University of Reading

An investigation into cognitive and linguistic variables that influence the learning and production of formulaic sequences by undergraduate students via a speaking task in an English-as-a-second-language context.

Khuloud Alali

Thesis submitted for the degree of Doctor of Philosophy in Applied Linguistics

Department of English Language and Applied Linguistics

School of Literature and Languages

University of Reading

January 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Khuloud Alali

Acknowledgement

First and foremost, all praise goes to Allah The All Provider for blessing me with completing this journey of knowledge and for always being there for me.

I would like to dedicate this thesis to my supervisor, Professor Jeanine Treffers-Daller, who not only provided great support and guidance throughout the writing process but also offered invaluable emotional support during times when I needed it the most. Jeanine, I want to express my heartfelt gratitude for your constant support, belief in me, and for instilling even more belief in myself. I would also like to express my deep gratitude to my second supervisor, Professor Parvaneh Tavakoli. Words cannot fully express how exceptional she is.

This thesis would not have been possible without the support of my beloved country, Kuwait, and the collaboration of colleagues who facilitated access to participants. Of course, my heartfelt gratitude goes to all the participants who willingly took part in this study.

I am deeply thankful to my father, who is the main motivation behind my pursuit of my academic studies, first for encouraging me to complete my master's degree and then to take on a PhD. His constant words of encouragement and unwavering determination have always echoed in my ears. Thank you, Dad, for imparting in me the strength to aim for the stars and pursue my dream. My mother has been incredibly supportive throughout my PhD journey. She has taken care of my two children, whom I welcomed into the world during my PhD journey. My beautiful daughter, Lulwa, and my charming son, Abdulrazzaq, you are my main source of strength, and I am grateful every day for having you in my life. Thank you for being mine.

Finally, I would like to thank the love of my life, my best friend, my mentor and my husband, who is also about to submit his PhD thesis, for everything. Mohammad, you have been by my side through thick and thin, and I could not have imagined navigating this journey without you. I am eagerly looking forward to embarking on more adventures with you. How about you?

Table of Contents

List of tables and figures

List of Abbreviations

A: Accuracy

A-SA: Accuracy of speech

AcaVoc: Academic vocabulary

ACCESS: Automatisation in Communicative Contexts of Essential Speech Segments

AS units: Analysis of Speech units

AWL: Academic Word List

BNC: British National Corpus

C: Complexity

C-SA: Complexity of speech

CAF: Complexity, Accuracy and Fluency

CBS: College of Business Studies

CEFR: Common European Framework of Reference for Languages

CLT: Communicative Language Teaching

COCA: Corpus of Contemporary American English

DSF: Digit Span Forwards

DSB: Digit Span Backwards

DSS: Digit Span Sequencing

DVs: Dependent variables

EFL: English as a foreign language

ESP: English for specific purposes

F: Fluency

F-SA: Fluency of speech

FS: Formulaic sequence

FSs: Formulaic sequences

H: Hypothesis

IELTS: International English Language Testing System

IVs: Independent variables

L1: First language

L2: Second language

LDS: Longest Digit Span

LHQ: Language History Questionnaire

MI: Mutual Information

NVLT: New Vocabulary Levels Test

SLA: Second language acquisition

PAAET: Public Authority for Applied Education and Training

PPP: Present-practice-produce

PVST: Phrasal Vocabulary Size Test

QPT: the Quick Placement Test

RQ: Research Question

RQs: Research Questions

SLA: Second-language acquisition

SRep: Sentence repetition task

TAALES: Tool for the Automatic Analysis of Lexical Sophistication

TF-FSs: Test of familiarity of FSs

VIF: Variance Inflation Factor

VST: Vocabulary Size Test

WCR: Weighted Clause Ratio

WM: Working memory

WMC: Working memory capacity

Abstract

A key problem for second language (L2) learners in many contexts is to improve the fluency of their speech (Tavakoli & Wright, 2020). As fluent speakers often make use of fixed expressions – i.e. formulaic sequences (FS), improving our understanding of these expressions will help us understand how they impact language ability in general and fluency in particular. Although many studies focus on the complexity, accuracy and fluency (CAF) of L2 learners' output, there are still few in-depth studies that examine the CAF of the FSs themselves as used by L2 learners. Therefore, this study aimed to address this gap by investigating the correlation between various linguistic and cognitive variables and (a) the CAF of entire speech and (b) the CAF of FSs elicited via a speech sample.

Participants in the study were adult L2 learners of English in Kuwait. The sample under study (N = 51) were mostly at A1 to A2 level according to the CEFR, with some participants at B1 to C1 levels. Participants carried out a monologic speaking task revolving around the topic of giving self-introduction in a work-life context. Participants also carried out the following tasks: a general proficiency test (the Quick Placement Test), a vocabulary test (New Vocabulary Levels Test), a FSs test (the Phrasal Vocabulary Size Test), a test of familiarity of FSs (operationalised as expressions learned from the participants' textbook) and working memory tasks (Wechsler Adult Intelligence Scale IV).

This study proposes a novel complexity index of FSs, derived from TAALES indices (Kyle et al., 2018), and monolinguals' judgements of the transparency of each FS and their complexity. The analysis of fluency of FSs is based on a study of the pauses before and after each FSs, while accuracy of FSs examination focuses on errors and error-free chunks within FSs.

Results from hierarchical regression analysis indicate that working memory was significantly associated with the complexity of overall speech, aligning with recent studies (Awwad & Tavakoli, 2022). This suggests that learners with higher working memory capacity (WMC) can store and retain more complex speech compared to those with lower WMC. However, no significant associations were found between learners' working memory and CAF of FSs, suggesting that for learners with lower working memory, the use of FSs serves as a compensatory strategy to overcome the challenges linked with limited WMC. These findings highlight the strategic role of FSs in enhancing speech production, addressing complexities in language use, accuracy and fluency.

