortega (2000) found that 49 studies were methodologically comparable to each other. The meta-analysis confirmed that generally explicit types of instruction were more effective than implicit types of instruction. The main findings can be summarized as follows:

1. form-focused instruction produces considerable gain in acquisition of the target forms
2. the effect seems to be sustained over time
3. explicit instruction creates positive results compared to implicit instruction

Norris and Ortega (2000) also acknowledged that a myriad of moderator variables (such as learner factors such as age and gender, cognitive factors such as aptitude and degree of noticing, linguistic factors such as the type of the target structure and pedagogical factors such as the intensity of instruction and the extent of integration of the intervention to the syllabus) have various roles in L2 learning (p. 502) and SLA researchers should consider conducting replication studies that operationalize the central constructs in SLA in a consistent, systematic and empirical way.

In a recent study, Roehr-Brackin (2014) examined the long-term development (over 3 years) of adult learners' use of two L2 constructions (the German Perfekt of *gehen* and *fahren*). The results showed that explicit knowledge had a positive effect on the participant's L2 learning and use. After 3 years of explicit instruction, the participants were able to develop beyond the conventional bottom-up grammar learning routine and take up a top-down approach to the target structures. This indicates that explicit instruction in the long-term can result in implicit processing of L2.

In a doctoral study, Nguyen (2014) investigated the compared effectiveness of explicit instruction of FS with Mandarin speaking students. The results showed that Intermediate level learners highly benefited from form-focused instruction of FS. The learners were not only able to use this productive knowledge in new situations, but also were able to retrieve it in the long run. Nguyen explains that the learners retained new information easily because they were able to encode the meaning of the FS thanks to explicit instruction.

In a 2014 study, Boers, Demecheleer, Coxhead and Webb implemented an intervention of verb-noun collocations with ESL students. After receiving focused instruction of collocations that mostly contained matching exercises, the participants were able to use the recently learnt collocations, but not to a great extent. Boers, Dang and Strong (2017) conducted a partial replication of this study by adding more exercise types to the intervention and found that learners showed greater gains with the exercises that required the use of intact expressions rather than assembling and reassembling words in collocations.

Bardovi-Harlig, Mossman and Vellenga (2015) investigated the effect of instruction of pragmatics and the use of FS with 26 L2 learners. After 4 hours of focused instruction with samples from MICASE corpus, the students completed an oral production test that included agreement, disagreement and clarification scenarios. The results showed that the participants in the experimental group, when compared to the comparison group, produced significantly more speech acts and correct target forms. In the oral production test, the participants used new FS that they were not presented in the instruction.

Peters and Pauwels (2015) found that explicit instruction of FS increased EFL business students' their use of FS in writing tasks. The instruction involved FS recognition in texts and choosing the FS and choosing certain FS in academic tasks. The instruction was supported by fill-in-the-blanks and rephrasing exercises. The FS recognition gain as mostly observed in cued output and spontaneous use of FS in academic tasks, indicating that activities involving production rather than recognition should be used in EAP classes while teaching FS.

Nassaji and Fotos (2007) propose that in-class activities should contain a considerable amount of form-focused activities (specifically those requiring output) in order to gain an increase in the linguistic development of the learners in terms of accuracy and fluency (p. 15). Ellis, N. (2007) also agrees with this view and suggests that L2 instruction should overcome the habit of employing L1 learning principles (implicit learning) and leave some room for explicit instruction. Another line of researchers (Ellis, 2002a, 2002b; Ellis, Basturkmen and Loewen, 2001; Fotos and Hinkel, 2007; N. Ellis, 2007; R. Ellis, 2003; Swain, 2000, 2005) suggest that in addition to explicit instruction, opportunities for output in classroom is another important factor in L2 proficiency gains.

Hyland (2012) discusses that FS have a specifically positive role in EAP and has a pedagogical value for EAP teachers. He warns that there are not many instructional studies to demonstrate how FS should be taught in EAP setting. One example is Jones and Haywood's 2004 study. Jones and Haywood (2004) instructed intermediate level L2 students in an EAP setting. The participants were exposed to repetition tasks and FS. The results showed that the participants developed an increased awareness about FS and were able to use them in writing tasks. Hyland (2012) concludes that more empirical studies are needed to be conducted by adopting systematically derived inventories such as Simpson-Vlach and Ellis' *Academic Formulas List* (2010).

Simpson (2004) also argues for the explicit instruction of FS in EAP setting and concludes that FS should be presented in context. Simpson also warns that EAP students should practice several FS serving a similar purpose. In a similar vein, Simpson-Vlach and Ellis (2010) created an Academic Formulas List and classified them according to their pragmatic functions. The present study takes these into

consideration. The instruction of FS that the participating EAP students were exposed included a long list of FS categorized according to their functions and presented in concordances that came from a frequently researched spoken academic corpus MICASE.

2.3. Speaking in L2

Speaking is "the verbal use of language to communicate with others" (Fulcher, 2003, p.23). The manifestation of speech is through sounds waves, which is also one of the properties that distinguish it from the other productive language skill, i.e. writing. Speaking is also different from writing in that it involves less formal use of language, fewer full sentences, contains repetitions, hesitations and pauses and has more conjunctions instead of subordinations, accompanies turn-taking behaviors, involves frequent use of fixed expressions that help conversations, varies in accordance with the social context of speech event (Fulcher, 2003; Thornbury and Slade, 2006). In addition to all these properties, Fulcher (2003) mentions the psychological aspects of speech that makes it different from writing. The writer has time to plan, produce and revise texts, whereas the speaker has to plan, formulate and articulate with considerable speed. The rate of speech that is appropriate to the context will depend on various factors such as speakers' control over the structure of the language (Fulcher, 2003, p. 24), their lexical range (Stengers et.al., 2011) and the ability to use formulaic expressions (Boers & Lindstromberg, 2009; Bygate, 1987). All of these factors lead to automatization of speech that no longer requires conscious attention.

Learning how to speak in a foreign language is even a more difficult task. Even though learners may possess a strong understanding about the grammar of the target language and a good amount of target language vocabulary, it does not guarantee to be able to speak in L2 in an efficient way. Research shows that it is necessary for language learners to learn and use formulaic sequences of the target language in order to be efficient L2 speakers (Lewis, 1993). Indeed, according to Ellis (1997), native speakers "speak idiomatically using frequent and familiar collocations", so the job of the second language learner should be to learn and use these sequences (p. 129). Actually, formulaic language is what one expects to hear in genuine communication (Schmitt & Carter, 2004). This indicates the essential role of

formulaic sequences in authentic use of the target language, both by native and non-native speakers.

In their seminal paper, Pawley and Syder (1983) discussed the role of formulaic language in native speakers' language use. They asserted that what made native speakers so efficient was the knowledge of several hundred thousand lexicalized sentence stems. As these stems are automatically accessed and processed, native speakers can allocate other cognitive processing sources to interactional issues. Indeed, the use of FS does not only allow efficient working memory use in native speakers, but also make them more accurate (Taguchi, 2007) and more fluent (Boers et al, 2006; Wood, 2002). So, returning back to Ellis' assertion (1997) about the use of FS by native speakers, one can say that L2 speaking can only be efficient when formulaic sequences of the target language are learned and used automatically in speech. So, as put by Ellis (1997), efficient L2 speaking is only possible when a good amount of FS is built upon sufficient knowledge of L2 grammar and vocabulary.

How FS help automatic processing and fluency was explained by Wood (2002) as follows: Prefabricated pieces are interwoven to each other, categorized into certain communicative situations. The act of speaking requires a good deal of planning in order to manage long stretches of speech. So, a large store of formulaic sequences and speech acts help automatic access and processing of these linguistic items. In the end, these items are pulled from the memory without much effort and fluency is enhanced. This also helps efficient planning of speech and vocabulary use since redundancy of attention is prevented by the use of readily stored, accessed and processed FS. In other words, FS lead to more efficient L2 speaking performance, because they reduce attention and processing burdens and enhance fluid and efficient communication.

L2 speakers, especially those with limited proficiency, have difficulties in allocating their attention on both form and meaning of linguistic items because they usually lack of a large store of FS. This situation forces them to make a choice between these two dimensions while speaking: they either focus on meaning or they focus on form (Anderson, 1995; VanPatten, 1990). However, when they have a chance to plan their oral production, they can overcome these processing limits, and as a result, the quality of their performance is enhanced (Yuan & Ellis, 2003). This shows that L2 speaking is a difficult skill to develop because there are so many

different linguistic components (i.e., grammar rules, vocabulary knowledge, FS etc.) and cognitive processing limitations involved (i.e., working memory capacity, attention etc.).

2.4. Aspects of Linguistic Performance: Fluency, Accuracy, Complexity

Fluency, accuracy and complexity, as the three basic aspects of linguistic performance, are "part of the vocabulary of language teaching" (Fulcher, 2003, p. 30). Actually, language teachers have an intuitive understanding of these terms and tend to classify classroom activities as "fluency" or "accuracy" based (Brumfit, 1984; Ur, 1981). According to Fulcher (2003), in terms of language testing, accuracy and fluency are related to, but not dependent on each other in that they are seen as being at opposite ends of a continuum where in one end there is accurate and dysfluent speech and in the other end is the inaccurate but fluent speech. However, together they constitute the construct definition of L2 speaking (Fulcher, 2003, p. 27), both from the perspective of L2 performance (Skehan, 1996, 1998; Ellis, 2003; Ellis and Barkhuizen, 2005) and from the perspective of L2 proficiency (Housen, Kuiken and Vedder, 2012) that interact during the course of L2 development.

In his earlier papers dating back to 1996, Skehan (1996, 1998) distinguished between three aspects of speech (fluency, accuracy and complexity) in a series of publications. According to him, fluency refers to the learner's capacity to produce speech with less hesitations and pauses. Complexity refers to elaboration of the current language system. It is associated with subordination and lexical richness of speech (Yuan & Ellis, 2003). Accuracy, on the other hand, concerns the degree the utterance corresponds to the target language norms.

Norris and Ortega (2000) also emphasized that further research showed that complexity, accuracy and fluency are distinct and interrelated constructs. According to Leonard and Shea (2017), these three dimensions of L2 performance interact during the course of L2 development regardless of the learning context (i.e. natural or instructional setting) and are helpful to explain the processes and the subskills underlying automaticity of the speaking act; however, they were also observed to develop at different paces during interlanguage development (Spoelman and Verspoor, 2010; Housen, Kuiken and Vedder, 2012).

2.4.1. Accuracy

Language teachers are aware that their students make errors while speaking. While some of these errors are negligible because they do not interfere with communication, some of them are rather serious in that they interfere with the intended message. According to Gilquin and De Cock (2011), establishing the norms for accuracy in written language is a thorny issue. Götz (2013) adds that it is even more difficult to establish the norms for accuracy in spontaneous spoken language because speaking requires additional cognitive planning and pressure on the speaker.

Fulcher (2003) listed the common types of errors that are penalized or ignored in testing situations according to their gravity (the degree of interference with communication). According to Fulcher (2003), the most serious error that L2 speakers make while speaking in English is word order and omission of words. This type of errors is known to decrease rapidly as the learner improves beyond the stage of beginner (Fulcher, 2003). The second type of errors is misuse of pronouns and relative clauses. These are not as serious as word order errors but they may cause a problem in the coherence of the utterance. Following these two types of errors, tense errors and misuse of prepositions are less serious type of errors L2 speakers make.

This information is useful in formulating the operationalized definition of accuracy of speech. In previous studies, accuracy was associated with correct use of certain structures in speech. For example, Yuan and Ellis (2003) measured accuracy as the percentage of accurately used verb forms and the percentage of clauses that did not contain any error. The first accuracy measure considered the percentage of correctly used verb forms in terms of tense, aspect, modality and subject-verb agreement. In the second accuracy measure, they considered all errors related to syntax, morphology and collocations. Ahmadian and Tavakoli (2011) operationalized accuracy of speech as the number of error-free clauses and correct verb forms in speech.

2.4.2. Complexity

Complexity refers to the elaboration of the current linguistic system (Skehan, 1996). According to Skehan (1996), it involves taking risks in language production and restructuring the interlanguage.

Researchers associate complexity of speech with the existence of subordination in utterances (Clercq and Housen, 2017). Previous studies mostly dealt with the relationship between task planning conditions and complexity of speech. According to Foster and Skehan (1996), pre-task planning results in greater complexity of language production. Mehnert (1998) found that complexity in speech is observed only when planning time is longer than 10 minutes. In the one-minute planning condition, the participants produced accurate speech which was less complex than 10-minute planning condition. Yuan and Ellis (2003) argue that accuracy and complexity of speech is enhanced in online planning conditions. In online planning conditions, the participants have to allocate their attentional resources into the ongoing task and have to prioritize form over meaning. However, this results in a decrease in fluency. According to Ahmadian and Tavakoli (2011), complexity is also enhanced when learners are given a chance to repeat a task.

As can be seen, the three aspects of speech are interrelated, but still need to be distinguished from each other because they require different type of processing on the learner's part (Skehan, 1996). As human beings have a limited capacity to process information, they are not able to attend to all aspects of a task. This occurs when learners have difficulty in allocating attentional resources and have to prioritize one aspect over the others (Anderson, 1995; Skehan, 1996; VanPatten, 1990).

For example, from the perspective of second language teaching, encouraging learners to produce language more accurately allows controlled rather than automatic processes (Ahmadian & Tavakoli, 2011), so for the sake of speaking in an accurate fashion, fluency is forfeited.

On the other hand, in some tasks, speakers may draw on formulaic language which results in enhanced fluency; however, they may not be able to refer to the acceptable language norms so efficiently and they may fail to produce accurate and complex speech (Skehan, 1996; Yuan and Ellis, 2003). In fact, when learners are producing more fluent language they are prioritizing meaning over form (Ellis & Barkhuizen, 2005).

To sum up, all three aspects of speech cannot be fully achieved in linguistic tasks, some degree of loss is always expected (Skehan, 1996, 1998; Yuan & Ellis, 2003).

2.5. L2 Fluency

A clear definition of L2 fluency has been a challenge for researchers. Koponen and Riggenbach (2000) define fluency as smoothness of speech. They also argue that historically speaking, the term *fluency* has evolved to be typically attributed to nonnative rather than native speech since native speakers are perceived to be "eloquent or articulate more often than fluent" (ibid, p. 9).

To this date, different types and categorizations of fluency have been proposed. For example, as one of the earliest cases of this categorization, Fillmore (1979) argued that there were four types of fluency: 1) purely temporal sense of fluency in which the speech flows rapidly; 2) producing statements in a logical and coherent manner; 3) not to be lost for words, being socio-pragmatically fluent; 4) being linguistically witty; especially used in imaginative, creative writing. Fillmore's classification indicates that there should be different levels of fluency that varies according to the social environment, situation and topic. These categories were developed according to L1 fluency types; however, they can be adapted into L2 speaking as well.

Lennon (1990) describes two basic types of fluency: a broad sense of fluency which corresponds to oral proficiency, and a narrow sense of fluency which refers to the speed of oral production. Emphasizing that temporal variables such as the existence and the quality of pauses and hesitations in speech is just the tip of the iceberg as indicators of fluency, Lennon (2000, p. 26) further discusses that a working definition of fluency should be "the rapid, smooth, accurate, lucid and efficient translation of thought or communicative intention into language under the temporal constraints of online processing" (p. 26). Such a definition calls for an understanding of fluency as the outcome of the interaction between linguistic knowledge, performance and processing. Indeed, Lennon (ibid, 28) also talks about the concept of "false fluency", a term that describes linguistics performance where the utterance is

smooth and fast, but poor from the conceptual, communicative and functional perspective.

Tavakoli and Hunter (2018) reformulated the definition of fluency types by Lennon (1990) and came up with the following framework:

1. *Very narrow* definition defines fluency in terms of temporal qualities such as speed, existence and frequency of pauses and self-corrections. Mostly researchers use this approach since it allows a systematic measurement.
2. *Narrow* definition refers to ease and fluidity of speech as independent from accuracy and complexity of speech.
3. *Broad* definition refers to fluency reflects L2 speaking ability and confounds elements such as pronunciation, self-confidence and accuracy of speech.
4. *Very broad* definition associates L2 fluency with overall L2 proficiency.

Tavakoli and Hunter think of these definitions as a pyramid where the very broad definition is at the bottom and the very narrow definition is at the top of the pyramid.

Lennon (2000) also describes the term "fossilized fluency". Fossilized fluency refers to fluent performance where inaccuracies in speech became automatized. In such cases, the speaker would choose to slow down her speech so that she could monitor her speech and make fewer errors. But, this time, listening experience is impeded.

According to Fillmore (1979), sources of fluency differences are mainly 1) vocabulary size and elaborative use of vocabulary; 2) repertory of idioms and other formulaic expressions and the ability to use them in appropriate contexts; 3) mastery in creative/ generative use of language forms and vocabulary in given situations; 4) knowledge of interactional schemata for conversations and appropriate ness of particular words. Fillmore's definition and classification of fluency types calls for an understanding of fluency that varies among different contexts. So, a proficient speaker would demonstrate different levels of fluency at different situations since different contexts would have different requirements for fluent performance. In other words, there are contexts where a hypothetical proficient speaker would display her most

fluent performance through use of formulaic expressions; however, there are also some other contexts where she could do her best by using the language in a creative/generative way.

For some researchers fluency is merely a performance variable, whereas for some researchers it is a cognitive skill (Segalowitz, 2010). According to Lennon (1990), fluency is purely a performance-based phenomenon and there is not a "fluency store" in the speaker's mind (p. 391). This purely performance-based view of fluency requires marking of temporal variables in fluency-related research studies. In language assessment, fluency is considered as a dimension of oral proficiency which reflects smoothness and continuity of speech that indicates lack of excessive hesitations and pauses (Koponen & Riggenbach, 2000). The length of fluent runs is also considered as an indicator of oral proficiency in some studies (see for eg., Towell, Hawkins & Bazergui, 1996). However, for some researchers, considering fluency as an indicator of fluency is not efficient in explaining current linguistic competence of language learners (Wood, 2012).

According to the latter perspective (i.e. fluent oral performance as an indicator of fluent cognitive processing), as an automatic procedural/ cognitive skill, fluency is defined as producing speech at a speed that requires little effort or attention (Schmidt, 1992). So it involves psycholinguistic processes of speech planning and production that function efficiently (Lennon, 1990). In accordance with this perspective of fluency, Schmidt (1992) made a distinction between procedural knowledge and procedural skill and placed speech fluency in the latter category. To summarize, the different perspectives L2 fluency has been examined are fluency as a characteristic of the speaker's speech (utterance fluency) and fluency as a characteristic of the speaker (cognitive fluency) (Segalowitz, 2010).

Segalowitz (2010) suggested that a consistent definition of L2 fluency should go beyond the audible aspects of language and attempt to include a wider context of scientific inquiry, i.e. cognitive sciences. According to Segalowitz (2010), L2 fluency is an aspect of L2 use that emerges from the complex interplay of many factors in a dynamic system. The components of this dynamic system are neurocognitive basis of speech production, a motivational system that supports engagement in L2

communication, the social context in which L2 communication is embedded and the fluency related perceptual and cognitive experiences.

Segalowitz (2010) is more interested in exploring fluency as a part of cognition. However, he emphasizes that the "fluency gap" between L1 and L2 speech is what makes the study of L2 fluency a significant research area. If this gap was not existent, there would be no need for researching L2 fluency because it would be just a part of human cognition that does not differ across languages or according to acquisition order. It is clearly observed in L2 speakers' inability to achieve the same level of fluency in their L2 as in their L1 (p. 168). According to Segalowitz, it is also important to note the large variances between the L2 fluency levels of individuals in comparison to the difference between L1 fluency levels.

Segalowitz (2010) further discussed that there are three distinct senses of fluency: *cognitive fluency*, *utterance fluency* and *perceived fluency*. *Cognitive fluency* concerns the speaker's ability to control the underlying cognitive systems that feed into production of utterances. This involves monitoring the cognitive processes efficiently to ensure that the utterance involves the intended message and speech runs smoothly. So this process includes not only lexical retrieval and information access that will support the message but also tailoring the utterance in accordance with the intended message and articulation processes. All these processes should be conducted almost simultaneously and with great efficiency to guarantee the smoothness of speech. *Utterance fluency*, on the other hand, deals with the features of utterances such as the amount of pauses, hesitations, repetitions and repairs in speech. However, utterance fluency does not concern how the utterance is perceived by a listener, so it is the actual performance of a speaker as measured in terms of temporal variables (such as the number of syllables or words uttered in a minute). *Perceived fluency* is related to the inferences of listeners about speakers' utterances. It refers to the conditions where a listener (most of the time, who is an examiner) is supposed to listened to a person or a recorded speech sample (in an online or an offline task) and judges the speaker as fluent or dysfluent according to impressions he/she draws from the speech sample.

In the present study, what is referred to as *L2 oral fluency* falls under *utterance fluency* of Segalowitz's classification of senses of fluency. So it is a performance

related variable that needs to be analyzed through temporal features. Tavakoli and Hunter (2018) asserted that although examining fluency from the perspective of only utterance fluency, it is the one way that allows the most systematic and reliable way to observe and measure fluency.

For researchers who investigate fluency as a performance variable, it is important to find ways of defining fluency that allows it to be measured precisely. Several quantitative measures that define fluent performance were proposed in previous studies. These measures include speech features like pauses, hesitations and repetitions in speech, rate of speech and a combination of these (Riggenbach, 1991). Detailed information about how these temporal variables are defined and measured is given in the following section.

2.5.1. L2 Oral Fluency

The term fluency has been associated with many different aspects of speech such as accuracy of speech (Ahmadian & Tavakoli, 2011), ease and smoothness of speech or lack of pauses and hesitations in speech (Mizera, 2006), and even good command of a language and overall L2 proficiency (Chambers, 1997; Derwing, 2017; Fulcher, 2003; Stengers et al., 2011; Wood, 2012). According to Lennon (1990) and Gut (2009), there are two possible definitions of oral fluency. Broad definition of fluency reflects the general oral proficiency, whereas narrow definition refers to speed and smoothness of oral delivery. In other words, it is "rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (Lennon, 2000, p. 26). In the first view, fluency is measured as perceived by a listener, whereas the second view requires measurement of specific temporal variables which will be discussed below.

2.5.1.1. Oral fluency as an indicator of proficiency

Segalowitz (2010) discusses that in the bilingualism neuroimaging research, *proficiency* and *fluency* are two terms that are used interchangeably. According to his definition, proficiency is the possession of linguistic knowledge, whereas fluency refers to the ability to use that knowledge. The knowledge aspect of proficiency is reflected in accuracy, i.e. errors in speech. Fluency, on the other hand, is reflected in oral performance, and it is different from proficiency in that even if one possesses the

linguistic knowledge, he may not be able to produce language smoothly at an acceptable speed rate. Segalowitz suggests that even if these two terms are significantly distinguished from each other, from the neurocognitive point of view, it is not totally unacceptable to use them interchangeably. However, Wood (2012) warns that it is important to distinguish them, from the same perspective, since even if all linguistic knowledge is stored in the brain, there is no fluency store or center.

For SLA researchers, fluency is an aspect of L2 proficiency as well (Schmidt, 1992; Segalowitz, 2010; Wood, 2012). However, research reveals that there are not universally applicable measures of oral fluency that measure one's speech in an absolute fashion. For instance, Derwing et al. (2004) examined the relationship between Mandarin speakers' oral fluency in English as L2 and listeners' judgments of the participants' oral fluency. They found that there was a relationship between the number of pauses and pruned speech rate and perceived fluency; however, relationship between perceived fluency and other fluency measures were not strong. This result indicates that judgments of fluency are based on only some aspects of fluency and are mostly selective. This result has important implications for language testers who aim to assess L2 test takers' oral proficiency through listener judgment.

Recently, there has been a growing interest among researchers to investigate the links between oral proficiency and L2 oral fluency. These studies focused on the perceptions of listeners' judgments of oral fluency and proficiency and how it affects oral proficiency assessment. However, perceptive fluency is affected by several factors such as intonation and pronunciation. Especially native speakers perceive non-native speakers' speech as dysfluent due to their poor pronunciation (Götz, 2013).

For instance in a 2013 study, De Jong et al. investigated whether L2 fluency measures can be predicted by L2 cognitive skills. They used a series of timed tasks to measure the participants' L2 processing skills and a vocabulary and a grammar task to gauge their declarative knowledge of L2. They found a correlation between only some of L2 oral fluency measures and participants' scores on L2 linguistic knowledge (grammar) and processing skills tasks. This interpreted that language testers should be concerned with the highly correlated fluency measures while assessing L2 proficiency.

Kang (2012) examined whether prosodic indices of speakers' accentedness influenced raters' judgment of oral proficiency. 70 undergraduate students rated the speaking and teaching proficiency of 11 international teaching assistants. The results showed that prosodic fluency was the most powerful predictor of oral proficiency ratings. This indicates that language test criteria should be sensitized according to prosodic features such as speech rate, syllable stress, and pitch range.

2.5.1.2. Oral fluency as a temporal quality

L2 oral fluency, as a temporal quality, has been associated with several different components of speech. Although research on L2 oral fluency is scarce, researchers investigating oral fluency as a performance variable share similar views on the types of temporal variables to be measured (Wood, 2012, p. 12). Today it is commonly agreed that fluent speech is diagnosed by longer fluent runs, increased speech rate and fewer hesitations and pauses (Kormos, 2006; Lennon, 1990; Segalowitz, 2010; Wood, 2004, 2012). For this reason, operational definitions of L2 oral fluency are focused on these phenomena. However, the major problem with operational definitions of oral fluency is caused by the lack of precise measurements of the mentioned speech variables such as the amount of pauses and hesitations, amount of repetitions etc. According to Segalowitz (2010), it is difficult for oral fluency researchers to decide on the cut point that determines what is referred to by fewer pauses and hesitations, length of uninterrupted runs and so on.

For example, Yuan and Ellis (2003) operationalized L2 oral fluency as the number of syllables per minute and number of meaningful syllables per minute. According to Wood (2012), oral fluency is associated with the rate of speech, repair strategies (repetitions, restarts, insertions, corrections), pauses/hesitations, length of fluent runs, and clause chaining. Wood (2009) examined the speech rate as syllables uttered per minute and mean length of runs as the mean number of syllables uttered between hesitations in the participant's speech in order to calculate the fluency score of the participant. In a 2014 study, Wood operationalized speech fluency as speech rate (as number of syllables divided by duration of speech), phonation time ratio and mean length of run (as number of syllables divided by number of pauses). As can be seen, so many measures have been used to examine oral fluency in previous studies and as also stated by Segalowitz (2010) there is still no consensus on which measures

are the most suitable ones that could cover all aspects of fluency as a performance variable.

In an earlier study, Cucchiarini, Strik and Boves (2002) examined the relationship between ratings of perceived fluency and temporal measures of speech quality (speech rate, articulation rate, number and length of pauses, number of dysfluencies, mean length of runs, and phonation/time ratio). 20 native and 60 non-native speakers of Dutch read aloud texts and their read-aloud performance was scored for fluency by nine experts and was then compared to the scores retrieved by an automatic speech recognizer. The results showed that ratings of perceived fluency correlated highly with the six temporal measures calculated by the speech recognizer. Additionally, the rate of speech was found to be the best predictor of raters' fluency scores. Cucchiarini et al. (2002) also found that read-aloud fluency was in high correlation with articulation rate and the number of pauses in speech and speculated that this finding was not surprising because rate of speech embodies these two temporal variables.

In a recent study, Rohr (2016) examined the relationship between prolongations and utterance and perceived fluency of native and non-native speakers. Data came from a narrative task and nine distinct temporal measures were used for fluency: mean length of run, speech rate, mean length of syllable, mean length of silent pause, and rates of silent pause, filled pause, self-corrections, repetitions, and prolongations. It was found that native speakers were generally more fluent and non-native speakers produced more prolonged syllables, which indicates dysfluent speech. Rohr also found that both native and non-native speakers produced similar number of filled pauses, whereas native speakers spoke faster as expected.

As can be seen, many studies pertaining to utterance fluency used speech samples as data. A limited number of studies also looked into the perceptions of the participants' oral fluency to gain further insight about L2 oral fluency phenomena. One of these studies is Kahng's 2014 study. Kahng examined the differences between speech fluency of L1 English speakers and L2 Korean speakers. Kahng (2014) collected both quantitative and qualitative data on the participants' fluency. The quantitative data came from speech samples while the qualitative data came from stimulated recall interviews. The qualitative findings showed that lower proficiency

learners elaborated on their thoughts during speech production more easily, indicating that they need to allocate attentional sources more intensely since their speech production process is not automatized.

In order to examine the oral fluency as a phenomenon in a more comprehensive way, Skehan (2003) and Tavakoli and Skehan (2005) classified aspects of oral fluency according to the three basic phenomena (speaking rate, pause phenomena, hesitations and repairs in speech) that constitute fluent oral performance: breakdown fluency, speed fluency, and repair fluency. *Breakdown fluency* refers to the flow of speech and is usually measured by counting the number of filled and unfilled pauses and calculating their length. *Speed fluency* refers to the speed of speech and is measured by counting the number of syllables per second. *Repair fluency* is related to repetitions, false starts and corrections in speech. Skehan's classification of aspects of fluency takes three phenomena (pause, hesitation, repair) into consideration which are described in detail below.

2.5.2. Pause phenomenon

Pauses are an important aspect of speech that makes it fluent or dysfluent both from the speaker's and the listener's point of view (Blake, 2006). Most researchers make a distinction between *silent pauses* and *filled pauses*. Silent pauses are, as indicated by the name of the term itself, silent; filled pauses include "ehh" or "uhmm"s.

Most pause studies examined the role pauses in speech, how often and where they are used at different levels and how they are linked to proficiency level and speech planning processes. For example, Kormos and Denes (2004) found that advanced L2 learners produced fewer silent pauses compared to lower level students. Iwashita, Brown, McNamara and O'Hagan (2008) also found that higher level learners spoke faster and produced less silent pauses. In an earlier study, Riggenbach (1991) found a correlation between unfilled pauses and rate of speech.

As mentioned earlier, pauses are also important for listening comprehension as well. In a recent study, Kang et al (2010) found that increase in pauses predicted comprehensibility of speech on the part of the listener. It was specifically the

frequency of *silent pauses* that created a positive impression in terms of fluency among listeners. Kormos and Denes (2004) also found that it was not the number or type of pauses but the length of the silent pauses that predicted fluency on the part of the listener. Kang's (2008) study also showed that the number of silent pauses was a strong predictor of listeners' judgment of oral proficiency.

Although pause phenomenon is an important part of fluency and perceived oral proficiency, as can be seen in the aforementioned studies, there has been considerable debate on the cut-off point for pauses. As Blake (2006) pointed out earlier research considered that this minimum length for the detection of pauses in speech had to avoid the unwanted inclusion of breathing and articulation pauses in the research data. According to Riggenbach (1991) pauses shorter than 0.2 seconds are considered micro-pauses and are not regarded as hesitation phenomena. The majority of researchers suggest 0.25 seconds as the cut-off point (Bosker, Pinget, Quene, Sanders & De Jong, 2013; De Jong et al., 2013; Kormos, 2006; Towell, Hawkins & Bazergui, 1996). However, according to Riggenbach (1991), there should be several cut-off points in accordance with the types of pauses: Micropauses are pauses of 0.20 seconds or less, hesitations are pauses between .3 and .4 seconds and unfilled pauses are those between 0.5 and 3 seconds.

2.5.3. Repetition and repairs

Self-repair refers to speech production phenomenon that includes variables such as repetitions, re-starts, false starts, and cutoffs in speech. This phenomenon is observed both in native and nonnative speech. According to Freed (2000), self-repair can show itself in the four ways:

1. Repetitions of exact words, syllables or phrases
2. Reformulations, false starts: to rephrase the speech part if the speaker failed to transmit the intended message
3. Corrections/grammatical repairs: usually directed onto specific forms
4. Partial repetition: occurs when the speaker is searching for words

According to Segalowitz (2010), under the domain of utterance fluency, self-repair a measurable feature of speech along with other temporal features that shows

the existence of self-monitoring during speech production. In a 2008 study, Seyfeddinipur, Kita and Indefrey found that L1 speakers did not immediately interrupt their speech to correct their errors, but did so when they felt they were ready to make a correction. Segalowitz (ibid) interpreted that self-correction is an indicator of an exchange between maintaining accuracy and fluency. He also concluded that this process should be more demanding in L2 speech in accordance with the proficiency level of the speaker and the circumstances of the context.

In a study by De Jong et al. (2013), it was found that L2 fluency is predicted by L1 fluency behaviors to a great extent. Specifically 57% of hesitation phenomenon can be explained by the impact of L1 fluency. In another study Bosker et al. (2013) found that repair measures did not predict perceived fluency as much as other aspects of fluency (breakdown and speed fluency). Bosker et al. interpreted that as of yet there is no consensus of the role of repairs in speech perception.

Recent research showed that among all three dimensions of speech develop differently and self-repairs are more susceptible to task repetition. Lambert, Kormos and Minn (2016) provided evidence for this premise. They examined the effect of task repetition on three dimensions of fluency and found that Japanese learners of English demonstrated considerable gain in their speech fluency, but in time pauses in speech started to decrease in the second performance and self-repairs decreased after the fourth performance.

As oral fluency involves speech-specific phenomenon such as pauses and hesitations, assessment of oral fluency requires the use of measures that consider the discussions around these phenomena. For instance, Kormos (2006) prepared a list of some commonly used temporal speech components (see Table 2). Although Kormos' list does not include all the previously used measures, it indicates that for L2 fluency researchers, fluidity is the most significant indicator of fluency and it also acknowledges the existence of the basic phenomena related to oral fluency: hesitation, repairs/self-corrections and length and number of pauses in speech.

Table 2
An Overview of Measures of Fluency

Measure Definition[1]	Definition
Speech rate	The total number of syllables produced in a given speech sample divided by the amount of total time required to produce the sample (including pause time), expressed in seconds. This figure is then multiplied by sixty to give a figure expressed in syllables per minute.
Articulation rate	The total number of syllables produced in a given speech sample divided by the amount of time taken to produce them in seconds, which is then multiplied by sixty. Unlike in the calculation of speech rate, pause time is excluded. Articulation rate is expressed as the mean number of syllables produced per minute over the total amount of time spent speaking when producing the speech sample.
Phonation-time ratio	The percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample (Towell, Hawkins & Bazergui, 1996).
Mean length of runs	An average number of syllables produced in utterances between pauses of 0.25 seconds and above.
The number of silent pauses per minute	The total number of pauses over 0.2 sec divided by the total amount of time spent speaking expressed in seconds and is multiplied by 60.
The mean length of pauses	The total length of pauses above 0.2 seconds divided by the total number of pauses above 0.2 seconds.
The number of filled pauses per minute	The total number of filled pauses such as uhm, er, mm divided by the total amount of time expressed in seconds and multiplied by 60.
The number of disfluencies per minute	The total number of disfluencies such as repetitions, restarts and repairs are divided by the total amount of time expressed in seconds and multiplied by 60.
Pace	The number of stressed words per minute (Vanderplank, 1993).
Space	The proportion of stressed words to the total number of words (Vanderplank, 1993).

[1] *(adopted from Kormos, 2006, p. 163)*

Although this list has been found to be comprehensive (Segalowitz, 2010), recent trend in oral fluency research is to make use of measures that do not confound (for eg. Bosker, et al., 2013; Bosker, et al., 2014; De Jong et al., 2013; Lahmann, Steinkrauss and Schmid, 2017). De Jong et al. (2013) propose that earlier studies made use of measures that refer to all three dimensions of speech (speed, pauses and repairs) in their measures and suggest the need for using non-confounding measures of oral fluency in L2 oral fluency research. The present study takes this premise into account (for detailed explanation, see Methodology).

When we look into arguments on the distinctive definition of fluency, we can see that most definitions are majorly concerned with temporal variables associated with fluent speech such as pauses, hesitations, length of uninterrupted runs, and the amount of fillers. As Lennon (1990) asserted, fluency deserves a distinct definition from other aspects of oral proficiency since it is a performance-based variable, rather than mentally stored knowledge. According to Lennon (1990), fluency depends on the judgment of the listener about whether the speaker's production occurs easily and efficiently. If so, how can we trust reliability of measures that test oral fluency? This key issue will be covered in the following sections.

2.6. Assessment of Oral Fluency

Fluent oral performance in a second language is widely accepted to be a significant indicator of oral proficiency in a second language (Rossiter et al., 2010). For this reason, L2 oral fluency is of significant importance for high-stakes examiners (Fulcher, 2003). In terms of assessing oral fluency, researchers investigating L2 oral performance are interested in measuring temporal qualities such as the number of syllables, percentage of hesitations and pauses in speech, whereas high stakes examiners evaluate how much fluent the oral product is according to predefined criteria.

Table 3 below displays the fluency scale for the Common European Framework. As can be seen, this scale is designed for consulting an expert listener/rater's perception of an oral production of an L2 speaker.

As can be seen in the descriptors, oral fluency as described by the Common European Framework is also perceived to be associated with effortless speech, lack of pauses and hesitations, smoothness and ease of speech, speaking speed and rate, and repair strategies to keep the communication going. Such scales are commonly used by high stakes examiners to evaluate L2 oral performance of test takers. However, the problem with this kind of assessment is that it is not reliable since listeners' judgments change from person to person (Wennerstrom, 2000). It is also not precise since oral performance may fluctuate during an oral task. For example, speech rate may change minute by minute which makes it difficult for the rater to give a reliable decision that could be generalized to the overall performance of the speaker. Another difficulty is to define what constitutes pauses and hesitations and how important they are to different listeners in terms of fluency (Mizera, 2006).

Table 3
The Fluency Scale for the Common European Framework

Level		Descriptor
Proficient User	C2	Can express him/herself at length with a natural, effortless, unhesitating flow. Pauses only to reflect on precisely the right words to express his/her thoughts or to find an appropriate example or explanation.
	C1	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.
Independent User	B2+	Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of speech.
	B2	Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he/she searches for patterns and expressions, there are few noticeably long pauses. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party.
	B1+	Can express him/herself with relative ease. Despite some problems with formulation resulting in pauses and "cul-de-sacs", he/she is able to keep going effectively without help.
	B1	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.
Basic User	A2+	Can make him/herself understood in short contributions, even though pauses, false starts and reformulation are very evident.
	A2	Can construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts.
	A1	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.

Assessing oral fluency through temporal measures has been a challenge for researchers (Segalowitz, 2010) since there is no consensus on which variables to involve in analysis of spoken data. For example, Yuan and Ellis (2003) investigated

the effect of online and pre-task planning on L2 oral fluency. They asked 41 undergraduate students to complete a narrative task under no planning, pre-task planning and online planning conditions. Fluency was calculated as the number of syllables and meaningful syllables per minute. In the same vein, Ahmadian and Tavakoli (2011) measured the number of syllables produced per minute of speech and number of meaningful syllables per minute of speech in speech samples.

Towell et al. (1996) suggest that oral fluency measures should involve:

1. the length of fluent runs,
2. the number of syllables between pauses,
3. length of pauses,
4. phonation/time ratio,
5. articulation rate.

In a 1995 study, Freed used the following measures:

1. amount of speech,
2. rate of speech,
3. unfilled pauses,
4. filled pauses,
5. length of fluent runs between pauses
6. repairs
7. clusters of dysfluencies

As of yet, there is no consensus on a certain cut-off point for pauses. However, researchers suggested that different cut-off points could be used for different research purposes. For instance, De Jong et al. (2013) suggest 0.25 seconds or below to measure fluent speech of intermediate L2 speakers, whereas Kang (2012) suggest 0.1 seconds for measuring prosodic features in speech. However, assessment of L2 oral performance is still a thorny issue when it is based on perceived fluency since there is no consensus on the role of repairs in speech perception although pauses and speed seem to predict perceived fluency (Bosker et al., 2013). Indeed, all these phenomena (pause, hesitation, repairs and so on) are naturally observed in both L1 and L2 speech.

For this reason it is important to cover all speech-related phenomena in L2 oral fluency assessment and to avoid measures that overlap each other.

In a seminal paper, Lennon (1990) listed frequently used temporal qualities of fluent L2 speech which were used in L2 speaking related research. Table 4 summarizes these measures below.

Table 4
Temporal Qualities of ESL Speech[2]

1. Words per minute (unpruned)
2. Words per minute (pruned)
3. Repetitions per T-unit
4. Self-corrections per T-unit
5. Filled pauses per T-unit
6. Percentage of repeated and self-corrected words
7. Unfilled pause time as percentage of total delivery time
8. Filled pause time as percentage of total delivery time
9. Mean length of speech "runs" between pauses
10. Percentage of T-units followed by pauses (filled and unfilled)
11. Percentage of pause time at T-unit boundaries (filled and unfilled)
12. Mean pause time at T-unit boundaries (filled and unfilled)

As can be seen, Lennon was not only interested in fluency but also complexity of speech because his list contains the percentage of t-units in speech. Another list of temporal variables to measure oral fluency has recently been proposed by Kormos (2006). As can be seen in Table 5 below, the list contains 10 temporal variables that were consistently used by SLA researchers investigating oral fluency.

[2] As described by Lennon (1990)

Table 5
An Overview of Measures of Fluency[3]

1. Speech rate
2. Articulation rate
3. Phonation-time ratio
4. Mean length of runs
5. The number of silent pauses per minute
6. The mean length of pauses
7. The number of filled pauses per minute
8. The number of dysfluencies per minute
9. Pace
10. Space

Kormos (2006) argue that researchers have not been consistent in the way they operationalized these temporal variables. For example, in some studies speech rate was measured as syllables per second while in some studies it was measured as words per minute.

However, De Jong et al. (2013) warn that there is a danger in involving a series of measures to assess fluency since they may overlap with each other, which causes imprecision of fluency assessment. For instance, a study taking up speech rate (syllables per minute) as an oral fluency measure is not only measuring speed fluency but also breakdown fluency. For this reason, De Jong et al. (2013) suggest caution in using variables that may confound. In their 2013 study, they used the following measures to overcome this problem:

1. Speed fluency: mean syllable duration
2. Pause phenomenon (breakdown fluency): mean length of silent pauses within and mean length of pauses between speech units
3. Hesitations phenomenon (Repair fluency): number of repetitions per second speaking time and number of corrections per speaking time.

De Jong et al. (2013) explain that for all frequency measures of fluency (number of silent pauses, filled pauses, repetitions, and repairs) they excluded the pausing time and divided the frequencies of pauses, repairs and corrections by the total speaking time so that frequency and duration measures would confound at

[3] *(Kormos, 2006, p. 163)*

minimal level. They warn that dividing the frequencies by total time including pause time would cause the measures to overlap with each other.

Bosker et al. (2013) also warns that there is a danger in using several temporal measures of oral fluency since it increases the chance of having many overlapping variables. In order to overcome this problem, they categorized the temporal measures in their study according to the phenomena they are related to and eliminated measures that confound. The final list of temporal measures used in their study was as follows:

1. Speed fluency: mean length of syllables
2. Pause phenomenon (breakdown fluency): Number of silent pauses, Number of filled pauses, mean length of silent pauses
3. Hesitations phenomenon (Repair fluency): number of repetitions, number of corrections

The abovementioned classification of fluency measures according to fluency-related phenomena originates in Skehan (2003) and and Tavakoli and Skehan (2005). They highlight that oral fluency is composed of three aspects: breakdown fluency, speed fluency, and repair fluency. *Breakdown fluency* refers to the flow of speech and is usually measured by counting the number of filled and unfilled pauses and calculating their length (De Jong et al., 2013). *Speed fluency* refers to the speed of speech and is measured by counting the number of syllables per second. *Repair fluency* is related to repetitions, false starts and corrections in speech. According to De Jong et al., (2013), most studies dealing with oral fluency conducted analysis with confounding measures. For instance, a study taking up speech rate (syllables per minute) as an oral fluency measure is not only measuring speed fluency but also breakdown fluency. However, using confounding variables causes imprecision in measurement.

To avoid this problem, in accordance with Skehan's classification of dimensions of fluency, the present study adopted temporal measures that correspond to breakdown fluency, speed fluency and repair fluency and also that do not overlap each other. Specific information about the temporal variables is given in Methodology.

2.7. Instructional Studies on Oral Fluency

Oral fluency is an important component of effective communication. However, in L2 classrooms and in SLA research this phenomenon is neglected. Rossiter, Derwing, Manimtim and Thomson (2010) examined 14 EFL coursebooks and found that the majority of the activities in the books were free communicative activities and there was an apparent lack of activities involving focused instruction FS and their practice. They suggested that EFL teachers can adapt their own fluency activities by way of using presentation-practice-production technique cycle, task-based instruction or other ways of teaching they prefer.

Tavakoli and Hunter (2018) surveyed 84 EFL teachers about their opinions on how to promote fluency in L2 classrooms. The participating teachers associated fluency mostly with confident manner of speaking and being able to communicate in L2 with intelligible pronunciation and with general L2 oral proficiency that contains ability to use accurate, lexically enriched and culturally acceptable language use, overall referring to the broad sense of fluency (Lennon, 1990). The teachers expressed that they were confident in themselves about how to effectively promote fluency in their classrooms and they were familiar with fluency research to a great extent. However, the teachers also expressed that they did not frequently use activities that are more effective in enhancing fluency such as repetition, FS instruction and fluency strategy training; instead they used free communicative activities and the type of activities that aim to promote overall L2 proficiency such as reading and writing exercises and listening to native speakers. 25% of the participants stated that they did not use any activities to specifically enhance L2 oral fluency. . The results showed a reasonable gap between the fluency practices of teachers and SLA research on L2 oral fluency, indicating immediate need for more instructional research on L2 fluency that would enlighten EFL teachers about how to enhance L2 oral fluency in EFL classrooms.

As one of the researchers who focused on oral fluency in instructional contexts, Wood (2004, 2009) is interested in how teaching of formulaic language affects fluent oral production in L2. In a longitudinal study, Wood (2004, also reprinted in 2012) investigated oral production development of 6 ESL high beginner level students in an intensive English program in Canada. He instructed the

participants to retell the story of silent animated films. The oral production of the participants were examined to have phonation/time ratio (PTR), speech rate (SR), articulation rate (AR) and the mean length of run (MLR) along with lexical phrase/run ratio (LP/R). Wood (2004) hypothesized that in 13 week semester, participants' oral production will be faster, include less pauses, exhibit longer runs between hesitations and lexical phrases will appear more frequently in the longer runs. The results showed that four of the participants' oral production was much faster at the end of the term compared to their oral production in the beginning of the term. There was some evidence that such an increase in speed may be resulted by the use of FS. The participants also exhibited less pauses in their speech over time and 2 of the 6 participants exhibited longer runs, as suggested in the second and third hypotheses. However, the participants performed differently from each other in terms of FS use. Not all the participants showed an increase in the number of FS they used in their speech. However, the participants who showed an increase in their speech rate also made use of FS in their speech more frequently than the other participants. The results of the study suggest that use of FS help L2 learners develop their L2 oral fluency to some degree. However, this was a small scale study and there is need for much more evidence to concretize the effects of learning and using FS in oral fluency development.

Wood (2009) also investigated the effect of focused instruction of FS and fluency on narratives in English as L2. This study was a case study. A female Japanese learner of English as L2 was exposed to 6 weeks of focused instruction of FS along with a fluency workshop. The fluency workshop included 4 stages: in the first stage, the participant listened to the recording of a native speaker twice, following a transcript and marking hesitations. The instructor drew her attention to the FS. This was the "input stage". In the second stage, she shadowed the recording with the transcript in the lab. While doing this, she also paid attention to FS and hesitations in the narrative. She was also exposed to dictogloss, jigsaw and chat circle exercises. This was the "automatization stage". In the third "practice and production" stage, the participant prepared and recorded her own narrative. In the final stage, she participated in a free talk session with classmates commenting and receiving comments about each other's performance, hesitations, and speed (Wood, 2009, p.48-50). Spontaneous narrative speech samples in the beginning of the term and six weeks

after the onset of fluency workshop were analyzed for speech rate (SR) and mean length of runs (MLR). According to the results of data analysis, the treatment worked: "a clear fluency gain was seen in the measures of speech rate and mean length of runs from sample one to sample two, and overall use of formulaic sequences was more extensive and complex in the second narrative" (Wood, 2009, p. 53). Although a small scale study, the results have so much to suggest for those interested in focused instruction of fluency and FS.

Boers et al. (2006) conducted an instructional study with 32 college students. All participants were exposed 22 hours of authentic listening and reading materials. In addition, the experimental group received instruction of "standardized word combinations", i.e. formulaic sequences. During the course of the treatment, the experimental group was first exposed to "exploration stage" in which the participants were encouraged to pay attention to FS in the reading and listening materials. In this stage, transcripts of the listening materials were used for gap-filling exercises. The experimental students examined the co-text of FS while the comparison group students did not focus on phraseology. After exploring the FS in the transcripts, the participants in the experimental group were asked to identify "the useful language", i.e. the FS and share their findings with classmates (Boers et al., p. 251). The comparison group students, on the other hand, underlined some other vocabulary than FS that they found useful because they did not explore the FS in the exploration stage. In the end of the term, both groups were interviewed by two blind judges to gauge their oral proficiency. Both of the judges found the experimental group to be more proficient than the comparison group. They also counted the number of FS in the interviews. Statistical analysis showed that the count of FS correlated well with oral proficiency ratings of the judges. The overall results of the study suggested instructional focus on FS can help L2 learners come across as proficient speakers of L2.

In a 2011 study, De Jong and Perfetti investigated the effect of fluency instruction on L2 oral fluency development of ESL students. They used a 4/3/2 task in which the participants completed an oral elicitation task first in 4 minutes, then in 3 minutes and in 2 minutes. 4-3-2 task is built around the idea that speech repetition increases cognitive fluency because it, even if temporarily, increase the availability of

sentence structures and lead to shorter pauses and less repairs and repetitions. De Jong and Perfetti (2011) had two experimental groups in their experiment: 1) the group that repeated the same task for 4, 3, and finally 2 minutes (Repetition group); 2) the group that spoke on three different topics (No-Repetition Group). They found that both Repetition Groups had gains in fluency.

Thai and Boers (2016) used the 4/3/2 task in a recent study. 20 Viatnemese EFL students performed fluently after the fluency instruction, but no significant increase was observed in accuracy and complexity of their speech.

In another recent study, Tavakoli, Campbell and McCormack (2016) investigated the effect of fluency training in an instructional setting with 37 EAP students enrolled in an UK university. In the experiment, the experimental group received awareness raising activities and practiced fluency strategies in and outside the class while the comparison group received EAP speaking and listening activities. In the fluency intervention, the experimental group listened to a native speaker's speech and rated her fluency in terms of speed, pausing, and repair measures and examined the transcripts of her speech and studied the dysfluencies in the transcripts. In class they did retelling exercises as a part of their fluency training. Outside the class, they recorded their speech and examined their dysfluencies by listening to their performance and re-recording their performance. They were also instructed to use fillers (such as well) and formulaic sequences in their speech. After 4 weeks of fluency instruction, the experimental group gained fluency in terms of speech rate and longer runs.

Wood (2016) recently investigated the relationship between willingness to communicate and L2 oral fluency. The data came from a stimulated recall interview with a Japanese learner of English. The results showed that there was a match between the level of willingness to communicate (WTC) and fluency of speech. Wood discussed that lower levels of willingness to communicate led to more breakdowns in speech. Emphasizing the role of culture effects on L2 oral performance, Wood called for more research on willingness to communicate in L2.

As can be seen, instructional research on L2 oral fluency is scarce. However, L2 oral fluency is a goal that every language learner wishes to attain and language

programs around the world need more scientific evidence on what kind of methods and linguistic structures to implement in L2 speaking classrooms. This is the impetus for the present study.

3.Methodology

In this study, the effect of focused instruction of formulaic sequences on L2 oral fluency was examined. In accordance with the aims of the study mentioned in Introduction, in the present study, the following research questions were answered:

1.What is the effect of explicit instruction of formulaic sequences on L2 oral fluency (when L1 oral fluency and L2 working memory capacity is controlled for)?

1. a. Does the effect of the explicit instruction of formulaic sequences change over time (when L1 oral fluency and working memory capacity is controlled for)?

2. Is there a relationship between the number of formulaic sequences produced by the participants and their L2 oral fluency scores?

3. What is the relationship fluency, accuracy and complexity of the participants' L2 oral performance after receiving explicit instruction of formulaic sequences?

3. a. What is the effect of explicit instruction of formulaic sequences on L2 oral accuracy?

3.b. What is the effect of explicit instruction of formulaic sequences on L2 oral complexity?

4. What are the perceptions of the participants about the effectiveness of the treatment?

The study adopted a nonequivalent comparison group research design with pretest and posttest and also two delayed posttests. A true experimental research design was not possible since the participants came from two intact freshmen English speaking classes. Thus, the sampling method of this study was convenient sampling. The conditions in the instructional environment (such as materials, teacher, type and number of activities) were kept constant for both groups. Table 6 below summarizes the research design.

Table 6
Research Design

Experimental group	Pretest	Instruction of formulaic sequences	Posttest	Delayed Posttest
Comparison group	Pretest	Academic vocabulary instruction	Posttest	Delayed Posttest

Two intact freshmen EAP classes were randomly assigned as the experimental group and the comparison group. The experimental group received explicit instruction of formulaic sequences while the comparison group received explicit instruction of academic vocabulary. The instruction for each group was conducted for five weeks within the same timeframe. Both the experimental and the comparison group in the study received a total of 10 hours of instruction, differing only in the targeted structures they were taught. In addition, all instruction-related and teacher-related variables were kept constant for both groups as the researcher taught both groups and exposed them to similar instructional practices during the intervention.

3.1. Participants

The participants were 37 freshmen students in a private university in Istanbul. The experimental group consisted of 18 students and the comparison group consisted of 19 students. The participants were enrolled in a speaking course titled ENG 102 (English 102) which was a four-credit EAP course for freshmen students. The students were from a variety of disciplines: Business Administration, Banking and Finance, Entrepreneurship, Economics, International Business and Trade, Management Information Systems, Hotel Management, Psychology, Law, Computer Science, Mechanical Engineering, Industrial Engineering, Electrical and Electronics Engineering.

The participants were 18-22 years old (Mean= 19.67, range= 18-22). 21 participants were male and 16 were female. Students repeating the course, speakers with speaking or hearing impairment or perceived speaking anxiety, the international students enrolled in the course and students whose L2 is different from English (i.e. students learning English as L3) were excluded from the study. However, all students enrolled in both classes, including the non-participants, were exposed to explicit

instruction of formulaic sequences and academic vocabulary as it was integrated into the regular instruction. Participation in the study was voluntary and the participants signed an informed consent form (see Appendix N).

The participants were of proficiency level above or near the minimum university entrance requirement of TOEFL IBT 80 or IELTS academic module 6,5. Proficiency levels of the students are determined by the university preparatory program the following steps: Students who enter the university first take the Michigan State University English Placement Test. This test is widely used by several universities around the world for the purpose of placing students into similar ability levels in English as a second language. The test consists of 100 multiple-choice items which cover four language areas: listening comprehension, grammar, vocabulary and reading. Students who are not successful in the placement exam are placed into Elementary, Pre-Intermediate, Intermediate and Upper-Intermediate Remedial English classes. Students who pass the placement test are eligible to take the proficiency exam (Test of Readiness for Academic English). This test consists of a writing section and 100 multiple-choice items that assess listening comprehension, grammar, vocabulary and reading comprehension. The proficiency exam determines which students are eligible to study in undergraduate program. Students who fail the proficiency exam are placed into the Advanced module in the Preparatory English Program. To summarize, the participants in this study were Advanced level learners who were found eligible to continue their academic studies in undergraduate programs.

3.2. Instruction

Instruction for each group was designed as two class hours a week and was implemented in five consecutive weeks. All classes were taught by the researcher in the same time frame. Both groups were exposed to the same type of activities during the course of the experiment following the same teaching method.

Since this was an instructional study, one of the instructional goals was to expose all participants to authentic language use as much as possible. This was accomplished by use of concordances while presenting and practicing the target items. The instruction was also learner-centered and production-oriented. The classroom activities were designed as regular EAP activities that included watching videos and

examining pictures pertaining to world problems, doing in-class research, pair and small group discussions and presentations and stating personal reflections about the reading texts, videos and pictures.

Figure 1 and 2 below shows sample concordances used in this study. In both groups, the students were given handouts that contain the list of items to be studied in the session and the concordances that showed how each item was used in real-life academic spoken discourse (see a sample handout in Appendix Q). The rationale behind this was to make students familiar with the context in which each target item is used in academic spoken discourse and make learning of these items more meaningful for them. The concordances were elicited from MICASE website. Repetitions and filled pause markers in the transcripts such as *erm* and *uhh* were eliminated in order not to encourage students to imitate them.

Figure 1.
Sample concordance for experimental group (formulaic sequences)

that was quite different, right? it may change.	**that's my point**	it might change, I say.

Figure 2.
Sample concordance for comparison *group (academic vocabulary)*

it doesn't matter if it's high carbon dioxide, it doesn't matter if it's an	**accumulation**	of waste products, it just simply doesn't matter.

During the implementation of the instruction, the participants in each group was first introduced the target structures (FS in the experimental group and academic vocabulary in the comparison group) in concordances. At this stage, the teacher and the students talked about the context of the utterances in the concordances and then they practiced the target structures (FS in the experimental group and academic vocabulary in the comparison group) in productive academic tasks such as group discussions, group and pair research and oral summary activities and pair and group presentations. At the end of two hours, the students in both groups were given

worksheets that contain the target structures to be done in pairs (see samples in Appendix Q). A sample lesson outline is displayed in Table 7 below. Sample lesson plans for the experimental and comparison group are also given in Appendix P.

The instructional materials and activities were identical for both the experimental and the comparison group. In both groups, students read the same academic texts and watched the same videos on the topic of the week. The content, requirements and instructions of the in-class tasks were also parallel in both groups to ensure validity of the instruction. The EAP course that the intervention was implemented in was originally three hours a week. However, in this study, the instruction was conducted for two hours a week. In the last hour, the students worked on their term projects.

Table 7

Outline of A Sample Lesson Plan

Warm-up Students examine pictures demonstrating global problems and answer questions related to the issues discussed in the weekly assigned reading (presented on a screen via projector). *Presentation* Teacher introduces a list of 7-10 target structures on a slide (FS in the experimental group and academic vocabulary in the comparison group) by projecting them on a screen. She explains that they can use these words and expressions while answering comprehension questions. The students are also given a handout that contain the target structures of the session and the concordances that contain the target structures (sample handouts can be found in Appendix Q). *Practice* The teacher presents a series of slides that display the target structures in concordance as they can be found in academic spoken discourse. She reads them aloud. Students are also given worksheets that include the list of the target structures. The teacher instructs the students to read them aloud. They examine their use and discuss their use in real-life speech in small groups or pairs. They work in pairs to reflect on the issues discussed and answer comprehension questions about the assigned text. They are encouraged to use the target structures in their answers. For example, the teacher instructs them to use 5 of the target items while answering the questions. *Production* Students also conduct small group discussions, research activities and prepare presentations. They are encouraged to use the target structures while giving report on their answers to the whole class. For example, the teacher instructs them to use 5 of the target items while doing their pair presentation.

As mentioned above, each group was exposed to a different set of target structures (FS in the experimental group and academic vocabulary in the comparison group). Further information about the item selection process is given in the next section.

3.2.1. Selection of Target Items

3.2.1.1. Selection of Formulaic Sequences

According to Ellis et al. (2008), it is important to triangulate instructor insight and corpus data in order to determine which linguistic items to include in instruction. In the light of this suggestion, the present study followed the steps below to systematically elicit the list of FS to be used in the instruction of the experimental group:

1. Firstly, seven course books focused on teaching speaking in English were scanned manually to detect the list of FS and other useful phrases suggested for explicit teaching in Academic English (Advanced Level) classrooms. A total of 739 items were elicited. However, some of the items appeared more than once and also some of them were not multi-word chunks; for this reason, they were eliminated. At the end of this step, there were a total of 328 FS in the list.
2. Second, the items in the list of FS were scanned in two spoken corpora (namely MICASE and BASE) to determine whether they are actually used in real-life academic spoken discourse. The items found in both of the mentioned two corpora were included in the FS list.
3. Next, the academic formula teaching worth (FTW) of each FS was calculated. As also described on page 19 above, this measure was first put forward by Ellis et al. (2008) and it shows EAP instructors' opinions about teachability of a list of formulaic expressions. In order to calculate the formula teaching worth of the target items, the researcher asked six EAP teachers to complete an FTW scale. The FTW scale asked the experienced EAP teachers to judge (1) whether these items constitute a formulaic expression, phrase or chunk, and (2) whether they are worth teaching in the current EAP context. There was a mostly high intercoder reliability among the teachers (maximum reliability Cohen' *k*= 98 and minimum intercoder reliability *k*=0.67). The final list of FS consisted of 80 FS. As can be seen, the list of FS included items that exist in two spoken academic corpora and are worth teaching in the given EAP context.
4. The items included in the final list of FS were categorized according to their function (see Appendix A for the list of FS and their functions). In this way,

the target FS were taught with the sense of a purpose. In this stage, expert opinion was taken from two experienced EAP teachers.

3.2.1.2. Selection of Academic Vocabulary

The academic vocabulary items that were instructed in the comparison group were systematically selected as described below:

1. The reading texts used in the course were examined by two experienced language instructors to elicit the academic vocabulary that were worth teaching in the given EAP context. A total of 201 items were marked by the instructors.
2. The vocabulary items the instructors highlighted were then checked in the Academic Word List (Coxhead, 2000). 82 of these vocabulary items were found in the Academic Word List. These 82 academic vocabulary items were used in the comparison group's intervention sessions. The final list is in Appendix B.

3.3. Data Collection

The present study adopted a non-equivalent comparison group experimental research design to examine the effect of explicit instruction of FS on L2 oral fluency. 37 freshmen students were delivered a pretest, a posttest and a delayed posttest during one academic semester. A second delayed posttest was also conducted 6 weeks after the delayed posttest.

Pretest and posttests used similar type of tasks. Table in Appendix R summarizes the data elicitation methods and tasks that were used in previous fluency-related studies. As can be seen, most studies used re-tell tasks (also named oral narration in some studies) and short conversations to collect L2 speech data.

In this study, the participants' oral performance was examined by using tasks in which speakers demonstrate transactional speech behavior (as in case of monologues versus dialogues). In accordance with this aim, the explicit instruction in the experimental group involved FS that can only be used in monologues. The tasks used in this study were selected from this list of available task types in accordance with the participants' familiarity with them. Expert opinion was taken from EAP

teachers who teach the other groups taking the speaking course. De Jong et al. (2012a) operationalized a three-way distinction in task type used in speaking tests as being complex (simple vs. complex), formal (informal vs. formal), and according to discourse mode (descriptive vs. persuasive). As the participants of this study come from Advanced level EAP students, expert opinion by the EAP teachers were not to use simple and informal tasks and also to include tasks that would allow the use of both discourse modes (descriptive and persuasive). Finally, three task types were determined to be used in the study by referring to comments by the EAP teachers of other EAP groups in the same language school. Moreover, the six EAP teachers also contributed to the selection of the content, words, pictures and prompts used in the oral elicitation tasks. Table 8 below summarizes the task types used in the study to collect data.

Table 8

Task types used in the study

Task Type	Qualities
Task 1: Oral Narration Task	complex, formal, descriptive
Task 2: Picture Description Task	complex, formal, descriptive
Task 3: Oral Argumentation Task	complex, formal, persuasive

Detailed information about the data collection tasks is given below.

3.3.1. Oral Elicitation Tasks

Three different types of tasks were conducted to examine oral performance of the participants in the pretest, posttest and two delayed posttests: an oral narration task, a picture description task and a ranking task. It is widely known among SLA researchers that different task types have different cognitive demands on the learner (Bialystok, 2001). Segalowitz (2010) also warns that oral fluency is sensible to different task types; under different conditions, speakers will exhibit different amounts and kinds of utterance fluency. For this reason, the researcher collected data through three different types of oral elicitation tasks.

In order to record the participants' speech, the researcher accompanied each participant to a silent room. Before recording their speech, the researcher ensured the participants that their speech would be recorded and analyzed for research purposes only, they would not be graded for their performance and these interviews were not to be shared by a third party.

Participants were given 3 minutes to complete each task. They were also given 30 seconds to plan their speech before each task. Planning time and other task conditions such as speech time allocated to each task were determined by corresponding to previous studies. During data collection sessions, the participants were given a digital chronometer that shows the remaining minutes and seconds for each task. The data collection sessions for the posttest and two delayed posttests took around 15-20 minutes for each participant.

3.3.1.1. Oral Narration Task

In this task, the participant was instructed to read and orally summarize a 600 word text on environmental problems. The participant was given five minutes to read the text and then 30 seconds to plan his/her speech. Planning time and other task conditions such as speech time allocated to each task were determined by corresponding to previous studies, six experienced EAP teachers' opinions and the pilot study. When the participant said she/he was ready, the researcher took the article back and gave her/him a simple outline of the article (around 20 words) that summarized each paragraph in one or two words (Appendix F, I, L). The outline was given to control the working memory load that could interfere with the participants' speaking performance. The researcher informed each participant that they had three minutes to finish this task. However, the researcher did not force the participants to talk more if they finished their speech earlier than the recommended time limit. She was also cautious not to interrupt the participants' speech during recording sessions. The participants were warned that they should speak uninterruptedly. Before each recording session, they were told to direct their questions to the researcher.

3.3.1.2. Picture Description Task

Picture description tasks in the pretest, posttest and delayed posttests contained pictures of protests from all around the world. The participants were given 30 seconds to examine the picture and 30 seconds to plan their speech (Appendix E, H, K).

3.3.1.3. Oral Argumentation Task

Oral argumentation tasks have been used in several studies to elicit speech samples (Foster & Skehan, 1996; Kormos & Dörnyei, 2004; Skehan & Foster, 1997). In such tasks, the participants are given a topic or list of items to engage in a problem solving activity.

In the oral argumentation task used in this study, the participants were given a list of concepts and instructed to evaluate their significance and rank them according to their perceived importance. The participants were instructed to speak for three minutes on their opinions (Appendix G, J, M).

3.3.2. Other Data Collection Tasks and Methods

The present study used L1 fluency and L2 working memory as adjusting variables because the literature on L2 speaking suggest that these sources of individual differences have to be taken into account (Segalowitz, 2010) when oral elicitation tasks are used. Data on L1 fluency was collected in a picture description task and L2 working memory was measured through Daneman's Speaking Span Test. Information about the tests are given below. This section also gives information about Familiarity test that measured their knowledge of the target items before and after the intervention and the content of semi-structured interviews.

3.3.2.1. L1 Fluency Measure

L1 fluency was measured in order to gain a baseline to control for individual differences in utterance fluency that are not specific to the L2. According to Segalowitz (2010), using L1 baseline data in L2 fluency research is rare, but necessary for operational definitions of L2 oral (utterance) fluency. To optimize the benefit of controlling L1 fluency, speech elicitation task should be similar to L2 oral fluency

task (Segalowitz, 2010, p. 41). For this reason, the L1 fluency test was a picture description task, similar to the one used in the L2 fluency test (Appendix D).

The participants were instructed to describe a picture in Turkish (L1) in three minutes. They were given 30 seconds to plan their speech. *Pruned speech rate* was calculated to elicit L1 fluency scores of the participants. *Pruned speech rate* is the number of syllables minus the number of repetitions, repairs and filled pauses divided by phonation time. This measure was used because it contains information about all the utterance-fluency related phenomena (rate of speech, the number of repetitions, repairs and pauses) and widely used in previous studies as an overall measure for utterance fluency (De Jong et al., 2013; De Jong et al., 2015; Derwing et al. 2004; Derwing et al., 2009; Freed, 1995; Lennon, 1990; Riggenbach, 1991; Rossiter, 2009; Towell, 2002).

3.3.2.2. Speaking Span Test

Daneman's (1991) speaking span test was adapted to measure the participants' working memory capacity in English. The test includes 100 words organized in a set of two to six words, consisting of five words for each sub-set. In this test, each word is displayed on a computer screen at one second intervals. After each set of words is displayed, a blank screen appears. The participant generates (by speaking aloud) a grammatically, syntactically and semantically correct sentence by using the each given word in the displayed set. An example is given below:

Given words in a three-word set:

pencils *observe* *journey*

The participant might produce:

"Those pencils are mine"

"I've observed it"

"The journey is long"

In the speaking span test, there is no restriction on the sentence length and complexity. There is also no time limit. The participants exercise with three practice trials before conducting the original test. To prevent the participants to rely on a fixed syntactic pattern, they are discouraged to repeat the same sentence patterns (such as "I saw the kitchen" and "I saw the farmers"); since such strategies would tax the processing component of working memory (Mizera, 2006). All the words included in the test are seven letter words with two syllables.

Six experienced EAP instructors working in the same institution evaluated the test items to take out the words the participating students may not be familiar with. The items that were marked as "not familiar" by at least two instructors were taken out from the list and replaced by new items. The total score of this test is 100 and a participant's speaking score is the number of acceptable sentences they could produce with the given words.

3.3.2.3. Familiarity test

In the first and the 8th week of the semester, both the experimental and the comparison group took a familiarity test to assess their current knowledge of the target items. The purpose of using this test was to get secondary data about the students' learning of the items. While the experimental group received a familiarity test containing the target FS, the comparison group received a familiarity test containing the target academic vocabulary. Familiarity tests were fill-in-the-blanks type of tests. The test statements were extracted from MICASE and converted into test items. The purpose of using this test was to examine whether the participants were familiar with the use of the target items in academic discourse. The test items were also checked by two native English speaking EAP instructors to ensure that the statements sounded natural when read aloud and made sense in English.

3.3.3. Interview

Another data collection method was semi-structured interviews. After the instruction was finalized, the participants were interviewed by the researcher about their opinions on the instruction they received. The researcher interviewed the

participants in Turkish. The participants answered the Turkish version of the following questions:

1. What do you think about the instruction of chunks/academic words?
2. Do you think you have benefited from this instruction? Why/why not?
3. Do you think you will use the chunks/academic words you have learned in this instruction in your academic studies? Why?

The interviews took place in a silent room one week after the instruction was finalized. The participants' speech was recorded via an audio recorder. Participation in the interview sessions were on voluntary basis; all the participants from the experimental and comparison group volunteered to be interviewed. Reliability analysis for coding procedure will be given at the end of this section.

3.4. Data Collection Procedures

3.4.1. Pretest and Posttest

The posttest included the same type of tasks as the pretest: an oral narration task, a picture description task and an oral argumentative task. However, the text for the oral narration, the picture for the picture description task and the items to be ranked in the oral argumentative task were different, though the theme was kept constant (glob-al problems) (Appendix E,F,G).

3.4.2. Delayed Posttests

To examine the long-term effect of explicit instruction of FS on oral fluency, a delayed posttest (Appendix H, I, J) was conducted after seven weeks (at the 15th week of the semester) and a second delayed posttest (Appendix K, L, M) (six weeks after the end of the semester). Delayed posttests contained the same type of tasks and also the theme of the tasks was kept constant (global problems). However, due to low number of participants, the data coming from the second delayed posttest was not used in the final analysis.

3.4.3. Number of Participants in Each Test

Table 9 below summarizes the number of participants that were present in each data collection session. Since the first delayed posttest was conducted in the end of the academic semester and the second delayed posttest was delivered in the Summer semester, there was a considerable decrease in the number of participants.

Table 9
Number of Participants in Data Collection Procedures

	Experimental Group	Comparison Group
Pretest	18	19
Posttest	18	19
Delayed posttest	10	9
L1 fluency	18	19
Working Memory	18	19
Interview	18	19

3.4.4. Data Collection Timeline

The dates of data collection procedure are listed below and displayed in Table 10:

1. In the first week of the 2013-2014 Spring semester, the pretest was conducted. Each participant was completed the oral elicitation tasks in a silent study room and students' responses were recorded via sound recorder.
2. In the third week of the semester, the instruction began. The interventions were given during regular class time for five consecutive weeks for two hours in a week.
3. In the eighth week of the semester, the posttest was conducted in a silent study room.
4. The following week, the participants from both groups were interviewed about their perceptions on the intervention. Interviews were conducted in a silent room.
5. In the fifteenth week of the semester, a delayed posttest was conducted to investigate whether the explicit instruction of FS have a long-term effect on L2 oral fluency. Six weeks after the delayed posttest, a second delayed posttest

was conducted. However, due to low number of participants, data coming from this test was not used in the study.

Table 10
Data Collection Timeline

Timeline	Data collection procedures	Variables
Week 1 & 2	Pretest	L2 oral fluency, accuracy and complexity
	L1 fluency	L1 fluency
	Speaking span test	Working memory capacity
Week 3	First Week of Instruction	
Week 7	The End of Instruction	
Week 8	Posttest	L2 oral fluency, accuracy and complexity
	Interviews	Opinions about the instruction
Week 15	Delayed posttest	L2 oral fluency, accuracy and complexity
Week 21	Second delayed posttest	Data not used due to low number of participants

3.4.5. Piloting

The present study went through two pilot studies in two consecutive academic semesters. The first pilot study was conducted in 2011-2012 Summer semester. 20 students in two intact freshmen speaking classes participated in this pilot study. While one group was exposed to 25 formulaic sequences elicited from five speaking course books and checked in MICASE corpus, the other group continued their lessons with regular EAP activities such as reading texts, preparing and conducting individual and oral group presentations, doing research and reporting to the class. The working memory test was also piloted with a group of 10 students. The result of this pilot study

showed that introducing FS without a specific framework was not very efficient. The FS items were categorized according to their functions.

In the Fall semester of the 2012-2013 academic year, the second pilot study was conducted. In the second pilot study, the teaching method and materials for both groups and data collection procedures were tested for feasibility and finalized before the conduction of the original study. However, due to time limitation, teaching of only one quarter of the target items was piloted in the second pilot study. The content, the format and conditions of the tasks were adjusted after the pilot study. The words the participants were not familiar were eliminated. Also the room the participants were interviewed in was organized in accordance with the feedback coming from the participants in the pilot study (lighting, position of the chairs, extraneous factors such as noise coming from outside etc.) and prepared for the experiment.

3.4.6. Elicitation of Fluency, Accuracy and Complexity Scores

In this study, the following data analysis procedures were completed:

1) After the delayed posttest data was collected, background noise of the audio files was cleared by using Audacity 2.0.5.
2) All recordings done in the pretest and the posttest were transcribed for complexity and accuracy calculations.
3) Interviews conducted in the participants' L1 were transcribed.
4) Working memory scores and L1 fluency scores were calculated and processed on Microsoft Office Excel 2010 and Statistical Package for Social Sciences (SPSS) 16.0.0.
5) The number of formulaic sequences used in the posttest and two delayed posttests was calculated manually.

3.4.6.1. Oral Fluency Scores

Currently there is no consensus on how to operationalize oral fluency; however, there are several measurements that were used in previous studies. There is an agreement among researchers that it is not possible to diagnose a person's L2

speech as fluent through uni-dimensional measures (Kormos, 2006; Wood, 2012). In addition to speech rate phenomena, other factors such as hesitations, pauses and repetitions affect fluency of speech (Thornbury, 2005). However, as also pointed out by Kormos (2006, p. 162), there is no consistency between the ways fluency related variables have been operationalized in the literature. For example, silent pauses have been defined with different levels of duration by different researchers (ranging mostly from 200 milliseconds to 1 second). In terms of speech rate, some researchers measured it in syllables per second or minute, while other researchers measured words per second or minute.

The present study adopted seven temporal measures. The measures did not confound each other. That means one measure dealt with only one aspect of fluency and did not in any way overlap with the others.

Pruned speech rate (PSR) was the global measure for oral fluency. It encompassed all three aspects of fluency in itself (namely *speed fluency, repair fluency and breakdown fluency;* see Skehan, 2003; Tavakoli and Skehan, 2005). *Pruned speech rate* was consistently used in previous studies as a global measure for L2 oral fluency (Derwing et al. 2004; Derwing et al., 2009; Freed, 1995; Lennon, 1990; Riggenbach, 1991; Rossiter, 2009; Towell, 2002). As PSR covers all aspects of oral fluency, it was used as the oral fluency measure for the research questions that did not require detailed fluency analysis; namely Research Question 2 that investigated whether there is a correlation between the number of FS used in oral fluency; and also all other research questions that controlled for L1 fluency. In other words, while investigating the controlled effect of L1 oral fluency on L2 oral fluency, only PSR measure was used.

To measure *speed fluency, articulation rate* (AR) was used. It is a measure that was consistently used in previous studies (Kang, Rubin and Pickering, 2010; Kang, 2012). It gives direct information about speed fluency but does not overlap with breakdown or repair fluency.

To measure *breakdown fluency*, three measures were used: 1) *Number of silent pauses,* 2) *duration of silent pauses* and 3) *number of filled pauses*. These three measures give information about pause phenomena (*breakdown fluency*).

To measure *repair fluency*, two measures were used: 1) *Number of repairs* and 2) *number of repetitions*. These two measures do not confound other fluency-related phenomena.

As can be seen, all measures used in this study reflected one aspect of fluency (in other words one fluency-related phenomena). The adopted measures also made use of phonation (spoken) time (excluding silences) instead of total time (including silences). The pause criterion in this study was set at 0.25 seconds as also used in previous studies (Towell, Hawkins &Bazergui, 1996; De Jong et al., 2015) and also according to Riggenbach (1991), pauses shorter than 0.25 seconds are considered micro-pauses and cannot be regarded as hesitation phenomena. The Table 11 below summarizes the temporal oral fluency measures adopted in this study.

Table 11

Operational Definitions of Temporal Measures of Fluency

Aspects of fluency	Measure	Operationalization
Overall fluency	1.Pruned speech rate (PSR)	The total number of syllables minus the number of repairs and repetitions and filled pauses divided by phonation time (Derwing et al. 2004; Rossiter, 2009).
Speed fluency	1.Articulation rate (AR)	The total number of syllables per phonation time (Kang, Rubin and Pickering, 2010; Kang, 2012).
Breakdown fluency	1.Number of silent pauses (NSP)	The number of silent pauses divided by phonation time (Bosker et al., 2013). The pause duration was determined as .25 seconds.
	2.Duration of silent pauses (DSP)	The total length of silent pauses divided by phonation time (Kang, 2012). The pause duration was determined as .25 seconds.
	3.Number of filled pauses (NFP)	The total number of filled pauses divided by phonation time (Kang, 2012).
Repair fluency	1.Number of repetitions (RPT)	The number of repetitions divided by phonation time (Bosker et al., 2013).
	2.Number of repairs (RPR)	The number of repairs divided by phonation time (Bosker et al., 2013).

PRAAT software (version 5.3.49) was used to detect pauses and syllables and calculate fluency scores in running speech. All fluency scores were calculated by using a script that was created by De Jong and Wempe (2008, 2009) and was used in several previous studies (De Jong et al., 2013; Daller et al., 2011; Lahmann, Steinkrauss and Schmid, 2017). The authors' permission was taken to use the script for data analysis. The script was used consistently with participants from different L1 background in previous studies (Bosker, et al., 2013; De Jong et al., 2013; Daller et al., 2011). The 25 ms cut-off point was determined following previous research by Riggenbach (1991) and De Jong et al. (2013).

The sound files recorded in data collection sessions were processed through this script which allows an automatic detection of syllables, number of silent pauses, total duration of speaking time and total duration of pausing time. With the help of this program, there is no need for transcription on PRAAT for fluency analysis. The data was transcribed for qualitative analysis, to detect the number of FS produced by the participants and to calculate accuracy and complexity scores.

3.4.6.2. L2 Oral Accuracy Scores

SLA researchers have different views about how accuracy can be best measured (Ellis and Barkhuizen, 2005). Some researchers have preferred to measure how accurately some specific structures and items were used in speech (for eg., Wigglesworth, 1997), while some others have preferred to calculate the percentage of error-free clauses (for eg., Skehan & Foster, 1997; Yuan & Ellis, 2003). Yuan and Ellis (2003) operationalized accuracy of speech as the percentage of error-free clauses and percentage of correct verb forms. The same method will be used in this study because this measurement was persistently used by previous researchers (Kormos & Dörnyei; 2004; Mackey & Gass, 2005).

The present study used two L2 oral accuracy measures: 1) *percentage of error-free clauses* and 2) *percentage of correct verb forms*. In this study, errors are defined as syntactic, morphological and collocation errors; correct verbs forms include accurately produced verbs in terms of tense, aspect, modality and subject-verb agreement. However, these principles were applied only to calculate the two accuracy

scores, not to categorize the errors in clauses and verb forms. Table 12 below summarizes the operational definitions of L2 oral accuracy scores used in this study.

Table 12
Operational Definitions of Temporal Measures of Accuracy

Measure Definition	Operationalization
Percentage of error-free clauses (ACC1)	The percentage of clauses that did not contain any error. Syntactic, morphological and collocation errors were taken into consideration (Yuan & Ellis, 2003).
Percentage of correct verb forms (ACC2)	The percentage of accurately used verbs in terms of tense, aspect, modality and subject-verb agreement (Yuan & Ellis, 2003).

Two experienced EAP teachers (the first rater was the researcher, the second rater was an experienced EAP instructor with an M.A. degree in ELT) independently calculated the ACC1 and ACC2 on 25 per cent of data. The external coder was trained by the researcher about how to do the calculations for accuracy and complexity scores. The intercoder reliability coefficient was calculating using Cohen's *kappa* (*k*). According to Landis and Koch (1977) *kappa* value 0.41 – 0.60 shows a moderate agreement between raters, while kappa value 0.61 – 0.80 shows substantial agreement and *kappa* value higher than 0.81 shows perfect agreement between raters. It was found that *kappa* value was 0.93 for ACC1 and 0.91 for ACC2, indicating a perfect agreement between the researcher and the external raters. The researcher calculated the accuracy scores from the rest of the data.

3.4.6.3. L2 Oral Complexity Scores

In this study, L2 oral complexity was operationalized in terms of three measurements: 1) *syntactic complexity*, 2) *syntactic variety* and 3) *mean segmental*

type-token ratio. These measurements were also used by Yuan and Ellis (2003) and found to be efficient by Mackey and Gass (2005, p. 241).

Syntactic complexity is the ratio of T-clauses to T-units. T-unit analysis was used because it is widely used in SLA research (Robinson 1995; Yuan and Ellis 2003; Kawauchi 2005; Iwashita et al., 2006). *Syntactic variety* is the number of different grammatical verb forms. *Mean Segmental Type-Token Ratio* is the number of different words divided by the total number of words in every 40-word segment.

Table 13
Operational Definitions of Temporal Measures of Complexity

Measure Definition	Operationalization
Syntactic complexity (COMP1)	The ratio of T-clauses to T-units (Mackey and Gass, 2005; Yuan and Ellis, 2003)
Syntactic variety (COMP2)	The number of different grammatical verb forms (Mackey and Gass, 2005; Yuan and Ellis, 2003)
Mean Segmental Type-Token Ratio (COMP3)	The number of different words divided by the total number of words in every 40-word segment (Mackey and Gass, 2005; Yuan and Ellis, 2003)

After the transcriptions were completed, two experienced EAP teachers (the first rater was the researcher, the second rater was an experienced EAP instructor with an M.A. degree in ELT) independently calculated the COMP1, COMP2 and COMP3 on 25 percent of data. *Kappa* value was found to be k= 0.91 for COMP1, k= 0.95 for COMP2 and k= 0.93 for COMP3, showing a perfect level of agreement between the researcher and the external raters. The researcher calculated the complexity scores from the rest of the data.

3.5. Statistical Procedures

Details about statistical procedures conducted in this study are displayed in accordance with the research questions in Table 14 below. As can be seen, this study made use of both quantitative and qualitative data analysis.

Table 14
Research Questions and the Corresponding Data Analysis Procedures

Research questions	Data collection and measures	Controlling variables	Purpose of the analysis	Data analysis
1.What is the effect of explicit instruction of formulaic sequences on L2 oral fluency (when L1 oral fluency and L2 working memory capacity is controlled for)? 1.a.Does the effect of the explicit instruction of formulaic sequences change over time (when L1 oral fluency and working memory capacity is controlled for)?	1. Pretest and posttest Fluency scores of experimental and comparison group 1.a. pretest, posttest, delayed posttest	1. Working memory capacity for L2 (as measured by speaking span test) 2. L1 fluency (PSR – "pruned speech rate score" coming from a picture description task in Turkish)	To compare the fluency gain from pretest to posttest for experimental and comparison group by adjusting for WM and L1 fluency	Two-way MANOVA (and MANCOVA) with repeated measures
2. Is there a relationship between the number of formulaic sequences produced by the participants and their L2 oral fluency scores?	Posttest Delayed posttest Only experimental group (N=18)		To examine whether fluency of the participants increases in correlation with the number of FS they used in their speech	Correlation (Spearman's rank correlation coefficient)
3.What is the relationship fluency, accuracy and complexity of the participants' L2 oral performance after receiving explicit instruction	3. Posttest		3.To examine whether the 3 components (fluency, accuracy and complexity) of	3.Correlation (Spearman's rank correlation

of formulaic sequences? 3.a. What is the effect of explicit instruction of formulaic sequences on L2 oral accuracy? 3.b. What is the effect of explicit instruction of formulaic sequences on L2 oral complexity?	3.a and 3.b. Pretest and Posttest		oral performance of the experimental group correlate with each other 3a and 3b. To compare the accuracy and complexity gain from pretest to posttest for experimental and comparison group	coefficient) 3.a. Two-way MANOVA with repeated measures 3.b. Two-way MANOVA with repeated measures
4. What are the perceptions of the participants about the effectiveness of the treatment?			To examine the participants' opinions about the instruction. To compare the fundamental findings coming from the experimental group to the comparison group	Open coding

3.5.1. Interview Coding Procedure

The semi-structured interview questions were designed in order to stimulate retrospective thinking on the explicit instruction of FS and academic vocabulary. The interviews were conducted in Turkish asking the participants' opinions about the vocabulary instruction, the extent of its possible benefits and whether or not they plan to use the items they learned in the future. As mentioned above, participation in the interview sessions were on voluntary basis. Interviews were recorded in a silent room via an audio recorder and scheduled one week after the instruction was finalized for both groups.

All of the participants from both groups volunteered to participate in the interviews to answer questions about FS (experimental group N= 18) and academic vocabulary instruction (comparison group N = 19). The interviews took place at week 8, after the instructions were completed.

The recordings were transcribed and coded systematically in order to elicit the most frequently occurring themes. Before analyzing the main body of data, intercoder reliability was calculated. According to Miles and Huberman (1994) and Denzin and Lilcoln (2000), it is imperative to ensure the credibility of open coding procedure by including an external rater. Therefore, one external coder participated in open data coding procedure. The external coder was familiar with qualitative coding procedures and was a PhD candidate in ELT at the time of data analysis. She coded 25 % of the data. The codes found by the volunteer coder was compared to the researcher's codes. Intercoder reliability coefficient was calculated by using Cohan' kappa (k). It found to be k= 0.87 on SPSS package program, an agreeable level of agreement according to Landis and Koch (1977, p. 165). Later, disagreements over the codes in the sample were solved through discussion and the researcher coded the entire data later on to detect reoccurring themes in the interview data. The themes were also confirmed by the external coder and another researcher with PhD in ELT.

After the coding procedure was finalized, an initial list of codes was prepared with 345 codes coming from the experimental group interviews and 230 codes from the comparison group. Later the overlapping codes were eliminated to help the pattern recognition procedure. Through negotiation with a researcher (who holds a PhD in

ELT) and the volunteer coder mentioned above, the researcher looked for patterns emerging in the codes listed in order to determine reoccurring themes. The final list of main themes and categories are given at the end of the Results section, along with sample responses of the interviewees.

4. Results

This chapter presents the results of quantitative and qualitative analysis for the research questions described in Methodology. Adopting an experimental research design, the study not only examined the immediate and the long-term effect of explicit instruction of formulaic sequences, but also aimed to provide in-depth information from the perspective of the participants by way of using interviews to enlighten the gaps that cannot be clearly explored under experimental conditions. Additionally, the third research question that asked whether there was a relationship between accuracy, complexity and fluency scores of the participants provided evidence that is not found in the current literature for advanced speakers, which made the research adopt only an experimental approach but also an exploratory one.

In a non-equivalent comparison group experimental research design, complete exploration of the experimental group and the comparison group participants' language and skills-related level and potential before the instruction is not possible. Therefore, conducting preliminary analysis of the participants' performance before the intervention is of great importance. For this reason, this chapter will first present the result of preliminary analysis comparing the means of pretest scores of both groups in order to determine the comparability of both groups on pretest time. Following that, findings of the study in relation to research questions will be presented.

Examination of L2 Oral Fluency Scores

Seven oral fluency scores were used in the study. *Pruned speech rate* (PSR) measures overall fluency, whereas the other six scores measure three distinct dimensions of L2 oral fluency and they do not confound each other: 1) *articulation rate* (AR) measures speed fluency; 2) *number of silent pauses* (NSP), *duration of silent pauses* (DSP) and *number of filled pauses* (NFP) measure breakdown fluency and 3) *number of repetitions* (RPT) and *number of repairs* (RPR) measure repair fluency.

Table 15 presents the result of factor analysis with scores on pretest coming from the experimental and comparison group (N=37). A principal component analysis with Varimax (orthogonal) rotation was used.

Table 15
Factor Loadings of Scores on Pretest (Rotated Component Matrix)

	Loadings		
	1	2	3
PSR	,975		
AR	,977		
NSP		,771	
DSP			-,416
NFP	-,416		
RPT			,805
RPR		,834	

PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

As can be seen in Table 15, Bartlett's test of sphericity (χ^2 (21) = 148.585, p < .000) indicated that correlations between items were sufficiently large for a pricipal component analysis. However, an examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was not factorable (KMO=.344), since the the KMO value was below the recommended value of .6. the communalities were all above .3, except from DSP (duration of silent pauses), confirming that each item shared some common variance with other items.

Three factors emerged from the component analysis. The initial eigen values showed that the first factor explained 30% of the variance, the second factor 21% of the variance, and a third factor 15% of the variance. The three factor solution explained 67.5 % of the overall variance. Factor 1 included PSR, AR and NFP. Factor 1 also included the highest factor loadings in the varimax solution (PSR= .9, AR= .9). Factor 2 included NSP and RPR, each of which measured a different dimension of fluency: breakdown and repair fluency. Factor 3 included DSP and RPT. The overall model suggests that the factors emerging from the pretest scores did not corresponded to the theories about types of fluency (Speed, breakdown and repair/hesitation).

Table 16 below shows the result of the principal component analysis with Varimax (orthogonal) rotation conducted on posttest scores of comparison and experimental group (N=37).

Table 16
Factor Loadings of Scores on Posttest (Rotated Component Matrix)

	Loadings		
	1	2	3
PSR	,886		
AR	,921		
NSP		,797	
DSP	,678		
NFP		,803	
RPT			,769
RPR			-,676

PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

As can be seen in Table 16, Bartlett's test of sphericity (χ^2 (21) = 76.737, p < .000) indicated that correlations between items were sufficiently large for a pricipal component analysis. However, an examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was not factorable (KMO=.573), since the the KMO value was below the recommended value of .6. the communalities were all above .3, confirming that each item shared some common variance with other items.

Three factors emerged from the component analysis. The initial eigen values showed that the first factor explained 30% of the variance, the second factor 20% of the variance, and a third factor 16% of the variance. The three factor solution explained 67% of the overall variance. Factor 1 included PSR, AR and DSP. Factor 1 also included the highest factor loadings in the varimax solution (PSR= .8, AR= .9). Factor 2 included NSP and NFP, both of which measured breakdown fluency. Factor 3 included RPT and RPR which measured the repair fluency. The overall model suggests that the pretest scores corresponded to the theories about types of fluency (Speed, breakdown and repair/hesitation).

By comparing of the two factor analysis tables, it can be seen that PSR and AR load onto the same factor and present the highest factor loadings in the rotated component matrix, indicating that overall fluency score (PSR) is highly associated with speed fluency (AR). However, scores related to breakdown (NSP, DSP, NFP) and repair fluency (RPT, RPR) showed an irregular pattern as these scores were loaded onto different factors on pretest and posttest. This could have resulted from the intervention the groups received.

Additionally, a regression analysis was used to test if the 6 L2 oral fluency scores (AR, NSP, DSP, NFP, RPT, RPR) significantly predicted the global L2 oral fluency score (PSR). Table 17 below shows the result of the multiple regression analysis that used the pretest scores coming from the experimental and comparison group (N=37).

Table 17
Multiple Regression Analysis for Variables Predicting Pruned Speech Rate

Model	R square changed	B	SE B	β	t	Sig.
AR	.973	,911	,022	,996	41,393	,000
NSP	-	-,004	,132	-,001	-,032	,975
DSP	.984	-,191	,068	-,066	-2,829	,008
NFP	.981	,271	,125	,053	2,161	,039
RPT	.978	-,905	,327	-,068	-2,765	,010
RPR	-	-,160	,344	-,012	-,464	,646

Notes: $R^2 = .98$ (* $p < .000$)
PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

The results of the regression showed that 98 % of the variability in the data is explained by the model ($R^2 = .98$, $F(6, 30) = 306{,}445$, $p < .000$), indicating a very good fit of the model. Although NSP and RPR were not predictors of L2 oral fluency, NFP was a good predictor at the significance level of .05 and DSP and RPT were good predictors at the significance level of .01. The strongest predictor of L2 oral fluency was AR ($p < .000$), which measured speed fluency. This result is consistent with the

factorial analyses conducted on pretest scores in which AR loaded onto the same factor as PSR.

Examination of Pretest Scores

In order to examine the differences between the pretest scores of the experimental and the comparison group, L2 oral fluency scores were submitted to a one-way Kruskall Wallis H test. Table 18 below summarizes the mean differences between the experimental group and the comparison group in terms of seven oral fluency scores in pretest.

Table 18
Mean Ranks of L2 Oral Fluency Scores on Pretest by Group

	Mean ranks			
	Experimental	Comparison	X^2	p value
PSR	17.33	20.58	.833	.362
AR	17.78	20.16	.447	.504
NSP	20.86	17.24	1.043	.307
DSP	19.56	18.47	.093	.761
NFP	16.28	21.58	2.226	.136
RPT	19.11	18.89	.004	.951
RPR	19.03	18.97	.000	.988

PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

The Kruskal-Wallis H test showed that there was not any statistically significant differences in pretest scores between the experimental and comparison group. This result shows that the two groups were at similar L2 oral fluency levels before the explicit instruction of FS. Therefore, an increase in fluency scores in the posttest will show that the fluency gain was caused by the intervention type, in this case instruction of FS or instruction of academic vocabulary.

Examination of Adjusting Variables

One of the adjusting variables in this study was L1 fluency. Segalowitz (2010) strongly emphasizes that oral fluency is not language-dependent; therefore, it is

necessary to control for individual differences in utterance fluency that are not specific to the L2.

Table 19
Mean Ranks of L1 Fluency Scores by Group

Groups	N	Mean ranks	Sum of ranks	U	Z	p
Comparison	19	21.34	405.50	126.500	-1.353	.178
Experimental	18	16.53	297.50			

The Mann Whitney U test showed that there was no significant difference between the experimental group and the comparison group in terms of L1 fluency (U= 126.500, p = .178). This result indicates that the comparison and experimental group participants were not significantly different in terms of oral fluency in their first language. As a result, further statistical analysis for L2 oral fluency can be carried out with confidence because the two groups have equivalent level of utterance fluency. In other words, any differences to be found in L2 oral fluency scores between the two groups will clearly demonstrate that any differences between the two groups in terms of an increase in gain of L2 oral fluency scores are caused by the intervention. Nevertheless, individual differences in oral production can have detrimental effect (Segalowitz, 2010) on the analysis of the data, so L1 fluency will be controlled for in further analyses to ensure reliability of the results.

Working memory is another important factor that needs to be controlled for while analyzing speech data, specifically in L2 in which more cognitive effort is needed during speech processing. Table 20 below shows the result of Mann Whitney U test that compared the mean ranks of two groups in terms of L2 speaking span.

Table 20
Mean Ranks of L2 Speaking Span Scores by Group

Groups	N	Mean ranks	Sum of ranks	U	z	p
Comparison	19	19.95	379.00	153.000	-.548	.599
Experimental	18	18.00	324.00			

The Mann Whitney U test showed that there was no significant difference between the experimental group and the comparison group in terms of speaking span capacity in L2 (U=153.000, p = .599). As individual differences in cognitive processing capacity can have a significant effect on the analysis of speech data, this variable will be controlled for in further analyses as well.

Summary of results of preliminary analyses

- L2 oral fluency scores showed a similar pattern in pretest and posttest. Speed fluency measure loaded into the same factor as the overall oral fluency on both pretest and posttest. However, scores pertaining to pauses and repairs/repetitions interacted with different scores on pretest and posttest.
- L2 oral fluency scores used in this study can predict the overall L2 oral fluency score.
- The experimental group and the comparison group were at a similar level of L2 fluency, across all temporal measures, before the intervention started.
- There was no significant difference between the experimental group and the comparison group in terms of speaking span capacity before the intervention started.
- The experimental group and the comparison group participants were equally fluent in their L1 (Turkish) before the intervention started.

Research Question 1: The Effect of Explicit Instruction of FS on L2 Oral Fluency

The first Research Question examined the effect of explicit instruction of formulaic sequences on L2 oral fluency. A repeated measures MANOVA test was conducted in order to determine the main effects of group (experimental / comparison) and time (pretest – posttest) and also the effect of interaction between group and time. MANOVA with repeated measures was chosen instead of several separate pair-wise t-tests (or ANOVA) in order to avoid committing type I error by multiplying the chance of error. Field (2009) discusses that conducting multiple ANOVAs instead of one MANOVA would inflate the familywise error rate (p. 376-377). As the number of dependent variables increase in a study, the researcher would need to conduct more ANOVAs, which would increase the chance of making a type I error. Field also suggests that MANOVA takes the relationship between the dependent variables into consideration and is able to detect the group differences along a combination of dependent variables.

The results of the multivariate statistical test Pillai's Trace were reviewed to determine if there were significant differences between the two groups on a linear combination of the dependent variables. The Pillai's Trace was used because it is considered the most powerful multivariate test to be used with a small sample (Field, 2009).

The assumption of equality of covariance matrices was first tested through Box's M Test. The result of this test was significant. To overcome such violations of equality of covariance matrices, Field (2009) suggests using the Pillai's Trace for test power. The assumption was then tested through Levene's test of normality. The results showed normal distribution of scores, eliminating the chance of making a Type I error.

Table 21 below shows the result of the repeated measures MANOVA and Univariate ANOVAs.

Table 21
Repeated Measures MANOVA Results on Oral Fluency Measures (Interaction Between Time and Group)

Multivariate tests [a]		Univariate ANOVAs [b]						
Source	F	F values for L2 Oral Fluency Measures						
	(overall)	(PSR)[c]	(AR)	(NSP)	(DSP)	(NFP)	(RPT)	(RPR)
Time (within-subjects)	14,076***	95,306***	89,952 ***	.486	4,600	1,730	6,826	4,047
Group (between-subjects)	1,996	2,349	7,730*	.308	1,078	5,147*	.178	.301
Time x Group (Interaction)	4,903***	7,481**	7,253*	2,002	.052	3,001	.096	.119

[a] df = 7
[b] df = 1
*p<.05 **p<.01 ***p<.001 (N=37)
[c] PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

The results of repeated measures MANOVA are shown in Table 21. The results showed that time [*F* (7, 29) = 14.076, *p* < .001, partial $\acute{\eta}^2$ = .773] had a statistically significant effect on the mean scores of the vector representing the seven oral fluency scores although the effect sizes were small across distinct scores. Follow-up univariate ANOVA statistics showed that time had a significant effect on *pruned speech rate* [*F* (1, 33) = 95,306 *p*<.*001*, partial $\acute{\eta}^2$= 740] and articulation rate scores [*F* (1, 33) = 89,952, *p*< .000, partial $\acute{\eta}^2$ = .720]. Table 22 below shows the descriptive statistics of the scores. As can be seen, the participants scored higher in *pruned speech rate* and *articulation rate* from pretest to posttest.

As for the effect of group, no significant difference was found between the groups. However, there was an interaction effect between time and group on L2 oral fluency scores [*F* (7, 29) = 4.903, *p* < .001, partial $\acute{\eta}^2$ = .545, Pillai's trace= .545]. According to follow-up univariate ANOVA statistics, the interaction between time and group had a significant effect on *pruned speech rate* [*F* (1, 33) = 7,481 *p*<.*001*, partial $\acute{\eta}^2$= .258] and *articulation rate scores* [*F* (1, 33) = 7,253, *p*< .05, partial $\acute{\eta}^2$ = .133].

This finding shows that the experimental group scored significantly higher compared to the comparison group on overall L2 oral fluency after receiving explicit instruction of FS. As discussed earlier, *pruned speech rate* score is widely used as a global score for L2 oral fluency and there was a significant increase in *pruned speech rate* scores of the experimental group when time and group interaction is taken into consideration. Another score that showed a significant increase was *articulation rate* which refers to the *speed fluency*. No significant change was observed for other *pause* and *hesitation* related scores in terms of the effect of the interaction between group and time. In terms of the effect of time, an overall effect was not found. However, the follow-up univariate ANOVA statistics showed that *articulation rate* [*F* (1, 33) = 7,730, *p*< .05, partial $\acute{\eta}^2$ = .603] and *number of filled pauses* [*F* (1, 33) = 5,147, *p*< .05, partial $\acute{\eta}^2$ = .540] differed significantly according to the groups. As can be seen in Table 22, *articulation rate* increased for both groups from pretest to posttest and *number of filled pauses* decreased in the experimental group whereas the comparison group produced more *filled pauses* per phonation time in posttest.

Table 22
Means and SD of 7 L2 Oral Fluency Scores

L2 Oral Fluency Scores	Experimental Group (N=18)		Comparison Group (N=19)	
	Pretest	Posttest	Pretest	Posttest
Pruned Speech Rate	2.08 (.39)	3.36 (.45)	2.21 (.44)	2.83 (.39)
Articulation Rate	2.17 (.43)	3.49 (.49)	2.31 (.49)	3.12 (.44)
Number of Silent Pauses	.29 (.09)	.19 (.08)	.26 (.07)	.26 (.11)
Duration of Silent Pauses	.32 (.13)	.35 (.16)	.31 (.16)	.34 (.14)
Number of Filled Pauses	.12 (.08)	.11 (.07)	.16 (.07)	.22 (.16)
Number of Repetitions	.10 (.03)	.09 (.02)	.10 (.02)	.09 (.02)
Number of Repairs	.10 (.02)	.10 (.03)	.10 (.03)	.10 (.02)

The Role of Adjusting Variables on L2 Oral Fluency Gain

The research question 1 also examined the role of specific covariates in the effect of explicit instruction of formulaic sequences: *L1 fluency* and *L2 speaking span capacity*. In order to control for the effect of the two adjusting variables, a MANCOVA with repeated measures was conducted with pretest and posttest scores of the experimental and the comparison group.

Table 23 below shows the results of the repeated measures MANCOVA test and univariate ANCOVAs with L1 fluency scores and L2 speaking span test scores used as covariates.

Table 23

Repeated Measures MANCOVA Results on Oral Fluency Measures (Interaction Between Time and Group) With L1 Fluency and L2 Working Memory as Covariates

Multivariate tests [a]		Univariate ANCOVAs [b]						
Source	F	F values for L2 Oral Fluency Measures						
	(overall)	(PSR)[c]	(AR)	(NSP)	(DSP)	(NFP)	(RPT)	(RPR)
Time (within-subjects)	.802	1,729	1,424	.297	.817	.636	.521	1,245
Group (between-subjects)	2,098	5,484*	2,278	1,338	.047*	7,913**	.008	.007
Time x Group (Interaction)	4,186**	9,260**	3,824	3,247	.096	2,420	.118	.280
Time x L1 Fluency (Interaction)	.851	1,213	.949	.907	.671	.688	.725	1,476
Time x L2 WM (Interaction)	.228	.000	.001	1,063	.459	.019	.029	.087

[a] df = 7

[b] df = 1

*p<.05 **p<.01 ***p<.001 (N=37)

[c] PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

As can be seen in Table 23 above, repeated measures MANCOVA results showed that time and group interaction when the role of the adjusting variables taken into consideration [F (7, 11) = 4.186, p < .001, Pillai's trace = .520] had a statistically significant effect on the mean scores of the vector representing the seven oral fluency scores, although the effect sizes were small across distinct scores. However, follow-up univariate ANOVA statistics showed that only *pruned speech rate* score differed significantly according to groups when the group and time interaction and the role of the adjusting variables taken into consideration [F (1, 33) = 9.260, p < .001, partial $\acute{\eta}2$ = .219].

No significant effect was found on any other multivariate tests. However, follow-up univariate ANOVA statistics showed that group effect was significant for three of the oral fluency scores: *pruned speech rate* [F (1, 33) = 5.484, p < .05, partial $\acute{\eta}2$ = .142], *duration of silent pauses* [F (1, 33) = .047, p < .05, partial $\acute{\eta}2$ = .001] and *number of filled pauses* [F (1, 33) = 7.913, p < .01, partial $\acute{\eta}2$ = .193].

When the MANOVA and MANCOVA analyses are compared, it can be seen that the adjusting variables, L1 fluency and L2 speaking span, has a significant role in adjusting group differences over time.

Research Question 1.a.: The Effect of Explicit Instruction of FS on L2 Oral Fluency Over Time

The sub-question 1a examined the effect of explicit instruction of formulaic sequences on the participants' oral fluency over time. A repeated measures MANOVA analysis was conducted with pretest and delayed posttest scores of experimental group (N=10) and comparison group (N=9). Table 24 below shows the result of MANOVA analysis.

As can be seen, time had a statistically significant effect on the mean scores of the vector representing the seven oral fluency scores [F (1,11) = 31,143, p < .001, Pillai's trace= .952]. Follow-up univariate ANOVA statistics showed that time had a significant effect on *pruned speech rate* [F (1, 17) = 102,676, $p<.001$, partial $\acute{\eta}^2$= 858], *articulation rate* [F (1, 17) = 204,846, $p<.001$, partial $\acute{\eta}^2$= 923], *number of filled pauses* [F (1, 17) = 7,947, $p<.05$, partial $\acute{\eta}^2$= 319] and *repetitions* [F (1, 17) = 95,306, $p<.05$, partial $\acute{\eta}^2$= 285].

No effect was found on the effect of interaction of time and group. However, according to follow-up univariate ANOVA statistics, the interaction between time and group had a significant effect on *articulation rate* [F (1, 17) = 12,953, $p<.01$, partial $\acute{\eta}^2$= .923]. This result shows that when the effect of time and group interaction is taken into consideration, *articulation rate* significantly increased from pretest to delayed posttest. As can be seen in Table 24, *articulation rate* increased for both groups from pretest to delayed posttest but more for the experimental group in delayed posttest.

Table 24
Repeated Measures MANOVA with Delayed Posttest Results on Oral Fluency Measures (Interaction Between Time and Group)

Multivariate tests [a]		Univariate ANOVAs [b]						
Source	F	F values for L2 Oral Fluency Measures						
	(overall)	(PSR)[c]	(AR)	(NSP)	(DSP)	(NFP)	(RPT)	(RPR)
Time (within-subjects)	31,143***	102,676***	204,846***	.335	1,645	7,947*	6,789*	3,990
Group (between-subjects)	1,782	.261	2,201	.486	.710	1,253	.000	.186
Time x Group (Interaction)	2,534	.419	12,953**	1,498	.134	1,894	.009	.175

[a] df = 1
[b] df = 1
*p<.05 **p<.01 ***p<.001 (N=18)
[c] PSR: Pruned speech rate, AR: articulation rate, NSP: Number of silent pauses, DSP: Duration of silent pauses, NFP: Number of filled pauses, RPT: Number of repetitions, RPR: Number of repairs

Table 25
Means and SD of L2 Oral Fluency Scores (pretest-delayed posttest)

L2 Oral Fluency Scores	Experimental Group (N=10)		Comparison Group (N=9)	
	Pretest	Delayed Posttest	Pretest	Delayed Posttest
Pruned Speech Rate	1.96 (.30)	3.34 (.40)	2.10 (.33)	3.32 (.42)
Articulation Rate	2.04 (.31)	3.71 (.30)	2.19 (.34)	3.18 (.41)
Number of Silent Pauses	.29 (.11)	.23 (.08)	.24 (.07)	26. (.06)
Duration of Silent Pauses	.27 (.13)	.20 (.06)	.29 (.15)	.25 (.12)
Number of Filled Pauses	.12 (.07)	.21 (.08)	.18 (.05)	.21 (.05)
Number of Repetitions	.10 (.03)	.13 (.03)	.09 (.02)	.13 (.03)
Number of Repairs	.10 (.04)	.10 (.02)	.13 (.02)	.13 (.03)

Research Question 2: The Role of FS Use in L2 Oral Fluency

The second research question asked whether there was a relationship between the number of formulaic sequences the participants used and their L2 oral fluency scores. The Shapiro-Wilk test was significant across the dependent variables (p.<001), suggesting a non-parametric distribution of scores across variables. For this reason, a Linear Regression analysis was not chosen to enlighten a possible linear relationship between the number of FS and L2 oral fluency. Instead, a Spearman' rank-order correlation was run on SPSS package program to determine the relationship between the number of FS used and dependent variables. Table 26 below shows the relationship between the number of FS and posttest scores.

Table 26
Spearman Correlation Coefficients of the Number FS Used in Posttest and Dependent Variables

	Number of FS used in posttest	Pruned speech rate in posttest	Articulation rate in posttest	Number of silent pauses in posttest	Duration of silent pauses in posttest	Number of filled pauses in posttest	Number of repetitions in posttest
Pruned speech rate in posttest	-.136						
Articulation rate in posttest	-.046	.970***					
Number of silent pauses in posttest	-.181	-.205	-.172				
Duration of silent pauses in posttest	.224	.610*	.560*	-.099			
Number of filled pauses in posttest	-.188	-.052	-.041	.181	-.194		
Number of repetitions in posttest	.009	.210	.255	-.405	.093	-.046	
Number of repairs in posttest	-.220	-.093	-.122	.288	.097	.203	-.304

*p<.05 **p<.01 ***p<.001 (N=18)

The Spearman rank-order correlation test showed that there was a high positive correlation between some of the dependent variables. For example, there was a positive high correlation between *pruned speech rate* and *articulation rate* (r_s =.970, p<.001), *duration of silent pauses* and *pruned speech rate* was in positive correlation (r_s =.610, p<.05) and *articulation rate* (r_s =.560, p<.05). However, there was not a significant relationship between the number of FS the participants used in their speech and their fluency scores.

Table 27 below shows the relationship between the number of FS and posttest scores collected at the end of the semester (7 weeks after the posttest).

Table 27
Spearman Correlation Coefficients of the Number FS Used in Delayed Posttest and Dependent Variables

	Number of FS used in delayed posttest	Pruned speech rate	Articulation rate	Number of silent pauses	Duration of silent pauses	Number of filled pauses	Number of repetitions
Pruned speech rate	-.192						
Articulation rate	-.131	.280					
Number of silent pauses	.633*	-.181	.003				
Duration of silent pauses	-.031	-.405	.183	-.064			
Number of filled pauses	.344	-.557	.049	.440	.413		
Number of repetitions	.025	.492	.254	.055	-.526	-.169	
Number of repairs	.246	.099	.454	.329	-.238	-.223	-.050

*p<.05 **p<.01 ***p<.001 (N=10)

The Spearman rank-order correlation test showed that there was a significant positive correlation between the number of FS the participants used in their speech and *number of silent pauses* (r_s = .633, p<.05).

Research Question 3: The Effect of Explicit Instruction of FS on L2 Oral Accuracy and L2 Oral Complexity

The third research question examined the effect of FS instruction on L2 oral accuracy and L2 oral complexity as compared to academic vocabulary instruction. First, the relationship between L2 oral fluency, accuracy and complexity was examined to measure the overall L2 oral performance of the experimental group on posttest time. Posttest scores of the experimental group were analyzed in a Spearman rank-order correlation test because the scores showed non-normal distribution as described in the previous subsection.

Table 28 below shows the result of the Spearman rank-order correlation test. As can be seen, oral fluency scores were not correlated with L2 oral accuracy scores. Among the 3 oral complexity scores, *syntactic variety* was in negative correlation with two oral fluency scores. There was a moderately high negative correlation with *pruned speech rate* ($r_{s=}$ -.532, p.<05) and *articulation rate* ($r_{s=}$ -.504, p.<05). In other words, in the posttest the experimental group participants' syntactic variety decreased as their overall L2 oral fluency and speech rate increased.

Table 28
Spearman Correlation Coefficients of L2 Oral Fluency, Accuracy and Complexity Scores on Posttest

	Pruned speech rate	Articulation rate	Number of silent pauses	Duration of silent pauses	Number of filled pauses	Number of repetitions	Percentage of error-free clauses	Percentage of correct verb forms	Syntactic complexity	Syntactic variety	Mean segmental type-token ratio
Articulation rate	.970***										
Number of silent pauses	-.205	-.172									
Duration of silent pauses	.610*	.560*	-.099								
Number of filled pauses	-.052	-.041	.181	-.194							
Number of repetitions	.210	.255	-.405	.093	-.046						
Number of repairs	-.093	-.122	.288	.097	.203	-.304					
Percentage of error-free clauses	-.156	-.149	-.117	-.155	.294	.103	-.396				
Percentage of correct verb forms	-.144	-.172	.449	-.050	.066	.068	.291	.290			
Syntactic complexity	.003	-.026	.086	-.094	.375	.185	-.002	.248	.096		
Syntactic variety	-.532*	-.504*	.132	-.389	.062	-.048	.314	.173	.357	.305	
Mean segmental type-token ratio	.014	-.015	-.065	-.084	-.292	.028	-.133	.137	-.025	.183	.292

*p<.05 **p<.01 ***p<.001 (N=18)

The main aim of the third research question was to investigate whether there was a significant gain in L2 oral accuracy and complexity after receiving explicit instruction of formulaic sequences as compared to the academic vocabulary instruction. A repeated measures MANOVA test was conducted in order to determine the main effects of group (experimental / comparison) and time (pretest – posttest) and also the effect of interaction between group and time.

The results of the multivariate statistical test Pillai's Trace were reviewed to determine if there were significant differences between the two groups on a linear combination of the dependent variables. The assumption of equality of covariance matrices was first tested through Box's M Test. The assumption was then tested through Levene's test of normality. The results showed normal distribution of scores, eliminating the chance of making a Type I error.

Table 29 below shows the result of the repeated measures MANOVA and Univariate ANOVAs.

Table 29

Repeated Measures MANOVA with Delayed Posttest Results on L2 Oral Accuracy and Complexity Scores (Interaction Between Time and Group)

Multivariate tests [a]		Univariate ANOVAs [b]				
Source	F	F values for L2 Oral Accuracy and Complexity Measures				
	(overall)	(ACC1)[c]	(ACC2)	(COMP1)	(COMP2)	(COMP3)
Time (within-subjects)	590,151***	136,926***	439,105***	335,444***	227,979***	2272,66***
Group (between-subjects)	387,547***	,465	,362	,446	,081	1387,7***
Time x Group (Interaction)	5200,76***	1,235	,061	1,130	9,087**	3197,08***

[a] df = 5

[b] df = 1

*p<.05 **p<.01 ***p<.001 (N=37)

s[c] ACC1: Percentage of error-free clauses, ACC2: Percentage of correct verb forms,COMP1: Syntactic complexity: The ratio of T-clauses to T-units, COMP2: Syntactic variety: The number of different grammatical verb forms, COMP3: Mean Segmental Type-Token Ratio.

The results of repeated measures MANOVA showed that time [F (5, 31)= 590.151, p < .001, partial $\acute{\eta}^2$ = .990] had a statistically significant effect on the mean scores of the vector representing the 2 oral accuracy scores and 3 oral complexity scores. The effect sizes were also large across distinct scores. Follow-up univariate ANOVA statistics showed that both groups made significant gains in L2 oral accuracy and complexity across all scores. Table 30 below shows the descriptive statistics of the scores.

As for the interaction effect between time and group, the FS and academic vocabulary instruction had a significant effect on L2 oral accuracy and complexity scores [F (5, 31)= 5200.76, p < .001, partial $\acute{\eta}^2$ = .999]. Significant gains were specifically observed in COMP2 (Syntactic variety: The number of different grammatical verb forms) and (COMP3: Mean Segmental Type-Token Ratio). When the means in Table 30 are taken into consideration, in can be seen that the experimental group outperformed the comparison group in terms of *Mean Segmental Type-Token Ratio* after receiving FS instruction [F (5, 31) = 3197.08, p < .001, partial $\acute{\eta}^2$ = .997, Pillai's trace= .990], while the comparison group outperformed the experimental group in terms of *syntactic variety* after receiving academic vocabulary instruction [F (5, 31) = 9.087, p < .001, partial $\acute{\eta}^2$ = .984, Pillai's trace= .990].

This finding shows that the experimental group scored significantly higher to the comparison group on *Mean Segmental Type-Token Ratio* (measured as the number of different words divided by the total number of words in every 40-word segment) after receiving explicit instruction of FS. Another score that showed a significant increase was *syntactic variety* which was operationalized as the number of different grammatical verb forms. No significant change was observed for any L2 oral accuracy scores and syntactic complexity in terms of the effect of the interaction between group and time.

Table 30
Means and SD of L2 Oral Complexity and Accuracy Scores

L2 Oral Accuracy and Complexity Scores	Experimental Group (N=10)		Comparison Group (N=9)	
	Pretest	Posttest	Pretest	Posttest
Percentage of error-free clauses	.77 (04)	.86 (.04)	.77 (.04)	.84 (.04)
Percentage of correct verb forms	.83 (.04)	.92 (.04)	.82 (.03)	.91 (.03)
Syntactic complexity	.58 (.06)	.67 (.05)	.59 (.05)	.69 (.03)
Syntactic variety	9.11 (1.71)	13.05 (1.83)	9.63 (1.42)	13.26 (1.40)
Mean Segmental Type-Token Ratio	76.27 (5.44)	83.78 (4.79)	76.84 (4.01)	83.26 (4.39)

Research Question 4: Perceived Effectiveness of Explicit Instruction of FS

The last research question examined the participants' perceptions about effectiveness of explicit instruction of FS. Data was collected through semi-structured interviews. The participants answered questions about the explicit instruction, the extent of its possible benefits and whether or not they plan to use the items they learned in the future. The inter-coder analysis was given in Methodology.

The research question is mainly interested in the perceived effectiveness of instruction received by the experimental group. However, comparison group interviews were also coded and analyzed to detect the opinions directly related to FS because the comparison group was also exposed to a similar type of instruction that involved communicative instruction and concordancing.

Table 31 below summarizes the number of themes emerged from experimental and comparison group interviews. As can be seen, 137 codes emerged from the experimental group interviews, whereas only 61 codes emerged from the comparison group interviews. As observed by the researcher, the interviewees in the comparison group were not as elaborate as experimental group participants about the instruction they received.

As can also be seen, similar patterns emerged from the experimental group and comparison group interviews, probably due to the similarities between the instructions the two groups received. The main difference is observed in the last theme. While the experimental group cited the important contribution of confidence they gained by the explicit instruction of FS, while the comparison group focused their opinions on the functionality of academic vocabulary instruction they received.

Table 31
Frequencies and Percentages of Themes

Themes	f	%
Experimental Group (N=18)		
Academic usefulness	65	48
Awareness about contextual use	39	28
Confidence	33	24
Comparison *Group (N=19)*		
Academic usefulness	25	40
Awareness about contextual use	26	42
Functionality	10	17

As can be seen in Table 31, 3 main themes emerged in both groups and they were similar in the first two themes: *academic usefulness* and *awareness about contextual use.* The groups differed only in the third theme. The experimental group stated that they gained *confidence* thanks to the explicit instruction of FS, whereas the comparison group did not mention about this effect of explicit instruction of academic vocabulary. Instead, the comparison group interviews yielded a distinct theme: *functionality*. This theme reflected the participants' positive opinions about the academic vocabulary instruction because it helped them to learn and remember new items easily.

Table 32
Frequencies and Percentages of Themes and Categories of the Experimental Group

Themes	f	%
Academic usefulness		
ESP/EAP needs	20	31
Meaningful/Easy to learn	22	34
Functions of chunks	12	18
Content quality	8	12
Lasting effect	3	4
Awareness about concordances		
Learning contextual use	17	44
Options/variability	13	33
Use of search engines	9	23
Self-confidence		
Motivation /will to speak	27	81
Content quality	6	19

(N=18)

Academic Usefulness

The participants' responses yielded a clustering round the central theme of Academic Usefulness for both groups. This theme included the most comprehensive category of codes for both groups. As for the experimental group, the most significant categories under this theme were the effect of FS instruction on the participants' EAP/ESP needs, its perceived meaningfulness and functionality, lasting effect and its positive effect on the perceived content quality of speech.

As can be seen in the Table 32 above, the participants in the experimental group stated that the explicit instruction of formulaic sequences helped them with their EAP/ESP needs. As described in the Methodology, the present study was an instructional study that integrated the formulaic sequences intervention into the regular EAP program. EAP programs aim to cater for EAP and ESP needs of students and this result indicates that the FS instruction catered for their needs and more. Participant 10 stated:

"I keep the notes of this course. I used them in the HR course last week. I had a presentation and used them there. I think my professor was also pleased".

The participant further stated that such structures should be taught in their other classes because they need them in all of their courses. Participant 8 agreed this idea stating that

"it is bad that FS are neglected in their other courses. We actually need them."

because they are very useful in their studies. Four of the 18 participants in the experimental group stated the same idea, emphasizing their need to learn more FS for their use in other classes. By referring to the use of concordances coming from spoken academic corpora, participant 7 explained:

"in department classes, we are required to do a lot of presentations, but we do not know so many chunks. We hear them often, but we do not use them. It (the instruction of FS) gave me idea to use them after seeing the examples from other classes".

The next category is related to the meaningful and easy way of presenting and learning FS. Some participants in the experimental group stated that their professors and teachers usually gave them papers and articles to read, but they did not include any FS and their teachers did not focus on teaching such meaningful and functional parts of speech. Participant 11 said that he

"found the instruction meaningful because FS are a part of natural speech and they are almost always neglected by other teachers".

Similarly, participant 2 stated that FS instruction was generally useful, but he did not like the fact that it was cut short:

"the FS instruction is very useful because it helped us to learn a great deal of FS, but I was surprised that we did not continue to learn more during the semester."

Some participants stated that seeing the FS in concordances helped them to realize how they are used in real life. According to them, use of concordances and

practicing the FS in communicative activities was a meaningful method and helped them to remember the FS easily while they were speaking. Participant 8 remarked

> *"I found it very easy to remember words (in the instruction) while speaking in class".*

Another category that emerged from the interview data analysis was Functionality. Some participants were highly satisfied with the presentation of the FS in accordance with their functions. Participant 2 stated:

> *"it is more meaningful to learn them (formulaic sequences) with their functions because that is how they are used in real life. There are a lot of words to remember while speaking. I memorized them very easily, I think, because they were in a list (that included functions)".*

Participant 13 said:

> *"it was good to see them like that (in concordances). Most of the time, for example, while listening to teachers or watching videos, we do not realize their full form or, sometimes (their) functions."*

Similarly, participant 4 stated that FS

> *"are not like single words or articles. It is difficult to catch them while listening, you have to see their functions and written form".*

For this reason, he appreciated the way they were presented in the intervention. Participant 4 explained that after the instruction he realized the importance of learning chunks in accordance with their functions; in that way he

> *"was able to remember them while doing presentations. I impressed my teacher".*

Content quality was another category that emerged from the interview data. Some participants stated that the use of FS in their speech increased the content quality of their speech. They mostly referred to the amount of words they used in their speech; for example, participant 12 stated that FS

> *"helped them to speak more and have more content in their speech."*

Participant 3 stated that FS they learned throughout the semester

> *"helped them to speak what they intended to speak not only in the EAP class, but also in other classes."*

Participant 15 elaborated on this by saying that the use of new structures (*FS*) in speech helped him to think about what he wants to say and to want to speak more. Not to add more to the speech without purpose, but to elaborate more on what he intends to say. He further explained:

> *"I can speak more because I can think more while speaking with the help of the FS he learned".*

Participant 6 emphasized the role of FS use in decreasing the planning effort in speech and said that

> *"chunks are useful. They serve me well while I am speaking. I am speaking effortlessly, rather than just trying to speak. I can say longer sentences and the content is good. I think my other teachers will also realize that."*

Similarly, participant 8 explained that learning FS will help their future studies in many ways:

> *"I needed this type of encouragement (to use FS) in an English class because the classes are going to get more difficult in years to come and the content of the classes and course requirements will get more difficult. By using FS, one can speak more and do even more difficult presentations".*

Lasting effect was the last theme that emerged in terms of academic usefulness. Some participants were more interested in learning and using FS more because they can remain in the mind for a longer period and they are not so easy to forget in contrast to single words. Participant 10 stated:

> *"I practiced the FS outside the class to use them in other classes and in my other studies; however, I do not have to spend a lot of time practicing them because they are easy to remember in the long term; the instruction also inspired me to continue practicing FS for my other studies. They have a long term effect."*

Moreover, three participants stated that forgetting the FS they learned would be very difficult because they learned them in context, emphasizing that they learned them by heart because they learned how each of them were used in real-life.

Awareness About Concordances

The second theme was centered round Awareness about Concordances. The majority of the participants in both groups stated that they were not aware of the use of concordances in EAP classes. For many participants, Learning the Contextual Use of FS was very important, because they were not fully aware of which contexts they were used in. Participant 1 responded to the interview questions by asking

"Why should we not learn them as they are used in real life? Now that we saw that they are used like this in real life, why don't we learn all chunks like this?"

Similarly, participant 3 commented:

"I had no idea about the existence of some of the FS presented in the instruction because I was just used to looking up words and phrases in dictionary, but I do not find this habit useful anymore."

10 participants appreciated learning about the real life use of FS and emphasized that concordancing helped them realized the real function of the new FS they come across in their studies and overall language learning experiences.

Variability was another category that emerged from the responses of the experimental group participants. Some participants stated that they benefited from the concordances because they realized that FS, even if they seemed conventional, actually had variability: they had different forms than they realized and they were used in different ways than they thought. Participant 5 said:

"I learned so many chunks and discovered their different uses. Not only the ones in lines (concordances), but also the ones I found after class. I think I learned that there is always more to learn because for every function, there is more than one to use."

Participant 16 stated that she was surprised to discover the many uses and variability of FS by saying:

> *"I was amazed to learn that some chunks were so short and some of them had different forms and different uses".*

As can be seen from the responses of the participants, explicit instruction of FS through concordances opened a new vista for the participants to explore the functions and variability of FS they come across in their studies. Some students stated that they perceived concordances as a new way of learning, which is linked to the next category.

The last category was the Use of Search Engines as a tool of learning chunks. Closely linked to the two categories cited above, the use of search engines was a popular strategy among the participants in both groups. Some participants of the experimental group stated highly positive opinions about concordancing and stated that concordancing was very similar to one of the strategies they used very often in their studies and overall language learning experiences. Participant 4 explained:

> *"search engines provide similar services. When I learn a new word or structure, I search it online to see how they are used in real life".*

However, each of the nine participants who mentioned this strategy in their responses stated that they searched for alternative FS when they hear or learn new chunks because the FS instruction inspired them to learn chunks according to their functions.

<u>Self-confidence</u>

Self-confidence was the last theme that the responses of the experimental group clustered around. The categories that emerged under this theme were increase in motivation to speak and maturity of speech. This theme is not found in the comparison group interviews.

Experimental group students generally were satisfied with the FS instruction because learning FS encouraged them to be more confident while speaking and helped

them deal with their shyness. Participant 9 explained how learning FS helped her to overcome her shyness as follows:

> *"I felt somehow more confident in presentations (in class). I am usually shy, you know. But I was able to think more on the content of what I would say next because chunks helped me."*

Participant 2, on the other hand, did not have the same problem, but she stated that learning FS has helped her to produce longer utterances by saying:

> *"I am not usually shy. But I think I feel I can speak for longer now because I know a lot of chunks".*

Participant 8 explained the reason behind this confidence:

> *"I think we all have a lot to say, but we can't because we do not have knowledge of conventional structures. We are not native speakers of English, you know. We need to know how native speakers speak so that we can speak so freely like them."*

Participant 1 remarked that students "can speak faster by using FS" because they feel "free", while another student said he was "more motivated to speak in this class compared to other English classes thanks to the raised awareness that native speakers use chunks a lot" and he happened to see that he can use them as well. Participant 16 said that he "was more confident now because he could see that he could speak almost as fast as a native speaker." Some students also added that they were happy that they could speak more and fast.

Another category that emerged under this theme was "Maturity" of speech, referring to the increased perceived proficiency as defined by the students. 6 students in the experimental students stated that FS made their speech more "mature". Participant 3 explained what she meant by maturity was being able to deal with speed and content quality of speech at the same time:

> *"I can think more about what I want to say when I use chunks. Thanks to the instruction, I learned so many chunks. While using them, I feel like I am more competent, more mature."*

Participant 11 explained what he means by maturity of speech by saying that his speech quality was balanced.

"It is now fast and good in terms of quality, more mature."

By these comments, it is clear that FS instruction helped students to have a more balanced view of speed of speech and content quality of speaking tasks.

Differences Between Experimental and Comparison Group Responses

Interview responses showed that experimental group and comparison group participants had similar views about the explicit instruction. As can be seen from the responses of the experimental group, the participants stated positive opinions about concordancing in that it helped them to gain awareness about the contextual use of chunks. Learning chunks in concordances was found meaningful because it was easy to learn and retrieve items while speaking, which resulted in a permanent learning of the target chunks. The comparison group stated similar responses about the explicit instruction of academic vocabulary.

The main difference between the responses of experimental and the comparison group was that the FS instruction helped the experimental group participants to speak more confidently and feel more native-speaker-like. Another different perception was the effect of learning FS according to their function as in real life language. That was not possible in the comparison group's instruction.

The last different component was the theme of Functionality in the comparison group responses. What the comparison group participants referred to as Functionality of the instruction was linked to their learning needs. As described in detail, in Methodology, the comparison group participants were only exposed to a list of academic vocabulary that was found in the reading texts used in the instruction. The participants learned the weekly vocabulary in concordances just as the experimental group. Since concordancing is not a common practice in EAP classes, the comparison group students found this practice useful and meaningful. Participant 5 explained that

"learning the functions of words in real communication was very useful. I felt like discovering the (English) language."

This theme also included the use of search engines as a learning strategy just as the experimental group. 9 participants stated that the use of search engines was a strategy they used while learning new vocabulary; however, the instruction through concordances was found to be more beneficial because the examples came from real conversations in university classrooms. It is important to note that in this category students stated their perceptions about the functionality of the instruction that can be reused by the students as a learning method as opposed to the experimental group's responses about the list of functions of FS.

Summary of Results

- There was a significant increase in pruned speech rate and articulation rate of the experimental group compared to the comparison group after receiving explicit instruction of FS. In other words, FS instruction improved overall utterance fluency and speed fluency; however, no significant decrease was observed in pauses or hesitations in speech after the instruction.
- When the adjusting variables (L1 fluency and L2 speaking span) taken into consideration, the experimental group outperformed the comparison group only in overall fluency.
- The long term effect of the FS instruction as compared to the academic vocabulary instruction is only observed in speed fluency.
- There was no significant relationship between the L2 oral fluency scores and the number of FS used in posttest and delayed posttest (except from number of silent pauses in delayed posttest).
- There was a negative correlation between L2 oral complexity and L2 fluency, specifically speed fluency.
- The experimental group outperformed the comparison group in terms of one L2 oral complexity measure (Mean Segmental Type-Token Ratio) after receiving FS instruction, whereas the comparison group outperformed the experimental group in terms of syntactic variety on posttest.
- Both the experimental group and the comparison group expressed positive learning outcomes from the vocabulary instruction through concordances; the FS instruction was differently perceived to increase their confidence while speaking.

5. Conclusion

The present study aimed to test a frequently argued hypothesis that focused instruction of formulaic sequences will increase L2 speakers' utterance fluency (Barfield & Gyllstad, 2009; Gholami et al. 2017; Segalowitz, 2010; Wood, 2007, 2009, 2012). In order to achieve this aim, a quasi-experimental research design was set up. Two groups of freshmen students enrolled in an EAP course were assigned either as the experimental or the comparison group. The experimental group received explicit instruction of FS whereas the comparison group received explicit instruction of academic vocabulary. The instruction was implemented for five consecutive weeks by the researcher in both groups. The items used in the instruction were elicited systematically by scanning academic speaking coursebooks and determining their frequency of use in two spoken academic corpora (Base and MICASE). Also EAP teachers rated the target items according to their teachability. The items were presented to the participants through concordances and practiced in communicative and academic speaking tasks. The students also completed worksheets containing the target items in concordances. Before and after the instruction, the participants completed a familiarity test that evaluates whether the target items were learnt.

The research questions concisely inquired 1) whether the explicit instruction of FS increased L2 oral fluency, 2) whether there was a correlation between the number of FS and the participants' L2 oral fluency scores after receiving FS instruction, 3) whether there was also an increase in the participants' L2 oral accuracy and complexity scores after receiving FS instruction, and 4) what the participants thought about the explicit instruction of FS.

Pretest, posttest and delayed posttest data was collected through a picture description task, an oral narration task and an oral argumentation test. The participants were also interviewed in their L1 (Turkish) to elicit their opinions about the instructional intervention they received. Seven oral fluency scores representing three aspects of fluent speech (*speed fluency, breakdown fluency,* and *repair fluency*) were elicited through automatic processing of speech samples and manual calculations. L2

oral accuracy and L2 oral complexity scores were also elicited from the transcriptions of the speech samples. The data coming from pretest and 2 posttests were analyzed in MANOVA and MANCOVA tests, in which fluency gains were adjusted for L1 fluency and L2 speaking span capacity.

The key findings of the study are discussed below in accordance with the research questions and sub-questions by corresponding to the studies in the SLA literature.

Research Question 1: The Effect of Explicit Instruction of FS on L2 Oral Fluency

The main aim of this study was to examine the effect of explicit instruction of FS on L2 oral fluency. The hypotheses for this research question was formulated in accordance with previous studies proposing that fluent performance is characterized by minimum amount of silent and filled pauses, self-corrections, hesitations, false starts and repetitions in speech and faster speech rate (Bosker et al. 2013; Cucchiarini et al. 2002; De Jong et al. 2013; Derwing, 2017; Kahng, 2014; Koponen & Riggenbach, 2000; Kormos, 2006; Lennon, 1990; Pawley & Syder, 1983; Riggenbach, 1991; Rohr, 2016; Segalowitz, 2010; Skehan, 2003; Tavakoli and Skehan, 2005; Towell et al., 1996; Wood, 2002, 2004, 2009, 2012, 2016; Wray, 2000; 2002). Therefore, possible fluency gains acquired as an outcome of explicit instruction of FS should reflect these properties of speech. Accordingly, in this study, it was hypothesized that focused instruction on FS would increase the EAP students' speech rate in L2 and would help to decrease the number of silent and filled pauses and also the length of silent pauses in their speech. It was also hypothesized that the explicit instruction of FS would decrease the self-corrections and hesitations in their speech in L2. In other words, the explicit instruction of FS was hypothesized to increase the participants' L2 *speed fluency* and decrease their *breakdown* and *repair fluency*.

The findings of the study showed that after receiving explicit instruction of FS, the experimental group, as compared to the comparison group who received explicit instruction of academic vocabulary, significantly improved in terms of *pruned speech rate* and *articulation rate*, i.e. in terms of *overall fluency* and *speed fluency.* In other words, FS instruction improved overall utterance fluency and speed fluency; however, no significant decrease was observed in *pauses* or *hesitations* in speech of the

experimental group after the instruction. Overall, these findings show that explicit instruction of FS for 5 weeks was not effective in decreasing *breakdown fluency* (pauses) and *repair fluency* of EAP learners; however, it was effective in increasing *speed fluency* and the *global fluency score (PSR).* Overall, there was an increase in pauses and repairs of the experimental group, but it was not significant when compared to the fluency gains of the comparison group who received focused instruction of academic vocabulary.

The explanation for not confirming the main hypothesis might be lying in the length of the instruction. According to Anderson (1989), it requires time and practice to proceduralize declarative knowledge. Ellis (2001) describes a fluent L2 speaker as someone who spent thousands of hours on linguistics tasks and processed millions of utterances. Accordingly, Segalowitz (2010) suggested that to promote L2 fluency development, communicative tasks should also be repetitive. The communicative tasks used in this study were not repetitive in nature, but they encouraged the use of the target items in different tasks. Even though the instruction in this study encouraged the participants to practice the FS in productive tasks and involved repetitive practice of the items, it took place only in 5 weeks. Some previous studies also found similar results. For example, Huensch and Ventura (2017) observed L2 speakers oral fluency over time and found that they did not show a persistent change in all of the L2 oral fluency scores in 6 test times. Repair fluency measures were the most resistant to the L2 fluency development; they developed significantly late and at minimum level compared to the other aspects of oral performance. This shows that even longer exposure to L2 may not be sufficient to promote L2 oral fluency at maximum capacity. In the case of this study, explicit instruction of FS seems to increase fluidity and automaticity of speech, compared to academic vocabulary instruction; however, as observed in lack of significant change in pauses, self-corrections and repetitions in speech, the participants' larger storage of FS did not positively support their planning efforts in L2 speech.

There are also contradictory findings on the effect of length of pedagogical interventions on L2 oral fluency. For example, Tavakoli, Campbell and McCormack (2016) implemented a fluency enhancement program in an EAP setting. As different from this study, the experimental group did a series of fluency-focused instruction in

which they worked on speech samples from MICASE corpus and rated native speaker's speech in terms of its speed, pausing, and repairs and studied the dysfluencies in the transcripts of her speech for 4 weeks. The results showed that after completing the fluency enhancement intervention, there was a significant improvement in fluency gains of the experimental group, but not in pausing behavior. Indeed, the experimental group produced more pauses in the posttest. Tavakoli et al. explained this finding by referring to De Jong et al.'s (2012) finding that pausing reflects a personal style of speech that can be traced back to L1 speech patterns. De Jong et al. (2015) evoked a familiar argument, proposing that pauses and repairs in L2 speech reflect a personal style rather than a hole in their L2 performance. Tavakoli et al. (2016) also discussed that the fact that the experimental group did not show significantly less pauses and repairs in posttest may be due to the developmental differences of pause and repair behaviors; in a way, these aspects of utterance fluency may be resistant to pedagogic intervention. This explanation can also be argued for the findings pertaining to the research question 1 of this study. The results showed that speed fluency gains can be achieved through short interventions of FS instruction. However, such a short intervention was not suitable for creating a significant change in pause and repair behaviors of the participating L2 speakers.

Indeed, previous studies often found contradictory results about the fluency-related phenomena and L2 fluency development. For example, Kormos and Denes (2004) did not find a strong correlation between speed fluency measures and perceived fluency whereas De Jong et al. (2013) found that speed fluency explained about 50 % of variance in L2 proficiency and that duration of pauses did not explain much of the variance in L2 fluency. As for repair fluency, there is even more contradiction between previous findings. For example, Bosker et al. (2013) found that perceived fluency was not predicted by repair fluency measures. After finding only a weak correlation between repairs and L2 fluency and no correlation between duration of pauses and fluency measures, Kangh (2014) argued that repairs occurred often in L2 as well as in L1 and might be reflecting a personal speaking style. As mentioned earlier, the same proposition is discussed by De Jong et al. (2015) as well. The apparent contradiction in these studies can be explained by the complexity of the construct of oral fluency. Similarly, the present study found no relationship between repair measures and other fluency measures. Examining fluency development of 24

L1 Spanish speaking learners over the course of 2 years, Huensch and Ventura (2017) found that speed fluency improved and pauses tended to decrease over time. However, repairs in speech did not show significant change from test 1 to test 2. They explained this result by referring to previous research that shows no steady, gradual improvement curve in repair fluency across different levels of proficiency. Lennon (1990) also found that repetition and filled pauses in speech were correlated with each other and only reflect planning in progress. Indeed he proposed that self-corrections were not critical indicators of fluency, contradicting with Skehan (2003) and Tavakoli and Skehan (2005). In a similar vein, Ellis and Barkhuizen (2005) discussed that repair phenomenon reflects choices about accuracy more than fluency.

Overall, it seems difficult to bring a fully satisfactory explanation to the different pace of development of different aspects of fluency due to lack of consistent findings about fluency development of higher proficiency level of L2 speakers. The present study is no exception. Towell et al. (2006) argued that overall improvement in speed fluency represents change in articulation, whereas changes in breakdown and repair fluency are always more difficult to interpret because they occur as a result of a change in formulation and conceptualization stages of the speech model by Levelt (1989). Levelt (1989) described a framework of oral production that occurred in three stages: conceptualization, formulation, articulation (Levelt, 1989). However, this model was inefficient to explain L2 speech and was adapted into L2 by De Bot (1992). De Bot asserted that L2 speakers mostly had difficulty from moving on from the formulating stage since they have to deal with dual lexicon. Segalowitz (2010, 2016) reformulated this framework and found 7 critical points of fluency vulnerability: conceptual preparation, grammatical encoding, lexical retrieval, morpho-phonological encoding, phonetic encoding, articulation and self-perception. According to Segalowitz (2010), in utterance fluency, mainly the last two points are of great importance: articulation and self-perception (monitoring of output) since they are directly related to the L2 oral production and less relevant to other L2 fluency components such as cognitive fluency. The findings of the present study are consistent with Segalowitz's observation. The higher proficiency learners in this study acquired significant fluency gains in terms of speed due to the FS instruction, but the instruction was not effective in creating a significant change in their pausing and repairs in speech. This may be due to the participants' monitoring of their speech

output. Again this is also congruent with previous studies. According to Lennon (1990), slips of tongue reveal the existence of self-monitoring during speech.

As mentioned earlier, the explanation of the findings pertaining to the first research question can be found in previous studies concluding that L2 fluency is a matter of style rather than of proficiency (De Jong et al., 2012, 2015; Kahng, 2014). The findings of this study can also be discussed in accordance with this explanation of L2 speech performance. Both groups in this study showed improvements in terms of utterance fluency; but surprisingly the experimental group did not show significant improvements in terms of pauses and repairs in speech after completing the FS instruction. Although some SLA theorists claim that fluent speech should include minimum amount of pauses and self-corrections, previous research found considerable amount of self-corrections and pauses in the speech of higher proficiency learners. For example, speed fluency was not found to be in strong correlation with articulatory skills in some studies (De Jong and Mora, 2017; De Jong et al., 2013, 2015; Kahng, 2014). De Jong and Mora (2017) explain that differences in L2 utterance fluency might result from other causes than L1 fluency or speaking style such as conceptualizing, formulating and monitoring (p. 11). Indeed, it is widely recognized in the SLA literature that self-corrections, false starts and hesitations of speech are natural part of articulation process and not indicators of flaw in L2 interlanguage development. L2 speakers use self-corrections and filled and silent pauses as devices characterizing their speech, just like L1 speakers. For example, Prefontaine, Kormos and Johnson (2016) found that duration of silent pauses was in strong correlation with rater's perception of fluency. They explained that the raters might be sensitive to the purpose of the pauses used by the L2 speakers and their existence, if not frequent, did not affect their judgment. Similarly, Davies (2003) suggested that L2 proficiency is characterized by not the amount of dysfluencies but their location and distribution in speech. Previous research found great differences between L1 and L2 pause phenomenon and their distribution across speech samples (for example, within or before clauses), but it is not possible to make a valid comparison since they did not use systematic measurement (De Jong, 2016b). For example, although L2 speakers paused more, both L1 and L2 speakers paused before starting a clause (De Jong, 2016b) in an effort to plan their speech ahead. This calls for the need for examining the speakers' actual purpose to interrupt their own speech. For example, using

stimulated recall interviews can shed light onto the pause phenomena from the perspective of the language user. There is also not a consistent pattern in results for different language pairs across cross-linguistic studies on L2 oral fluency. For example, De Jong and Mora (2017) found a strong correlation between L1 and L2 fluency measures; however, the duration of silent pauses was not significantly different in L1 and L2. They discuss that this was congruent with Towell et al. (1996) and De Jong et al. (2015). All of these findings show that oral fluency measures, except from speed fluency measures such as articulation rate, do not consistently correlate with each other across studies, indicating that they might be playing a role in creating differences in individuals' speaking style. In fact, not only fluency patterns but also the patterns of FS use in speech were also found to be characterizing personal speaking style. Towell et al. (1996) conducted a 4-year longitudinal study with a small group of college students and found no change in speed or pause length in speech over time. However, the fluency gains were observed in longer runs in speech. They also found that less fluent speakers used less FS, and fluent speakers did not use significantly more FS, but used them more effectively. They concluded that all L2 speakers used FS, but some of them achieved higher proceduralization that was observed in longer strings of utterances they were able to produce. Kormos (2006) further discusses that some speakers use filled pauses (umms, and uhhs in speech), hesitations and self-corrections and some very simple formulaic expressions (gap fillers) as a strategy for planning, indicating that not all hesitations should be perceived as a detriment in the L2 learners' competence.

The present study implemented a focused FS instruction in an EAP in which the students learned and practiced the target FS in concordances coming from a spoken corpus (MICASE). The comparison group was also exposed to a similar type of instruction, but they learned academic vocabulary instead of FS and they also learned and practiced the target FS in concordances and completed the same exercises. Thus, the design of this study allows for a comparison of the effect of FS and academic vocabulary instruction on L2 oral fluency. The participants completed a familiarity test before and after the implementation of the FS and academic vocabulary instruction to test to what extent they learned the target items. The results showed that the majority of the target structures were successfully acquired by the participants. Although academic vocabulary instruction traditionally takes place in

EAP classrooms, based on SLA theories and previous studies (Nguyen, 2014; Simpson, 2004; Towell et al., 1996; Wood, 2002, 2004, 2012; Wray, 2000; Wray and Perkins, 2000) it was hypothesized that the FS group would outperform the academic vocabulary group in terms of L2 fluency. However, this hypothesis was not fully confirmed. The FS group outperformed the academic vocabulary group only in terms of speed fluency. The effect of the FS persisted for a long time; however, it was not more effective than academic vocabulary instruction for improving the other aspects of utterance fluency, indicating that in the EAP setting, academic vocabulary instruction could be as effective as FS instruction, but major speed fluency gains should not be expected.

Lastly, one inference can be drawn from the findings about the preliminary data analysis. It was found that articulation rate is a strong predictor of pruned speech rate -- a measure that is used a global score that shows overall L2 oral fluency. Two factorial analyses also showed that articulation rate consistently loaded onto the same factor as pruned speech rate. This shows that speech rate is the primary determinant factor of L2 utterance fluency. This is consistent with previous research (Bui & Huang, 2018; Towell et al., 1996; and with its relation with perceived fluency see Prefontaine, Kormos and Johnson 2016). This result might also justify why there were fluency improvements in the experimental group in terms of speed, rather than pauses and repairs since the last two are not as important predictors as the first aspect.

The Role of Adjusting Variables on L2 Oral Fluency Gain

Experimental studies are susceptible to many variables that may affect the outcome; thus, it is important to take the effect of adjusting variables into consideration while explaining the effect of the experiment. The present study used L1 fluency and L2 working memory (i.e. L2 speaking span) as adjusting variables for their significant roles in L2 oral fluency. The MANCOVA analysis showed that when the two adjusting variables (L1 fluency and L2 speaking span) taken into consideration, the experimental group outperformed the comparison group only in overall fluency. That means that when their roles in L2 oral fluency is taken into consideration, the effect of the FS instruction still persisted.

L2 oral fluency shows great variability among L2 speakers and one source of this variability might be resulted from L1 speaking style and fluency (Segalowitz , 2010; De Jong and Mora, 2017). For this reason, L1 fluency should be used as a baseline data while examining L2 oral fluency (Segalowitz, 2010). Previous research found strong relationship between L1 and L2 fluency measures with different language pairs (Derwing et al., 2009; De Jong et al., 2015; De Jong and Mora, 2017) and L2 speech as less fluent compared to L1 (De Jong and Mora, 2017). When not taken into consideration, in an experimental setting, L1 fluency could pollute the observations made on L2 fluency. Also, WM plays a significant role between the planning and articulation stage of language production. L2 speakers usually find speaking in an L2 difficult because while they are planning the overall message, they have to deal with mental lexicon racing in their minds. This cognitive load that occurs during micro-planning is one of the reasons that make L2 speech less fluent (Segalowitz, 2010, 2016).

The MANCOVA analysis showed that there were some minor changes in F values across dependent variables; however, they were negligible. That means, when the role of L1 fluency and WM is taken into consideration, the effect of FS on L2 oral fluency persisted. However, the effect of FS instruction did not resist to the role of adjusting variables in terms of *articulation rate* score which reflects the speed of utterances in L2. This finding may be indicating that L1 fluency and WM combined has a diminishing effect on L2 speed fluency. This result is not surprising in that both of these factors were found to have a significant effect on L2 fluency. However, in this case, the effect was specifically observed in speed fluency gains after receiving FS intervention. Overall, the findings pertaining to the first research question suggest that FS instruction has a significant effect on overall L2 utterance fluency; however, it was not effective in creating a change in the participants' pausing and self-correction patterns in speech. When the specific role of the adjusting variables taken into consideration, 6 L2 utterance fluency measures were resistant to the FS instruction, while the *pruned speech rate* that measures overall L2 utterance fluency was unaffected.

Research Question 2: The Role of FS Use in L2 Oral Fluency

The second research question asked whether there was a significant relationship between participants' use of FS in speech and their L2 fluency scores after receiving FS instruction. It was hypothesized that after storing a large amount of FS, the participants in the experimental group would use them in posttest and delayed posttest tasks and would also acquire fluency gains. This hypothesis relies on the premise that FS are stored and retrieved as a whole, allowing automatized, effortless speech, with less pauses and hesitations and increased rate of speech (Conklin & Schmitt, 2008; Simpson, 2004; Wood, 2002; Wray, 2002) by allowing easy processing of linguistic information by the speaker and making it easier to plan the speech ahead (Skehan, 1998).

The present study found a significant fluency gain of the experimental group after receiving FS instruction, however, data analysis showed no significant relationship between the L2 oral fluency scores and the number of FS used in posttest and delayed posttest (except from number of silent pauses in delayed posttest). This finding is surprising because the main hypothesis of this study was that explicit instruction of FS would increase the participants' L2 oral fluency and logically a significant correlation was expected to exist between the number of FS used in speech and L2 oral fluency scores. This hypothesis was not confirmed. The fluency gains observed after the completion of FS intervention may be explained by the length of the instruction. It is possible that the participants simply did not use as many FS as expected since they did not have much time to learn, practice and proceduralize the use of FS autonomously in oral tasks. During the intervention, the EAP students practiced the use of FS and completed worksheets containing the target items in classroom. At the end of the intervention, they completed a familiarity test that assessed their knowledge of the target items and they were successful to a great extent; but it was a paper based test that did not show whether the participants proceduralized the use of FS or not. According to ACT theory by Anderson (1989), declarative knowledge can turn into procedural knowledge through practice. Therefore, this finding shows that the practice of target items in this study was not enough to turn declarative knowledge (explicit instruction of instruction) into procedural knowledge (FS acquisition).

This finding is also congruent with some previous research. For example, Boers et al. (2006) implemented a FS noticing instruction for 22 hours with higher proficiency learners, as opposed to only 10 hours of explicit, concordance-based instruction of FS instruction implemented in this study. They found that the participants built up a large repertoire of FS; however the experimental group did not produce more FS than the comparison group in the posttest interviews, indicating the need for more including activities containing mnemonic strategies. They explained that noticing activities were not effective in adding new FS into the participants' linguistic repertoire for active use. The same explanation can also be applied to the findings of this study. There are also some contradictory findings in the literature. For example, Wood (2007) examined the relationship between FS use and L2 oral fluency with Chinese ESL learners studying abroad. The participants had significant fluency gains and showed increased FS use after six months. Wood argued that FS use facilitated a considerable improvement in oral fluency scores.

Another possible explanation is that only the correctly used FS were taken into consideration during the analysis. Also variations of the original forms or approximations of chunks were not taken into account. That means even if the students used as complex and appropriate FS other than the ones given in the intervention, they were not included in the analysis. Ortaçtepe (2013) found that English L2 speakers with Turkish L1 background tended to use freely generated utterances more often than formulaic expressions. However, the method found in this study eliminated a large amount of approximations of chunks in data analysis, in the name of not polluting the distinctive effect of FS instruction implemented. A different research design might also have polluted the results by not ignoring the autonomous learners' independent learning of FS outside the classroom, nullifying the effect of instruction of FS on L2 oral production.

Research Question 3: The relationship between L2 oral fluency, accuracy and complexity

The motivation for postulating the third research question lies in the different approaches to defining fluency. Tavakoli and Hunter (2018) described fours approaches to define fluency for research and classroom practices, moving from the

very narrow sense that refers to temporal qualities as also was the focus of the first research question of this study, to the very broad sense that refers to the overall L2 proficiency. The third research question deals with a broad definition of L2 fluency with a systematic measurement of L2 oral accuracy and complexity. Research question 3 looked into the relationship between L2 oral fluency, L2 oral accuracy and L2 oral complexity and also the L2 oral accuracy and complexity gains due to focused FS intervention. This question is based on the literature pertaining to the three aspects of speech and the interrelations between them. It was hypothesized that FS instruction would have a facilitating effect on L2 oral accuracy and complexity. However, the results showed that the FS instruction did not significantly improve the L2 oral accuracy scores of the experimental group, as compared to the comparison group. On the other hand, L2 oral complexity was found to be significantly affected by instruction type. While the FS instruction facilitated *Mean Segmental Type-Token Ratio* of the experimental group, the academic vocabulary instruction significantly improved *syntactic variety* of the comparison group. This finding is not totally surprising when the operationalization of these measures is considered. *Mean Segmental Type-Token Ratio* was calculated by dividing the number of different words by the total number of words in every 40-word segment. In a way, this measure encapsulates aspects of speech attributable to speed. For this reason, it is not surprising that the experimental group who also acquired greater speed fluency gains by FS instruction also outperformed the comparison group on this measure. *Syntactic variety*, on the other hand, was operationalized as the number of different grammatical verb forms. As academic vocabulary instruction contained a great deal of new verbs for the comparison group to practice, it is also not surprising that the comparison group who acquired a large repertoire of new verbs thanks to the instruction outperformed the experimental group in terms of *syntactic variety*. *Syntactic variety* was also found to be in negative correlation with the *pruned speech rate* and *articulation rate* scores of the experimental group.

The finding pertaining to the lack of significant relationship between L2 oral fluency, accuracy and complexity scores can be explained by the nature of the interrelations between the three aspects of speech as widely discussed in the literature. As mentioned earlier, the motivation for this question emerged from the literature defending that fluency, accuracy and complexity should be the part of an underlying

system in which developing interlanguage system involves greater control of the language, reduction in error, development of appropriate monitoring systems and automatization (Skehan, 1998). In a language view such as this, fluency should be in correlation with other aspects (Foster & Skehan, 1996). Wood (2012) also agrees that these three aspects of speech may not be as distinct from each other and change in one aspect would affect another. Because human beings have limited attention capacity and L2 learners do not have full control of L2, a trade-off is expected to occur among accurate, complex and fluent performance of L2 learners especially in instructional settings (Foster & Skehan, 1996; Foster & Tavakoli, 2009). Indeed, Robinson (2011) discusses further that fluency contrasts with accuracy and complexity. In the present study, a moderate negative correlation was found between one L2 oral complexity score (*syntactic variety*) and L2 fluency, specifically speed fluency. This can be explained by the participants' limited attentional capacity in L2 to deal with both the articulatory skills and the content of their speech. According to Skehan (1996), overachievement on one aspect of speech results in a situation he calls *undesirable fluency* in which speakers fluently produce output for the sake of pretending to speak faster, by ignoring the accuracy, lexicalization and communicative appropriateness of their utterances. For effective fluency, a balance between all aspects should be maintained; additionally, to a large repertoire of FS should be acquired for active use.

The findings pertaining to the effect of instruction type on L2 oral accuracy. The results showed that FS instruction significantly improved the participants' L2 oral accuracy; but, when its effect was compared to the effect of academic vocabulary instruction, the effect did not persisted. This finding is also difficult to discuss since there are no instructional studies that investigated the effect of FS instruction on L2 accuracy and SLA literature contains a great amount of contradictory arguments about the relationship between aspects of speech and FS acquisition and use. For example, Jiang and Nekrasova (2007) argue that when learnt properly FS can increase fluency as well as accuracy in linguistic tasks. However, Segalowitz (2010) discusses that a trade-off between accuracy and fluency should be expected in L2 speakers due to the cognitive demand of self-monitoring. Kormos (2006) further argues that some speakers are fluent because they choose to compensate for their weaknesses in other areas of speech such as their accent and oral accuracy.

As for the effect of FS instruction on L2 oral fluency, a persistent effect was not found across all complexity measures. In fact, FS instruction and academic vocabulary instruction facilitated different complexity scores. FS instruction significantly improved the number of different words used per 40 words while academic vocabulary instruction group significantly improved in the use of different verb forms. In fact, when the nature of the scores is examined, this finding is not very surprising. The academic vocabulary instruction contained a large amount of academic verbs; so the participants in the comparison group outperformed the experimental group in terms of using different verb forms while speaking in L2. The finding that FS instruction significantly improved the number of different words in speech is also congruent with the literature. According to Pawley and Syder (1983) fluency involves multiple clause-chaining due to a large storage of FS (p. 202). In other words, fluency is interwoven with complexity of speech, as supported by the use of FS. When examined from this perspective, the finding about the effect of FS on facilitating the number of different words is not surprising. The acquisition of a large repertoire of FS helped the participants outperform the comparison group in terms of capability to add variety (and complexity) to their speech.

Research Question 4: Perceived Effectiveness of Explicit Instruction of FS

One aim of the present study was to explore the participating students' perceptions about the instruction they received. The interview questions were designed in order to stimulate retrospective thinking on the explicit instruction of FS and academic vocabulary. The interviews were conducted in Turkish and in a semi-structured fashion allowing the participants to elaborate on how they found the instruction, their opinions on the extent of its possible benefits and whether they would want to use the items they learned in their future studies.

The qualitative results showed that both the experimental group and the comparison group expressed positive learning outcomes from the vocabulary instruction through concordances. This result is congruent with the findings of previous studies. For example, Anğ (2006) found that freshmen students in an EAP setting perceived the use of concordances as useful tools to enhance their awareness about formulaic language use in research articles. In fact, three major themes that

emerged from Ang̃'s qualitative analysis were focused on similar themes with this study: increased motivation, language awareness and learner autonomy. Although the mentioned study dealt with teaching and learning FS through concordances to be used in writing tasks, these similarities between the participants' responses show that EAP learners feel that they gain positive benefits from concordancing for their productive L2 skills. Indeed, concordancing is also discussed to be a useful tool for supporting L2 vocabulary, especially FS instruction. For example, Stengers and Boers (2015) called for more contextualized vocabulary exercises to be presented in textbooks for language learning for their positive effect on learning outcomes. Similarly, Kormos and Prefontaine (2017) suggested that there is need for more research to examine L2 learners' perceptions about speaking tasks used in foreign language classrooms. One inference for language learning materials developers is that concordancing is found to be useful by learners as a tool that allows easy acquisition of FS. Also, in both groups, the students found the intervention useful and stated that they would go on using the new vocabulary in the future and in their other studies. The perceptions of the students about the effect of the intervention about their long-term language learning experience, however, was not in total congruence with the quantitative findings. The intervention had a persistent, long-term effect on only speed fluency.

The participants in this study were not fully aware of the aim and hypotheses of the study. Therefore, they did not stated comments directly addressing the research problems answered by this study. However, they expressed significant commentaries that supports and supplements the overall effect of the intervention. For example, the FS group specifically expressed that the FS instruction helped them to speak more confidently in L2. This is frequently discussed in previous research (Kormos, 2006; Pawley & Syder, 1983; Segalowitz, 2010; Wood, 2012). Due to a larger storage of FS repertoire, the participants felt that they could produce longer runs and could plan their speech ahead. This is also congruent with previous research that found the use of FS results in longer runs in L2 speech (Towell et al., 1996). The participants also stated that FS instruction increased their motivation to speak in L2. According to Segalowitz (2010), fluency acquisition is affected by "various socially grounded factors" (p. 120). Motivation and self-confidence are among these factors. As the participants stated positive opinions about the motivational effect of the intervention, it can be assumed that implementing concordance-based instruction with FS and

academic vocabulary could increase EAP students' motivation to learn and use new vocabulary, either it is FS or single, high-proficiency level words. Also, some participants in the experimental group stated that learning FS made them feel they were speaking faster and in a more "mature" way that allowed them to keep the balance between matching up the speed of their speech with effort to deliver content in an appropriate manner. In a recent study, Wood (2016) found that L2 speakers might present fluctuating levels of motivation while speaking in L2. A focused-FS instruction could help L2 speakers to catch up with demanding motivational requirements of L2 oral tasks to a great extent, as can be observed in the participants' justifications in this study.

Simpson (2004) suggested introducing a large amount of FS to EAP students according to their functions could promote better learning outcomes. This idea was implemented in the intervention in this study. It seems to be found functional by the participating EAP students as well. In fact, both groups expressed positive impression about the effect of the intervention on their EAP/ESP needs. Wood (2012) also suggested presenting new FS in context. A focused vocabulary instruction based on concordancing and productive and communicative tasks that requires repetitive use of the new vocabulary seems to be found a useful learning method by EAP students. A special focus on FS in such an instruction would comparatively create faster speaking performance and considerably increase the L2 speakers' self-confidence and motivation.

5.1. Implications

Many pedagogical implications can be drawn from the findings of this study. It is widely believed that fluency cannot be taught because it naturally and gradually develops as a result of exposure to L2 in and out of classroom (Chambers, 1997; Lennon, 1990); however, instruction can help enhancing it (Gatbonton and Segalowitz, 1988). According to Tavakoli and Hunter (2018), one line of fluency enhancing activities used in EFL classrooms are necessarily communicative, as also was executed in the present study. The participating students in the experimental group practiced FS in communicative tasks that required repetitive use of the new

items during the intervention. The findings revealed that they showed significant fluency gains after receiving FS instruction; however, the effect was not observed across pausing and repair measures. They also expressed in the interview that they found the focused FS instruction a useful tool for learning the contextual use of FS, for their role in increasing their motivation and self-confidence and meeting their EAP/ESP needs. As can be seen, EAP teachers who wish to enhance their students' L2 oral fluency can make use of focused FS instruction by way of using concordances to introduce new items and communicative tasks to encourage repetitive use of the target items. Such a program can enhance the speed fluency of their students. However, FS should be systematically elicited from academic spoken corpora and be presented to learners as they are used in real-life academic context (Wood, 2012) and also in accordance with their purpose and functions (Simpson, 2004).

However, some adjustments seem to be needed to complement such a program to facilitate all aspects of L2 utterance fluency. As suggested by Gatbonton and Segalowitz (2005) communicative EFL classrooms can enhance L2 oral fluency by including more fluency-focused activities. However, current coursebooks are reported to lack such activities (Rossiter et al., 2010). Tavakoli and Hunter (2018) suggest that EFL teachers should adapt additional fluency activities in their classrooms due to lack of fluency-related teaching materials. Rossiter et al. (2010) also suggest that EFL teachers who want to promote fluency in their classrooms should adapt fluency-focused activities by the method of their choice: presentation-practice-production (PPP) cycle, task-based instruction etc. In the present study, a PPP cycle was used to introduce and practice FS and academic vocabulary. The results coming from the familiarity test showed that the students learned the target structures after 5 weeks of explicit instruction by way of using PPP and communicative EAP tasks. So the instruction worked in terms of learning of the FS, but not in terms of increasing all aspects of fluency. In order to address other aspects of utterance fluency than mere speed fluency, interventions specifically addressing all aspects of fluency (speed, pauses, repairs, hesitations, self-corrections etc.) should be implemented. To overcome this problem, Boers et al. (2006) suggested teaching of mnemonic strategies, while De Jong and Perfetti (2011) suggested fluency-focused activities such as timed summary activities. Integrating FS instruction based on concordancing and

use of communicative EAP tasks with fluency-enhancing activities can enhance all aspects of L2 oral fluency.

Both groups in this study were exposed to a different type of vocabulary instruction. Therefore, pedagogical implications can be inferred from the comparison of fluency gains between FS and academic vocabulary groups. Keeping in mind that EAP students aim to acquire a great deal of academic vocabulary and FS in their studies, it can be inferred that EAP students could benefit both from focused academic vocabulary instruction and FS instruction to promote their L2 fluency development. FS instruction would help them speak faster. Additionally, FS instruction programs are suggested to be spread across the curriculum. As the findings suggested, a few weeks long interventions do not seem to have a long lasting effect on the acquisition of FS and their effect on L2 utterance fluency.

Similarly, it can also be inferred that mere FS instruction is not more effective than mere academic vocabulary instruction in increasing L2 speakers' oral accuracy and complexity. It was found that FS instruction significantly increased the number of different words used in speech while academic vocabulary instruction facilitated variety of verb forms used in speech. Instructors who aim to promote their students' L2 oral complexity can use both FS and academic vocabulary instruction.

One last implication could be the significance of articulation rate which seems to possess a power to measure speed fluency as well as pruned speech rate. It was found that articulation rate is a strong predictor of pruned speech rate, a measure that is used a global score that shows overall L2 oral fluency. Additionally, two factorial analyses conducted with pretest and posttest scores also showed that articulation rate loaded onto the same factor as pruned speech rate. This is congruent with previous research findings (Bui & Huang, 2018; Towell et al., 1996), implying that researchers should consider the dominant role of articulation rate in pruned speech rate measure which is consistently used in SLA research as a global L2 fluency score.

5.2. Suggestions for Future Research

Speaking in L2 is a major concern generally for EFL and specifically EAP teachers; however, as discussed in this dissertation, it has been highly neglected in SLA research. There is need for more research that would allow a deeper understanding of speaking in L2, which will help draw the framework for the future of EAP curricula.

It was acknowledged in the Introduction section of this dissertation that due to small sample size, generalizability of the findings of this study should be interpreted cautiously. Another major limitation of the study was the role of adjusting variables that were not included in the study; although the data were adjusted for working memory capacity in L2 and L1 fluency, there might be other variables that could affect the results of the study. Suggestions for future research were discussed below in accordance with the stated limitations.

The present study showed that a focused instruction of FS is not as effective as it might be thought in creating a change across all aspects of L2 utterance fluency for EAP students. As suggested earlier, future research can look into the effect of FS instruction supported by a fluency-enhancement instruction on L2 utterance fluency.

The participants of this study were freshmen students who were high-proficiency L2 learners with Turkish L1 background and the instruction took place in an EAP setting. Future research may consider examining the effect of teaching FS on L2 oral fluency of language learners of different levels of L2 proficiency and with different language pairs.

The present study examined the L2 oral fluency of the participants only from the perspective of temporal qualities, which is called utterance fluency by Segalowitz (2010). Although, it is suggested that it is the most effective approach to objectively examine the development of L2 oral fluency (De Jong, 2013; Skehan, 2003; Tavakoli & Skehan, 2005; Tavakoli & Hunter, 2018), future research may consider examining other aspects of fluency, for example perceived fluency, in an instructional study to examine L2 oral fluency development in a more comprehensive way. However, caution is suggested while doing this, since there is still not systematic and objective method to operationalize judgments on fluency.

Future research can also examine the qualities of the fluent performance after receiving FS instruction through stimulated recall interviews. In this way, participants can elaborate on the fluency vulnerability points in their speech, speaking motivation, the reasons for pausing and repair behavior during speaking. Taking the perspective of the learner would enlighten the actual interaction between the use of FS and the use of pauses, self-corrections and L2 oral fluency, complexity and accuracy.

References

Ahangari, S. & Abdi, M. (2011). The effect of pre-task planning on the accuracy and complexity of Iranian EFL learners' oral performance. *System, 29*, 1950-1959.

Ahmadian, M. J. & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity and fluency in EFL learners' oral production. *Language Teaching Research, 15* (1), 35-59.

Alderson, J. C., Clapham, C. & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research, 1*, 93–121.

Anderson, J. R. (1989). Practice, Working Memory, and the ACT Theory of Skill Acquisition. *Journal of Experimental Psychology 15* (3), 527 – 530.

Anderson, J. R. (1995). *Language and Memory: An Integrated Approach*. New York: Wiley.

Anğ, F. (2006). *Effectiveness of Corpus Consultation through Concordancing on the Formulaic Academic Language Use of Freshman ELT Students*. Unpublished master's thesis. Bogazici University.

Baars, B. J. (1997). *In the theatre of consciousness: The workspace of the mind*. Oxford: Oxford University Press.

Bardovi-Harlig, K. (2009). Conventional Expressions as a Pragmalinguistic Resource: Recognition and Production of Conventional Expressions in L2 Pragmatics. *Language Learning, 59* (4), 755-795.

Bardovi-Harlig, K. & Griffin, R. (2005). L2 pragmatic awareness: Evidence from the ESL classroom. *System, 33*, 401-415.

Bardovi-Harlig, K., Mossman S. & Vellenga, H. E. (2015). The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research 19* (3), 324–350.

Bardovi-Harlig, K. & Vellenga, H. E. (2012). The effect of instruction on conventional expressions in L2 pragmatics. *System, 40*, 77-89.

Barfield, A. & Gyllstad, H. (2009). Introduction: Researching L2 Collocational Knowledge and Development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 153-170). New York: Palgrave Macmillan.

Basturkmen, H. (2006). *Ideas and options in English for specific purposes*. New Jersey: L. Erbaum Associates.

Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition.* Cambridge: Cambridge University Press.

Biber, D. Conrad, S. & Cortes, V. (2004). If you look at... Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25* (3), 371–405.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Blake, C. G. (2006). *The potential of text-based Internet chats for improving ESL oral fluency*. Unpublished doctoral dissertation. Purdue University.

Boers, F. Dang, T. C. T. & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research 21*(3), 362–380.

Boers, F., Demecheleer, M., Coxhead, A, & Webb, S. (2014) Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*, 54-74.

Boers, F. & Lindstromberg, S. (2009). *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. New York: Palgrave Macmillan.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H. & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10* (6), 245-261.

Bosker, H.R., Pinget, A.-F., Quené, H., Sanders T. & De Jong, N.H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing 30* (2), 159-175.

Bosker, H. R., Quené, H., Sanders T. & De Jong, N.H. (2014).The Perception of Fluency in Native and Nonnative Speech. *Language Learning, 64* (3), 579–614.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy.* Cambridge: Cambridge University Press.

Bui, G. & Huang, Z. (2018). L2 fluency as influenced by content familiarity and planning: Performance, measurement, and pedagogy. *Language Teaching Research 22*, (1), 94-114.

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

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In A. Soler & S. Jordà (Eds.), *Intercultural language use and language learning* (pp. 41-57). Oxford: Oxford University Press.

Chambers, F. (1997). What do we mean by fluency? *System 25* (4), 535–544.

Conklin, K. & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29* (1), 72-89.

Clercq, B. & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. *The Modern Language Journal, 101*, 2, 315–334.

Council of Europe (2001). *A Common European Framework of reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics , 34* (3), 369–388.

Cucchiarini, C., Strik, H. & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America 111* (6), 2862–2873.

Daller, M. H., Yıldız, C., De Jong, N. H., Kan, S. & Başbağı, R. (2011). Language dominance in Turkish–German bilinguals: methodological aspects of measurements in structurally different languages. *International Journal of Bilingualism, 15*, (2) 215–236

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20* (6), 445-464.

Davies, A. (2003). *Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters.

Davies, M. J. (2014). An overview of formulaic language and its possible role in L2 fluency development. *Ritsumeikan Studies in Language and Culture 25* (3), 221-231.

De Bot, K. (1992) A Bilingual Processing Model: Levelt's 'Speaking' Model Adapted. *Applied Linguistics,* 13: 1-24.

De Jong, N. H. (2016a). Fluency in second language assessment. In: Tsagari D., Banerjee J. (Eds.) *Handbook of Second Language Assessment*.: Mouton de Gruyter. 203-218.

De Jong N. H. (2016b). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency, *IRAL* 54 (2), 113-132.

De Jong, N. H. & Mora, J.C. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition, 1-13.*

De Jong, N. H. & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61,* (2), 533-568.

De Jong, N. H., Steinel, M. P., Florijn A., Schoonen, R. & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics 34*, 893–916.

De Jong, N. H., Groenhout, R., Schoonen, R. & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36* (2), 223-243.

De Jong, N. H., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34* (1), 5–34.

De Jong, N. H., & Wempe, T. (2008). *Praat script speech rate*. Retrieved December 08, 2013, from sites.google.com/site/speechrate/.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41 (2),* 385-390.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty and M. H. Long. (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.

Denzin, N. K., & Lincoln, Y. S. (2000). *Handbook of qualitative research*. Thousand Oaks, Calif: Sage Publications.

Derwing, T. M. (2017). L2 fluency development. In (Ed.) S. Loewen & M. Sato, *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 246 -). NY: Routledge.

Derwing, T., Rossiter, M., Munro, M. J., & Thomson, R. (2004). Second language fluency: judgments on different tasks. *Language Learning, 54*, 655-679.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*, 533– 557.

Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System* 35.2, 271−280.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford, UK: Oxford University Press.

Ellis, N.C., (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt, M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Ellis, N. C. (2007) The weak interface, consciousness, and form-focused instruction: mind the doors. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honor of Rod Ellis* (pp. 55–69). Oxford: OUP.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35–47). Washington, DC: Georgetown University Press.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

Ellis, N. C., Simpson-Vlach, R. & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42* (3), 375- 396.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R. (2002a). "Does form-focused instruction affect the acquisition of implicit knowledge?" *Studies in Second Language Acquisition*, *24*, 223–236.

Ellis, R. (2002b). The place of grammar instruction in the second/foreign curriculum. In E. Hinkel and S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 17–34). Mahwah, NJ: Erlbaum.

Ellis, R. (2003). *Task Based Language Learning and Teaching.* Oxford: Oxford University Press.

Ellis, R. (2005a). Principles of instructed language learning. *System, 33,* 209–224.

Ellis, R. (2005b). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*, 431–463.

Ellis, R. & Barkhuizen, G. (2005). *Analysing learner language.* Oxford: Oxford University Press.

Ellis, R., Basturkmen, H. & Loewen, S. (2001). Preemptive focus on form in the ESL classroom. *TESOL Quarterly, 35* (3), 407-431.

Erman, B. and B. Warren (2000). The idiom principle and the open-choice principle. *Text 20*, 29–62.

Fan, M. (2009). An exploratory study of collocational use by ESL students – A task based approach. *System 37*, 110–123.

Field, A. (2009). *Discovering statistics using SPSS*. SAGE: London.

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler & W. S. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic Press.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, and

M. Swain (Eds), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 75-93). Harlow: Longman.

Foster, P. & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18,* 299-324.

Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency and lexical diversity. *Language Learning, 59* (4), 866–896

Fotos, S. & Hinkel, E. (2007). Form-focused instruction output for second language writing gains. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honor of Rod Ellis* (pp. 55–69). Oxford: OUP.

Freed, B. F. (1995). Do students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: John Benjamins.

Freed, B. F. (2000). Is fluency, like beaut, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor, MI: University of Michigan Press.

Fraenkel J.R. & Wallen, N.E. (2000). *How to Design and Evaluate Research in Education* (4th Ed.). Boston: McGraw Hill.

Fulcher, G. (2003). *Testing Second Language Speaking*. London: Longman/Pearson Education.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL quarterly, 22*, 473–492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: a focus on access to fluency. *Canadian Modern Language Review, 61*, 325-353.

Gardner, D. & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*, (3) 305–327.

Gholami, L., Karimi, M. N. & Atai, M. R. (2017). Formulaic focus-on-form episodes in adult EFL communicative interactions. *System, 68*, 72-86.

Gilquin, G. & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics 16* (2), 141–172.

Guillot, M. N. (1999). *Fluency and Its Teaching*. Clevedon: Multilingual Matters.

Gouverneur, C. (2008). 'The phraseological patterns of high-frequency verbs in advanced English for General Purposes: a corpus-derived approach to EFL textbook analysis' in F. Meunier and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.

Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam & Philadelphia: John Benjamins.

Gut, U. (2009). *Non-Native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt am Main: Peter Lang.

Harwood, N. (2002). Taking a lexical approach to teaching: principles and problems. *International Journal of Applied Linguistics, 12* (2), 139-155.

Hinkel, E. (2004). *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. Mahwah, N. J.: Lawrence Erlbaum Associates.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Philadelphia/Amsterdam: John Benjamins.

Hsu, J. T. (2010). The effects of collocation instruction on the reading comprehension and vocabulary learning of college English majors. *Asian EFL Journal, 12*, (1), 47-87.

Hu, G. (2011). Metalinguistic knowledge, metalanguage, and their relationship in L2 learners. *System, 39*, 63–77.

Hudson, T & Llosab L. (2015). Design Issues and Inference in Experimental L2 Research. *Language Learning, 65* (Suppl. 1), 76–96.

Huensch, A. & Ventura, N. T. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal 101*, (2), 275 – 293.

Hyland, K. (2012). Bundles in academic discourse. *ARAL, 32*, 150-169.

Iwashita, N. (2006). Syntactic Complexity Measures and Their Relation to Oral Proficiency in Japanese as a Foreign Language. *Language Assessment Quarterly, 3* (2), 151-169.

Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency : how difficult? *Applied Linguistics 29*, 24-49.

Jiang, N. & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language E-Journal, 91* (3), 433-445.

Jones, M. & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: an exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 269-300). Amsterdam: Benjamins.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning 64*, (4), 809-854.

Kang, O.,(2008). Ratings of L2 Oral Performance in English : Relative Impact of Rater Characteristics and Acoustic Measures of Accentedness. *Spaan Fellow. WorkingPapers 6*, 181-205.

Kang, O. (2012) Impact of Rater Characteristics and Prosodic Features of Speaker Accentedness on Ratings of International Teaching Assistants' Oral Performance. *Language Assessment Quarterly, 9* (3), 249-269.

Kang, O, Rubin D. & Pickering, L. (2010) Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *The Modern Language Journal*, *94* (4), 554-566.

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (ed.), *Planning and Task Performance in a Second Language, pp. 143-164*. Philadelphia: John Benjamins.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor, MI: University of Michigan Press.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal, 59,* (4), 322-332.

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

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*, 145–164.

Kormos, J. & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 9 (2), 1-19.

Kormos, J. & Prefontaine, Y. (2017). Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research , 21* (6), 699 – 716.

Krashen, S. D. & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning 28*, 283-300.

Lahmann, C., Steinkrauss, R. & Schmid, M. S. (2017). Speed, breakdown, and repair: An investigation of fluency in long-term second-language speakers of English. *International Journal of Bilingualism, 21* (2), 228–242.

Lambert, C., Kormos, J. & Minn, D. (2016). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39* (1), 167-196.

Landis, R. J. & Koch, G. G. (1977). The Measurement of Observer Agreements for Categorical Data. *Biometrics 33*, 159-174.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning, 40* (3), 387-417.

Leonard, K. R. & Shea, C. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal, 101*, 1, 179–193.Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge,MA: MIT Press.

Levelt, W. J. M. (1989) *Speaking: from intention to articulation*. Cambridge, Massachusetts: The MIT Press.

Lewis, M. (1993). *The Lexical Approach*. Hove: Teacher Training Publications.

Lewis, M. (2000). *Teaching collocation: Further development in the lexical approach.* Hove: Language Teaching Publications.

Liao, Y. & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English". *Language Learning 54* (2), 193–226.

Long, M. (1991). "Focus on form: A design feature in language teaching methodology". In K. De Bot, R. Ginsberg, C. Kramsch. *Foreign language research in cross-cultural perspective* (pp.39-52). Amsterdam: John Benjamins.

Mackey, A. & Gass, S. M. (2005). *Second Language Research: Methodology and Design.* New Jersey: Lawrence Erlbaum Associates.

Martinez, R. (2013). A Framework for the inclusion of multi-word expressions in ELT. *ELT Journal, 67* (2), 184-198.

Martinez, R. & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics 33*, 299–320.

McCall, R. B., & Kagan, J. (1994). *Fundamental statistics for behavioral sciences*. Fort Worth: Harcourt Brace College.

McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, 1 (1), 1-15.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20* (1), 83-108.

Miles M. B. & Huberman, M. (1994). *Qualitative Data Analysis: A Sourcebook of New Methods* (2nd Edition). Beverly Hills, CA: Sage Publications.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics, 32*, (2), 129-148.

Mirdamadi F. S. & De Jong N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research 31* (1), 105-116.

Mizera, G. J. (2006). *Working Memory and L2 oral fluency*. Unpublished doctoral dissertation. University of Pittsburg.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity and lexical density in L2 speech production. *Fragmentos, 24*, 69-104.

Nation, I.S.P. and Newton, J. (2009) *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesi, H. & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lecturers. *International Journal of Corpus Linguistics, 11* (3), 283-304.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24* (2), 223-242.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

Norris, J. & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning 50* (3), 417-528.

Nguyen, H. T. H. (2014). *The Acquisition of Formulaic Sequences in High-Intermediate ESL Learners*. Unpublished doctoral dissertation . University of Pennysylvania.

Ortaçtepe, D. (2013). Formulaic language and conceptual socialization: The Route to becoming nativelike in L2. *System 41*, 852-865.

Osborne, J. (2011). Fluency, complexity and informativeness in native and non-native speech. *International Journal of Corpus Linguistics 16* (2), 276–298.

Osborne, J. (2011). Oral learner corpora and the assessment of fluency in the Common European Framework. In Frankenberg-Garcia, A., Flowerdew, L. & Aston, G. (eds.) *New Trends in Corpora and Language Learning* (pp. 181-197). London: Continuum.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.

Peters E. & Pauwels P. (2015) Learning academic formulaic sequences. *Journal of English for Academic Purposes 20*, 28-39.

Prefontaine, Y., Kormos J. & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing 33*, (1), 53–73.

Qi, Y. & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System, 39*, 164-174.

Richards, J. C. (2002). Accuracy and fluency revisited. In Hinkel, E & S. Fotos (eds.), (2002). *New Perspectives on Grammar Teaching in Second Language Classrooms.* New Jersey: Lawrence Erlbaum Associates, Inc., Publishers: 35-49.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes, 14*, 423–441.

Robinson, P. (1995). Task Complexity and Second Language Narrative Discourse. *Language Learning 45,* (1), 99-140.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a compotential framework for second language task design. *IRAL, 43*, 1-32.

Roehr-Brackin, K. (2014). Explicit knowledge and processes from a usage-based perspective: The developmental trajectory of an instructed L2 learner. *Language Learning 64* (4), 771–808.

Rohr, J. L. (2016) *Acoustic and Perceptual Correlates of L2 Fluency: The Role of Prolongations.* Unpublished doctoral dissertation. University of Texas.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review, 65* (3), 395–412.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G. & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review, 66* (4), 583-606.

Sánchez Hernández, P. (2013). Lexical bundles in three oral corpora of university students. *Nordic Journal of English Studies,* 13, (1), 187-209.

Scheffler, P. & Cincia, M. (2011). Explicit grammar rules and L2 acquisition. *ELT Journal, 65*, 13–23.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14, 357-385.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. & Carter, R. (2004). Formulaic sequences in action: an introduction. In: Schmitt, N. (Ed.), *Formulaic Sequences*. John Benjamins, Amsterdam, pp. 1–22.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. New York: Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants . *International Review of Applied Linguistics in Language Teaching, 54* (2), 79 – 95.

Seyfeddinipur, M., Kita, S. & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition, 108,* 837-842.

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In U. Connor & T. A. Upton (Eds.), *Discourse in the Professions: Perspectives from corpus linguistics* (pp. 37 – 64). Amsterdam: John Benjamins.

Simpson, R. C., Briggs, S. L., Ovens, J. & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. The Regents of the University of Michigan.

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List (AFL). *Applied Linguistics, 31, 487-512.*

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A. (2017). Researching the teaching and learning of multi-word expressions. *Language Teaching Research, 21* (3), 289-297.

Siyanova-Chanturia, A. & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36* (5), 549–569.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17,* 38–62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2003). Task based instruction. *Language Teaching, 36,* 1–14.

Skehan, P. (2007). Task research and language teaching: Reciprocal relationships. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honor of Rod Ellis* (pp. 55–69). Oxford: OUP.

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1* (3), 1-27.

Spoelman, M. & Verspoor,M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. A*pplied Linguistics, 31*, 532–553.

Spöttl, C., & McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 191–225). Philadelphia: John Benjamins.

Stengers, H. & Boers, F. (2015). Exercises on collocations: a comparison of trial-and-error and exemplar-guided procedures. *Journal of Spanish Language Teaching , 2*, 152-164.

Stengers, H., Boers, F., Housen, A. & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *IRAL 49*, 321-343.

Swain, M. (2000). The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In J. P. Lantolf (ed.). *Sociocultural Theory and Second Language Learning.* Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In C. Doughty and J. Williams (Eds.), *Handbook on research in second language learning and teaching* (pp. 471-484). Mahwah, NJ: Lawrence Erlbaum Associates.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*(4), 433-457.

Tavakoli, P., Campbell, C. & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly , 50* (2), 447-471.

Tavakoli, P. & Hunter, A. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Languae Teacher Research.* Advance online publication. doi.org/10.1177/1362168817708462

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). Amsterdam: John Benjamins.

Thai, C. & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly, 50* (2), 369-393.

Thomson, H., Boers F. & Coxhead, A. (2017). Replication research in pedagogical approaches to spoken fluency and formulaic sequences: A call for replication of Wood (2009) and Boers, Eyckmans, Kappel, Stengers & Demecheleer (2006). *Language Teaching*, 1-9.

Thompson, P. & Nesi, H. (2001). The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research 5* (3) 263-264.

Thornbury, S. (2005). *How to Teach Speaking*. London: Pearson Education.

Thornbury, S. & Slade, D. (2006). *Conversation: From Description to Pedagogy*. Cambridge: Cambridge University Press.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *IRAL—International Review of Applied Linguistics in Language Teaching, 40,* 117–150.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84–119.

Tucker, G. R. (1997). Multilingualism and language contact: An introduction. *Annual Review of Applied Linguistics, 17,* 3-10.

Ur, P. (1981). *Discussions That Work: Task –Centered Fluency Practice*. Cambridge: Cambridge University Press.

Ushigusa, S. (2008). *The Relationships Between Oral Fluency, Multiword Units, and Proficiency Scores*. Unpublished doctoral dissertation. Purdue University.

VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287–301.

Vanderplank, R. (1993). Pacing and spacing as predictors of difficulty in speaking and understanding English. *English Language Teaching Journal, 47,* 117–125.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics 16,* (2), 180-205.

Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102-127). Ann Arbor, MI: University of Michigan Press.

Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics, 10,* 128-137.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*. 85-106.

Willis, D. (1990). *The Lexical Syllabus*. London: Harper Collins.

Wood, D. (2002). Formulaic Language in Acquisition and Production: Implications for Teaching. *TESL Canada Journal, 20* (1), 1-15.

Wood, D. (2004). An Empirical Investigation into the Facilitating Role of Automatized Lexical Phrases in Second Language Fluency Development. *Journal of Language and Learning, 2,* (1), 27-50.

Wood, D. (2007). Mandarin Chinese Speakers in a Study Abroad Context: Does Acquisition of Formulaic Sequences Facilitate Fluent Speech in English? *The East Asian Learner, 3* (2), 43-62.

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics, 12* (1), 39-57.

Wood, D. (2010). *Perspectives on Formulaic Language Acquisition and Communication*. London: Continuum.

Wood, D. (2012). *Formulaic Language and Second Language Speech Fluency : Background, Evidence and Classroom Applications*. London: Continuum International Publishing.

Wood, D. (2016). Willingness to communicate and second language speech fluency: An idiodynamic investigation. *System, 60,* 11-28.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21* (4), 463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Wray, A. (2013). Formulaic language. *Language Teaching, 46* (3), 316-334.

Wray, A. & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication, 20,* 1-28.

Wulff, S. (2008). *Rethinking Idiomaticity: A Usage-Based Approach*. London & New York, NY: Continuum.

Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24* (1), 1-27.

Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System 66,* 156-167.

APPENDIX A:
Target Formulaic Sequences Classified According to Functions

EXPRESSING PERSONAL OPINIONS

it seems to me
in my opinion
this is my personal view
what I think is
my point is
as far as I'm concerned
personally I think
that's my point

EXEMPLIFYING

as an example
let's say
to illustrate
an example of this is
a couple of examples of
is a good example of
here are some examples of
I'd like to mention
take for example

EMPHASIZING

this is important
more importantly
… is really important because
It's important to note that
but surely
this is crucial
the bottom line is
… is really interesting because
that's the bottomline

EXPLAINING FURTHER

what I mean when I say
what I mean by
what I really mean is
let me explain
by that I mean
what I'm trying to say is
I want to say

NOTING DEFINITIONS/REASONS

this is known as
the overwhelming majority
is the term for
that would be
and the reason is
for the following reasons

EXPRESSING CONCERNS AND SUGGESTIONS

there may be
I'm not sure if
I wonder
I'm sure that
I agree but
they should just
the best way
maybe we can

EXPRESSING IDEAS/EXPERIENCES

I'm saying that
in my experience
I realized that
I'd like to say
it became clear
I noticed that
I'll just point out that
I learnt that
I found out that

SEQUENCING IDEAS

last but not least
I want to focus on
this brings us to

for starters
to start with
let’s look at
let me tell you
let me talk about

CONNECTING IDEAS

in the same way
on the other hand
is caused by
as a result
in other words
that means
so it means

DRAWING ATTENTION

as you can see
we all know
you already know
I want to point out that
it was interesting that
might sound
is right because
there is definitely

APPENDIX B:
List of Academic Vocabulary Instructed in the Comparison Group

Week 1:

indicator
advocate
commitment
ultimately
annual
volume
integrity
devoted
discretion
refined
reliance
readjustment
accumulation
aggregation
alternatively
illustrative
analogous

Week 2:

overlapping
incorporation
parameters
output
prospective
brevity
commodities
incompatible
compounds
intensity
comprehensive
inconclusive
infrastructure
concurrently
non-conformist
monitoring

Week 3:

reconstruction
consumption
persistent
corresponding
convention
academia
accessibility
beneficiaries
distorted
enhancement
environmentalist
established
expertise
evidently
hierarchical
revenue

Week 4:

inconceivable
underlying
implication

institutionalize
intervention
consent
justification
maintenance
marginal
abnormal
occupied
precede
predominate
extraction
deviation
emergent

Week 5:
preliminary
disproportionate
attainable
conversely
authorship
pursue
qualitative
incentive
unanticipated
assembly
resourceful
scope
subsidies
survey
violation
implicitly
compilation

APPENDIX C:
Speaking Course Books Scanned for Formulaic Sequences

Beglar, D. & Murray, N. (2002). *Contemporary topics 3 : advanced listening and note-taking skills.* New York: Longman.

Blass, L. & Hartmann, P. (2007). *Quest 3: listening and speaking.* Boston: McGraw Hill.

Campbell, C. (2007). *English for academic study: vocabulary : course book*. South Street, UK: Garnet Education.

Dellar, H. & Walkley, A. (2006). *Innovations : a course in natural English : elementary workbook / Hugh Dellar and Andrew Walkley*. London : Thomson.

Maurer, J. (2005). *Focus on grammar 5 : an integrated skills approach*. White Plains, NY : Pearson Education.

Sarosy, P. & Sherak, K. (2007). *Lecture ready 2 : strategies for academic listening, note-taking, and discussion*. New York : Oxford University Press.

Tiberio, S. C. (2006). *Focus on grammar 3: an integrated skills approach : teacher's manual*. White Plains, NY : Pearson Education.

APPENDIX D:
Oral Narration Task (L1 Fluency Test)

APPENDIX E:
Picture Description (Pretest)

Dear participant,

In the first part of the interview, you will orally describe this picture. I will not lead you with questions once the recorder is turned on and I will not answer questions from that point on. You are expected to speak without interruption. Spend your time describing what you see in the picture (Who are these people? How many people are there? What are they doing? And why/why not? What are they wearing? Where are they?). It would be a good idea to assume that I have not seen it before so that you can describe it in detail to me.

You have one minute to examine the picture.

You have three minutes to describe it. You can use the chronometer.

APPENDIX F:
Oral Narration Test (Pretest)

The Comment Group: The hackers hunting for clues about you

If you had an email that looked like it was from your boss asking how your recent holiday went, would you open it? Most probably - and hackers know it.

One group in particular has used this simple technique to devastating effect, using it to spy on some of the world's biggest corporations. But who are they, and what are they looking for?

When security experts looked into some of the highest profile hacks in recent years - one particular criminal group kept on coming to their attention.

The Comment Group, which industry insiders say is based in China, offer hacking for hire - be it for individuals, corporations or governments.

But more recently, the Comment Group has become known for being particularly adept in one other important discipline of hacking: research.

"They find the weakest link in the company," explains Jaime Blasco, a security specialist.

"What they do is collect intelligence about the companies,"

Attack on different sides

It is an approach that has been devastatingly effective.

The group has been credited as being behind a vast range of attacks - everything from gaining access to user accounts at the EU to, according to BBC, targeting politicians.

In a document published by Wikileaks, the US government regarded the Comment Group as being one of the most serious of all hacking threats originating from China.

Soft drinks

One example which demonstrates the group's approach is that of Coca-Cola, which towards the end was revealed in media reports to have been the victim of a hack.

And not just any hack, it was a hack which industry experts said may have derailed an acquisition effort to the tune of $2.4bn (£1.5bn).

The US giant was looking into taking over China Huiyuan Juice Group, China's largest soft drinks company - but a hack, believed to be by the Comment Group, left Coca-Cola exposed.

How was it done? Bloomberg reported that one executive - deputy president of Coca-Cola's Pacific Group, Paul Etchells - opened an email he thought was from the company's chief executive.

In it, a link which when clicked downloaded malware onto Mr Etchells' machine. Once inside, hackers were able to snoop about the company's activity for over a month.

The Chinese government binned the acquisition soon after - citing competition concerns.

Coca-Cola has not officially commented on the hack. In a statement, the company told the BBC: "As a matter of practice, we do not comment on security matters."

But Alienvault's Mr Blasco explained how the attack was typical of the Comment Group's style.

"This Comment Group has been targeting a lot of companies that were in the process of being acquired or that a US company was trying to acquire in China," he said.

"I have seen that in dozens of industries. They are trying to gain access to financial information, and also they are compromising not only the companies but all the third parties, like lawyers, that are helping that company."

Highly organised

When you hear someone describing the Comment Group, it sounds like almost like any other firm, with groups of employees all assigned to different crucial bits of the business.

But rather than accounts, HR and sales - the Comment Group's components are designed to maximise efficiency in stealing information, Mr Blasco says.

"They have the guys working on exploits, you have the guys that are changing or programming the malware to gain access to the systems, and then you have the guys that are the operators.

"They don't know a lot about computers, what they do is operate the malware - they try to find the specific information, they collect intelligence from the victims and save that information for whatever purpose."

APPENDIX G:
Oral Argumentation Task (Pretest)

In this part of our interview, you will talk about YOUR IDEAS about the problems in the world.

Here is a list of problems that need attention as determined by world leaders in the year 2000 in a meeting organized by the United Nations:

- Children's health
- Mothers' health
- HIV/AIDS and other diseases
- Gender equality issues
- Extreme poverty and hunger
- Education
- Environmental sustainability
- Developing partnership between the world governments

Examine the list. Is there anything that you do not understand?

Now, will you tell me which one of these important problems needs the most immediate attention and which one needs the least immediate attention and WHY?

You have three minutes to explain your choice and reasoning.

APPENDIX H:
Picture Description (Posttest)

Dear participant,

In the first part of the interview, you will orally describe this picture. I will not lead you with questions once the recorder is turned on and I will not answer questions from that point on. You are expected to speak without interruption. Spend your time describing what you see in the picture (Who are these people? How many people are there? What are they doing? And why/why not? What are they wearing? Where are they?). It would be a good idea to assume that I have not seen it before so that you can describe it in detail to me.

You have one minute to examine the picture.

You have three minutes to describe it. You can use the chronometer.

APPENDIX İ:
Oral Narration Task (Posttest)

Only 38% of Americans Get Mental Health Care When They Need It, and For One Simple Reason

One of the biggest untreated problems in the United States affecting everything from social relationships to employment is mental health. Many do not receive the care they need, mostly for financial reasons. Luckily, things may improve in the next few years, as the Affordable Care Act (ACA) is implemented across America.

Around 25% of adults experience a mental health issue in a given year, yet less than 1 in 3 adults receives services. According to the report, around 50% of Americans will experience some mental health issues over their lifetimes. The rate of mental health issues in the U.S. is abnormally high, and lack of treatment options is only making this worse. The estimated impact in terms of loss of productivity in the workplace is around $63 billion. Only a small percentage of these diagnoses consist of severe problems such as schizophrenia, but the impact of other more treatable forms of mental disorders is undeniable.

The Substance Abuse and Mental Health Services Administration recently released the results from its 2011 national survey on mental health. One of the most disturbing results in this survey is that only 38% of individuals with mental health issues have received appropriate services.

A simple graphic on page 26 of the report outlines the reasons why individuals did not benefit from services. The single biggest reason for not receiving services is "Could Not Afford the Cost." When cost is combined with responses around under-insurance, over 65% cited money-related issues as the primary reason for not receiving treatment.

Lack of treatment impacts more than just productivity. Many untreated mental health issues lead to an increased likelihood of substance abuse, child abuse, and other domestic problems. The financial ripple effect is much greater than the loss of individual productivity, resulting in more services being consumed in other programs.

The simple question is, what can be done?

For a very long time, insurance companies have skimped or omitted coverage for mental

health services. Most plans pay 50% of the cost of services at best. Given the large number of uninsured Americans, it is easy to see why so many people cite money as the reason they have not received services.

Culturally, the U.S. has a common stigma against people who benefit from mental health services. Many people associate mental health services with neurotics like Woody Allen who have spent decades in "analysis" with no signs of improvement. The simple fact is that many people experience mental health issues during periods of great stress such as illness, financial worries, job loss, etc. Treatment is often an educational process, helping people to learn coping strategies to reduce their internal stress levels and to respond more appropriately to life's challenges.

Although the ACA has been arguably the most divisive issue in the Obama presidency, it does offer a ray of hope for improving mental health services in the United States. On January 1, 2014, all qualified health plans under the ACA will be required to cover mental health services. Not only does the ACA require coverage, it expands services to where they often needed but currently ignored.

As David Mechanic explains in a 2012 article in the journal *Health Affairs*:

"It promotes new programs and tools, such as health homes, interdisciplinary care teams, the broadening of the Medicaid Home and Community-Based Services option, co-location of physical health and behavioral services, and collaborative care."

While we still have a long way to go to address the unmet needs to mental health services in the U.S., the ACA should put us on track for lasting and meaningful change.

APPENDIX J:
Oral Argumentation Task (Posttest)

In this part of our talk, you will talk about YOUR IDEAS about the problems in Turkey.

Here is a list of problems that need attention as determined by world leaders in the year 2000 in a meeting organized by the United Nations:

- Children's health
- Mothers' health
- HIV/AIDS and other diseases
- Gender equality issues
- Extreme poverty and hunger
- Education
- Environmental sustainability
- Developing partnership between the world governments

Examine the list. Is there anything that you do not understand?

Now, will you tell me which one of these important problems needs the most immediate attention and which one needs the least immediate attention in Turkey and WHY?

You have three minutes to explain your choice and reasoning.

APPENDIX K:
Picture Description (delayed posttest)

Dear participant,

In the first part of the interview, you will orally describe this picture. I will not lead you with questions once the recorder is turned on and I will not answer questions from that point on. You are expected to speak without interruption. Spend your time describing what you see in the picture (Who are these people? How many people are there? What are they doing? And why/why not? What are they wearing? Where are they?). It would be a good idea to assume that I have not seen it before so that you can describe it in detail to me.

You have one minute to examine the picture.

You have three minutes to describe it. You can use the chronometer.

APPENDIX L:
Oral Narration Task (Delayed Posttest)

Switzerland guns: Living with firearms the Swiss way

Switzerland has one of the highest rates of gun ownership in the world, but little gun-related street crime - so some opponents of gun control welcome it as a place where firearms play a positive role in society. However, Swiss gun culture is unique, and guns are more tightly regulated than many assume.

Last month, in the French-speaking village of Daillon, 100km (62 miles) from Geneva, a psychologically disturbed man opened fire on locals, killing three people and wounding two others. Police had already got his weapons from the gunman in 2005, after he had been placed in psychiatric care.

Inevitably, his actions prompted a fresh wave of debate in Switzerland about its relatively liberal gun laws.

According to a Geneva-based research, there are about 89 civilian-owned guns for every 100 people who live in the United States. Switzerland ranks third in terms of gun ownership with 3.4 million guns among its population of nearly eight million.

Target shooting is a popular national sport but many of the firearms in Switzerland are military weapons.

All healthy Swiss men aged between 18 and 34 have to do military service and all are issued with rifles or guns which they are supposed to keep at home.

Twenty years ago the Swiss military was a force of around 600,000 soldiers. Today it is only a third of that size but until recently most former soldiers used to keep their guns after they had completed their military duties, leading to lots of weapons being stored in the cupboards of private Swiss households.

In 2006, the champion Swiss skier Corrinne Rey-Bellet and her brother were

murdered by Corinne's ex-husband, who shot them with his old rifle before killing himself.

Since that incident, gun laws concerning army weapons have tightened. Although it is still possible for a former soldier to buy his firearm after he finishes military service, he must provide a justification for keeping the weapon and apply for a permit.

In America then, gun ownership is about self-defence whereas in Switzerland it is seen more in terms of national security. To many traditionalists, a gun in the home has become a metaphor for an independent, well-developed Switzerland which has helped to keep the country out of two world wars.

Hermann Suter, vice-president of the Swiss lobbying group Pro Tell, is angry because of calls that the Swiss military should give up their guns and store them in a central store.

"It is a question of trust between the state and the citizen. The citizen is not just a citizen, he is also a soldier, " he reminds me. "The gun at home is the best way to avoid dictatorships - only dictators take arms away from the citizens."

Yet despite the possession of firearms, violent gun-related street crime is extremely rare in Switzerland.

In an average year here, there is one gun murder for every 200,000 of the population - in the US that figure is several times higher. But there are more domestic homicides and suicides with a firearm in Switzerland than pretty much anywhere else in Europe.

The army is not the only entity to have a tradition with guns however. About 600,000 Swiss - many of them children - belong to shooting clubs.

Swiss citizens - for example hunters, or those who shoot as a sport - can get a permit to buy guns, unless they have a criminal record. But their number is still so low compared to the US.

APPENDIX M:
Oral Argumentation Task (Delayed Posttest)

In this part of our talk, you will talk about YOUR IDEAS about these statements:

- "Education is the most powerful weapon we can use to change the world" Nelson Mandela
- "I believe that if your aim is to change the world, journalism is a more immediate short-term weapon" Tom Stoppard
- "The human race has one really effective weapon, and that is laughter" Mark Twain

You can spend two minutes to think about these statements.

As you can see several influential figures suggested different weapons can be used to create change in the world, such as education, journalism and laughter/optimism. Now, will you tell me with which one of these weapons you would rank as the most effective one and which one you would rank as the least effective one in creating a change in the world and WHY?

You will spend three minutes to explain your opinion.

APPENDIX N:
Informed Consent Form

Araştırmaya Dair Bilgilendirme Formu

Yeditepe Üniversitesi İngiliz Dili Eğitimi Doktora Programı'nda devam etmekte olan bir tez araştırmasına katılmaya gönüllü oldunuz. Aşağıda araştırmanın amacı ve koşulları hakkında bilgi verilecek ve bunları takiben katılımınızı onayladığınıza dair imza atmanız istenecektir.

ARAŞTIRMACI: Araştırmacının iletişim adresi: Ayşegül Nergis, aysegulnergis@gmail.com

AMAÇ: Araştırmanın amacı, Yaz dönemine dek katılımcılara bildirilmeyecektir. Üniversite yönetimi, araştırmanın öğrencilerin ENG 102 dersinde alacakları eğitimi ve performanslarını olumsuz yönde etkilemeyeceğini ve araştırma etiğine uygun olduğunu onaylamıştır.

KATILIM KOŞULLARI: Araştırmaya katılmak zorunlu olmayıp tamamen kişinin seçimine bağlıdır. Katılımdan dolayı ders içi veya dışı bir ödüllendirme yapılmayacaktır. Araştırma sonuçları yabancı dil eğitimi veren programların geliştirilmesi konusunda bilimsel veri elde edilmesini sağlayacaktır. Araştırmacı, sizinle dönem başında ve sonunda, uygun bulduğunuz zaman dilimi içerisinde birer görüşme yapacaktır. İlk görüşmenin 30 dakika, ikinci görüşmenin ise 15 dakika süreceği öngörülmektedir.

OLASI RİSKLER: Araştırma süresince herhangi bir risk oluşabileceği düşünülmemektedir. Ancak, katılımcılar araştırmaya katılımlarının kendilerini herhangi bir sebepten dolayı rahatsız ettiğini düşünürlerse sebep bildirmeksizin son görüşmeye gelmek zorunda değildirler.

GİZLİLİK İLKESİ: Toplanan tüm veriler gizli tutulacak, üçüncü bir şahıs tarafından paylaşılmayacaktır. Katılımcılara dair kişisel bilgiler araştırma raporuna yansıtılmayacaktır. Ancak, araştırmanın bütününden elde edilen bulgular bilimsel toplantı ve eserlerde paylaşılabilir.

GERİ ÇEKİLME HAKKI: Katılımcılar, hiçbir sebep bildirmeksizin araştırmaya katılmaktan vazgeçebilir. Bunun katılımcıya yansıtılacak herhangi bir sonucu yoktur.

Yukarıda tanımlanan araştırma koşullarını okudum ve araştırmaya katılmaya gönüllü oldum.

Katılımcının adı: Katılımcının imzası:

Tarih:

Araştırmacının adı: Araştırmacının imzası:

Tarih:

APPENDIX O:
Background Questionnaire

Adınız:

Yaşınız:

Cinsiyetiniz:

Mezun olduğunuz lise:

Üniversitedeki Programınız:

Hazırlıkta okuduğunuz sömestr sayısı:

Herhangi bir işitme ya da konuşma probleminiz var mı:

Eğer yurtdışında okuma deneyiminiz olduysa süresi ve yerini yazınız:

Kaç senedir İngilizce eğitimi görmekte olduğunuzu yazınız:

Bildiğiniz diğer yabancı dilleri ve seviyelerini yazınız:

(Bu bilgiler gizli kalacaktır)

Teşekkürler

APPENDIX P:
Sample Lesson Plan

Lesson Plan For Experimental Group (Week 3)

Materials: PPT, handouts of formulaic sequences and concordances, end-of-lesson tests
Core reading passage: *Corporate Social Responsibility: What It really Is, Why It's So Important, and How It Should Be Managed (retrieved from* http://www.ecrc.org.eg)

FIRST LESSON

TIME	PROCEDURES	TARGETED FORMULAIC SEQUENCES
0-5th minutes	Warm-up Teacher and students exchange greetings. Teacher takes attendance. Teacher asks whether the students found the article of the week interesting. She asks what they found the most interesting about the article.	EXPRESSING PERSONAL OPINIONS it seems to me in my opinion this is my personal view what I think is my point is as far as I'm concerned personally I think that's my point
5-8th minutes	Transition questions Teacher asks questions to the class about their shopping habits: what kind of things do they shop for? What are the main factors that affect their shopping preferences? Which companies do they usually buy from? What are the factors that affect the prestige of these companies?	
8-12th minutes	Transition questions The teacher shows a slide that contains pictures of the BODY SHOP logo, its shops and its spokesperson. She asks the class whether they knew about this company before reading the	

	passage, what they learnt about the company after reading it, who this lady is, what they know about her, what the chapter tells about CSR behaviors of the BODY SHOP.	
12-15th minutes	*Presentation*: Teacher projects the list of targeted formulaic sequences on the screen. She explains that these formulaic sequences can be used to express personal opinions. Students read them aloud one by one repeating after the teacher.	
15-23rd minutes	*Practice* Then she shows another slide containing concordances of the formulaic sequences to show in which context they are used. She reads them aloud one by one. Then she gives the students a worksheet that includes these formulaic sequences and concordances. She tells them to examine the concordances so that students can comprehend how each of these items is used in real world context. They read them to each other in groups. After that they read the FS aloud one by one repeating after the teacher.	
23-35th minutes	*Production* Question and answer/discussion: After a few minutes, she tells them to choose 4 of the formulaic sequences and state their opinions on the following questions by using these formulaic sequences. The questions are:	

	- Are you for or against animal testing for producing beauty products? - Do you think animal testing is the only option? - Should people go on buying products of companies like LOREAL that still uses animal testing?	
35-50th minutes	*Production* Group discussion/oral report: Teacher initiates a pair/group discussion activity. Students are presented she tells them to choose 4 of them and state their opinions on the following questions by using these formulaic sequences. The following questions: - Do you think companies do a good job of reporting about the test methods they use? - Can we trust the word of companies? - Who can guarantee their transparency? What can these institutions/decision makers do?	

SECOND LESSON

0-5th minutes	*Warm-up questions* Teacher asks students about their opinions on the rest of the article. They discuss the most interesting points and data given in the article.	
5-10th minutes	*Presentation* Teacher shows students another slide that contains the targeted formulaic sequences. Teacher explains that they can be used to express concerns and suggestions Students read them aloud one by one repeating after the teacher. Then the teacher shows a slide containing the concordances. Students examine the concordances given in the slides. Teacher reads the concordances aloud.	EXPRESSING CONCERNS AND SUGGESTIONS there may be I'm not sure if I wonder I'm sure that I agree but They should just The best way Maybe we can
10-20th minutes	*Practice* She gives them a sheet of paper that contains the formulaic sequences and concordances so that they can comprehend in which contexts these formulaic sequences are used in real life. They examine them in groups. Then teacher goes back to the previous slide. Then they read the list of FS aloud one by one repeating after the teacher.	
20-35th minutes	*Production* Research task: Teacher initiates a research task. In	

	groups of 3-4, students do research on the spokesperson of the Body Shop, Anita Riddick and a controversial event she recently experienced. "Does this event mean that CSR can be purchased?" Students must use at least 4 of the formulaic sequences given in the work sheets while reporting their conclusion to the class.	
35-50th minutes	*Production* Group discussion and presentation: The teacher assigns students into random groups of three or four. She assigns a case study report of LBG Associates to each group and delivers copies of case studies to each group member. She instructs them to read them carefully so that they can summarize the important points of the report for the whole class and present their opinions on what each of these programs aimed to achieve. The students discuss in groups the value/merit of each program: Were they useful? Which one of these programs were more valuable in terms of solving a problem in society/environment? The students prepare a 3-minute presentation that will be presented to the whole class in which they discuss their opinions on what the important facts/information about the report are. She reminds them to use at least 4 of the formulaic sequences in their	

	worksheets in their presentations.	

APPENDIX P.2.

SAMPLE LESSON PLAN

Lesson Plan For Comparison Group (Week 3)

Materials: PPT, handouts of academic vocabulary items and concordances, end-of-lesson tests
Core reading passage: *Corporate Social Responsibility: What It really Is, Why It's So Important, and How It Should Be Managed (retrieved from* http://www.ecrc.org.eg)

FIRST LESSON

TIME	PROCEDURES	TARGETED ACADEMIC VOCABULARY
0-5th minutes	Warm-up Teacher and students exchange greetings. Teacher takes attendance. Teacher asks whether the students found the article of the week interesting. She asks what they found the most interesting about the article.	indicator advocate commitment ultimately annual volume integrity devoted
5-8th minutes	Transition questions Teacher asks questions to the class about their shopping habits: what kind of things do they shop for? What are the main factors that affect their shopping preferences? Which companies do they usually buy from? What are the factors that affect the prestige of these companies?	
8-12th minutes	Transition questions The teacher shows a slide that contains pictures of the BODY SHOP logo, its shops and its spokesperson. She asks the class whether they knew about this company before reading the passage, what they learnt about the	

	company after reading it, who this lady is, what they know about her, what the chapter tells about CSR behaviors of the BODY SHOP.	
12-15th minutes	*Presentation*: The teacher projects the list of academic vocabulary in the reading text on the screen. She reads them aloud one by one to demonstrate how they are pronounced. Students read them aloud one by one repeating after the teacher.	
15-23rd minutes	Then she shows another slide containing concordances of the vocabulary items to show in which context they are used. She reads the concordances aloud by emphasizing the target word. Then she gives the students a worksheet that includes these vocabulary items and concordances. She tells them to examine the concordances so that students can comprehend how each of these items is used in real world context. In groups they read them aloud to each other. After that students read them aloud one by one repeating after the teacher.	
23-35th minutes	*Production* Question and answer/discussion: After a few minutes, she tells them to choose 4 of them and state their opinions on the following questions by using these academic words. The questions are: - Are you for or against animal testing for producing beauty products? - Do you think animal testing is the only option? - Should people go on buying products of companies like LOREAL that still uses animal testing? - Do you think companies do a	

	good job of reporting about the test methods they use? - Can we trust the word of companies? - Who can guarantee their transparency? What can these institutions/decision makers do?	
35-50th minutes	*Practice* Group discussion/oral report: Teacher initiates a pair/group discussion activity. She tells them to choose 4 of the target items that they did not use and state their opinions on the following questions by using these vocabulary items. The following questions: - Are you for or against animal testing for producing beauty products? - Do you think animal testing is the only option? - Should people go on buying products of companies like LOREAL that still uses animal testing? - Do you think companies do a good job of reporting about the test methods they use? - Can we trust the word of companies? - Who can guarantee their transparency? What can these institutions/decision makers do?	

SECOND LESSON

0-5th minutes	*Warm-up questions* Teacher asks students about their opinions on the rest of the article. They discuss the most interesting points and data given in the article.	
5-10th minutes	*Presentation* Teacher then returns back to the core reading of the week. The teacher shows another slide that contains the targeted vocabulary items. She reads them aloud one by one. Students read them aloud one by one repeating after the teacher. Then she projects the concordances. The students examine the concordances given in the slides.	discretion refined reliance pursue accumulation aggregation alternatively illustrative analogous
10-20th minutes	*Practice* She gives them a sheet of paper that contains the vocabulary and concordances so that they can comprehend in which contexts these words are used in real life. She reads the concordances aloud by emphasizing the target word. In groups students also read them aloud to each other. Then they read the academic words aloud one by one repeating after the teacher.	
20-35th minutes	*Production* Research task: Teacher initiates a research task. In groups of 3-4, students do research on the spokesperson of the Body Shop, Anita Riddick and a controversial event she recently experienced. “Does this event mean	

	that CSR can be purchased?" Students must use 5 of 4 of the academic words while reporting their conclusion to the class.	
35-50th minutes	*Practice* Group discussion and presentation: The teacher assigns students into random groups of three or four. She assigns a case study report of LBG Associates to each group and delivers copies of case studies to each group member. She instructs them to read them carefully so that they can summarize the important points of the report for the whole class and present their opinions on what each of these programs aimed to achieve. The students discuss in groups the value/merit of each program: Were they useful? Which one of these programs were more valuable in terms of solving a problem in the society/environment? The students prepare a 3-minute presentation that will be presented to the whole class in which they discuss their opinions on what the important facts/information about the report are. She reminds them to use at least 4 of the academic words in their worksheets in their presentations.	

APPENDIX Q:
Sample Handout and Worksheet

EXPERIMENTAL GROUP

Handout 3

Below are some useful phrases that you can use to express your personal response. Examine them in context and use them in your answers.

It seems to me
In my opinion
What I think is
My point is
Personally I think
That's my point

The following are some examples of how these phrases are used in real-life university classrooms.

it's the same reason that, i don't have a a content analysis on this so my number could be wrong but	**it seems to me**	that about eight out of ten movies that come out are centered in Los Angeles. now you could say
for limited health care dollars you can't ignore that. the economic competition is basically	**in my opinion**	is what is driving this attack on nurse practitioner practice. basically, organized medicine and
and you're trapped in between two worlds. one of them is not just the postmodernist world, but	**what i think is**	the myth of the modern world which, i think he falls for, to a certain extent.
he doesn't think you have a God. i opt for the atheist Hume, in all of this. but	**my point is**	simply to situate him, in this complicated game of defending
was working from was a seventy thirty split. that seventy percent was spent on law enforcement, and	**personally i think**	that's a great split, personally, as a member of jail.
that was quite different, right? It may change.	**that's my point**	it might change, I say.

APPENDIX Q.2.

SAMPLE WORKSHEET FOR EXPERIMENTAL GROUP

Fill in the blanks with the most appropriate sequences.

What I think is my point is in my opinion it seems to me personally I think that's my point

1. _____ quite different from authorities in the field. And it really makes a difference in how you look at what this second novel is about.
2. Album sales, like ticket sales for movies, are not, _____, something that it has high musical quality.
3. The local governors had criticized our actions but they were mostly jealous because we were successful. At least _____.
4. _____ there are a lot of mistakes in the paper. It lacks the visual data and graphs.
5. That movie was rather weak, but _____ overall it was fine because I could fine well-hidden social facts.
6. What looks evident to you may change. But _____. You know… the Earth revolves around the Sun. That's directly observable.

APPENDIX Q.3.

SAMPLE HANDOUT FOR COMPARISON GROUP

Handout 3

Below are words that were cited in the texts of this week. Examine them in context and use them in your answers.

refined
reliance
pursue
accumulation
visibility
alternatively
analogous

The following are some examples of how these words are used in real-life university classrooms.

but they in a sense become, a kind of ornament if you will, although extremely subtle, extremely	**refined**	, extremely spare. so in contrast with this idea of, high modernism something changes, in the sixties
plants, is the fact, there's a lot of, advantages we'll talk about, but, it may help us decrease our	**reliance**	on oil-derived chemicals. so lots of chemicals that're, additives in foods or, additives to plants
consumption in terms of, constraints given in the environment, that you create certain houses, you	**pursue**	certain economic strategies, because you are limited, given your environment, on what those strategies
if it's low oxygen, it doesn't matter if it's high carbon dioxide, it doesn't matter if it's an	**accumulation**	of waste products, it just simply doesn't matter, the metabolic response of the animal is identical,
well what we found in our assessment our program opportunities is that, we have more visibility. more	**visibility**	could strengthen, the partner, relationship and program attendance that's the opportunities they can
it differently. Kelly's phrase was what he called constructive alternativism. that you could	**alternatively**	construct the world and your experience, differently from moment to moment. it seems to me that
great freedom in building up and modifying the form. modelling is a lot like it is really in some ways	**analogous**	to oil painting. you have a chance, to fuss around with it you can change it. you can, you know

APPENDIX Q.4.

SAMPLE WORKSHEET FOR COMPARISON GROUP

Fill in the blanks with the most words.

refined reliance pursue accumulation visibility alternatively analogous

1. There is no guarantee that you will succeed in this test, although it is _____ to the previous one.

2. Good promotion can help _____ of the Project.

3. Today we are thankful for the _____ of technological evolution.

4. It has become almost impossible to find _____ resources.

5. As a species, our _____ on animals as a source of nutrition may lead to our end.

6. You can volunteer in an established organization, or if you are rich enough you can _____ establish for own organization.

7. Who would like _____ a career where there are limited promotion opportunities?

APPENDIX R

Oral Data Elicitation Methods Used in Previous Studies

Study	Tasks for measurement of oral fluency, accuracy, complexity and proficiency	Calculations	Level
Yuan & Ellis, 2003	• Picture description	*Fluency* Rate A: N of syllables per minute Rate B: N of meaningful syllables per minute *Complexity* Syntactic complexity: the ratio of T-clauses to T-units Syntactic variety: N of different grammatical verb forms Mean Segmental Type-Token Ratio: Narratives were divided into 40 word segments. The N of different words divided by the total N of words in the segment. *Accuracy* Error-free clauses: The percentage of clauses that did not contain any error Correct verb forms: The percentage of accurately used verb forms	Advanced (undergraduate students)
Boers et al., 2006	• Oral Proficiency Interview (parameters of fluency, accuracy, range of expression [lexical richness and syntactic complexity]) 1) conversation on a short article (7 minutes) 2) spontaneous conversation (7	Rater judgment on a scale of 0-20.	Estimated to be upper-intermediate to advanced (university students)

	minutes)		
Derwing et al., 2004	• Picture narration • 2-minute monologue • conversation (participants directed questions of the researcher about the researcher's happiest moment)	Temporal measures: pruned syllables per second Rater judgment on a 9-point scale	High beginner speakers (university students)
Stengers et al., 2011	• re-tell task (600 words. 3 minutes of preparation time)	*Fluency* : tokens-per-minute measure. *Range* : type counts were used (as a reflection of the diversity or width of students' repertoires of formulaic sequences). *Accuracy*: type counts	Advanced speakers (university students)
Ahmadian & Tavakoli, 2011	Oral narrative task (on a 15-minute video)	*Accuracy*: Error-free clauses and correct verb forms *Complexity*: Syntactic complexity (amount of subordination and syntactic variety (the total number of different grammatical verb forms used in participants performances) *Fluency*: number of syllables produced per minute of speech and number of meaningful syllables produced per minute of speech	Intermediate level young adults
De Jong et al., 2013	• 8 monologue speaking tasks: role play/picture description)	*Utterance fluency*: Number of silent pauses Mean duration of silent pause Number of Filled pauses, corrections, repetitions Mean duration of syllables	Intermediate learners (adult learners)