Interestingly, scores from the vocabulary test showed a less significant correlation with both the CAF of speech and CAF of FSs elicited via speech samples. This finding suggests several implications, including that low-proficiency learners may compensate for limited vocabulary by relying more on FSs, given their prior exposure to these sequences Similarly, general proficiency test scores showed a moderately positive correlation with CAF-FSs. This highlights the importance of general proficiency in language learning, although with a potentially less pronounced impact on the utilisation and integration of FSs into spoken language. These findings are important as they highlight the importance of concentrating on FSs in language learning, especially for this groups of participants – beginners.

Furthermore, while the study shows that scores from the test of familiarity of FSs significantly predict CAF of FSs scores derived from speech samples, the FS test derived from literature containing unfamiliar FSs had a weaker association with both CAF of entire speech and CAF of FSs in speech. This underscores the importance of prior exposure and familiarity with domain knowledge. It is argued that it may also be the case that the FS test derived from the literature was too advanced to accurately capture beginner learners' knowledge or to establish an association, considering that the participants were mostly lower-proficiency learners, even though this test was not restricted to a specific language proficiency level; instead, it was advocated for use by all L2 learners. These findings have several significant implications for L2 research, educators and those interested in pedagogy.

An important avenue for future research involves replicating this study with learners at higher proficiency levels to explore whether similar individual difference variables associate with CAF of speech and subsequently CAF of FSs. Additionally, investigating whether scores from test of familiarity of FSs maintain their explanatory power for CAF of FSs, compared to more general FS tests, would provide insights into the dynamics of language production across proficiency levels.

Chapter 1. Introduction

1.1 Rationale and focus of the study

Fluency of speech is a key issue for second language (L2) learners (Tavakoli & Wright, 2020). Many L2 learners struggle to interact in L2, even after having learned it for a long time. Ortega (2005, p. 79) mentioned the ‘awkwardness’ of learners in producing L2 speech in class due to the limited attention given to it by teachers. This ongoing issue is encountered by many L2 learners around the world, amongst them Kuwaiti ESL learners. From my experience as a learner, a secondary school teacher and then a language educator at college level, I have recognised teachers’ belief in the importance of implementing oral skills in the classroom; however, in practice, more time and focus is allocated to other domains (grammar, reading, writing and defining a vocabulary list). This may be attributed to teachers facing challenges in how to teach oral English, especially as learners often have limited vocabulary, grammatical and L2 communicative knowledge and the amount of available classroom time is limited. Cultural issues may be another reason for difficulties in speaking as learners would not want to lose face in front of peers, thus students can be fearful of criticism or simply shy. Although assessment is not the focus of this study, it cannot be denied that speaking is generally not assessed – as in quizzes and exams in the other domains mentioned above thus, classroom time is used to cover the given curriculum in order to meet exam requirements.

The use of traditional language learning approaches where a distinction is maintained between grammar and vocabulary is a typical example of teaching in various L2 contexts. Such a methodology may enable teachers to cover the syllabus in good time, but it places less emphasis on L2 interaction. The focus on grammar and compiling lists of vocabulary in L2 teaching is not facilitating effective communication for learners. Perhaps we have to look for a different avenue to help students to interact and engage in speaking practice, such as introducing fixed expressions – i.e., formulaic sequences (FSs) – to L2 curricula. As fluent speakers often make use of FS, improving our understanding of these expressions will help us understand how they impact language ability in general and fluency in particular and the analysis of which can help learners.

Speakers who use FSs seem to require less time to plan what they want to say (e.g., Christiansen & Chater, 2016; Khodadady & Shamsaee, 2012; Wray, 2002). FSs function as short-cuts for both the acquisition of target language aspects and processing of grammar and vocabulary (Wray, 2002) and make learners more confident to speak as formulating novel utterances is extremely demanding (Hall, 2010; Wray, 2002). Compared to the extensive research on FSs in written production (see Boers & Lindstromberg, 2012; Pellicer-Sánchez & Boers, 2019 for reviews), only a limited number of studies have explored the influence of FSs on both the learning and usage of FSs in spoken language. This study is motivated by the need to address this gap, specifically addressing beginner Kuwaiti learners of English who need to fulfil their needs— using the L2 for their future career. There is very little evidence from the Kuwait context on L2 learners' knowledge and use of FSs and how this relates to their fluency in English.

Therefore, the main aim of this study is to investigate to what extent there is a correlation and association between Kuwaiti students' oral fluency and their knowledge and use of FSs. To my knowledge, only a few studies are based on a correlational design that looks into the relationship between different variables and FSs produced via a speaking task. Khodadady and Shamsaee (2012) undertook one of the few studies into the relationship between the frequency of use of FSs and the learners' listening and speaking scores. However, Khodadady and Shamsaee's (2012) study involved correlating one variable–the frequency of FSs use–to another, the learners' level of speaking ability. A greater variety of variables, including general proficiency, vocabulary knowledge, knowledge of FS, working memory (WM) and variables from learners' language learning history, have not been explored to a great extent in a single study to determine whether, and, if so, to what extent, there is a correlation among these variables and the complexity, accuracy and fluency (CAF) of entire speech.

Secondly, reviewing previous research reveals a scarcity of studies utilising tasks that tests the familiarity of students' language domain, particularly tasks that are relevant to learners' future needs, in a scaffolded and supported context to enable them to perform at or near their maximum potential. Thirdly, a beginner learner level of proficiency is another factor that has not been examined with regards to variables that correlate with their production of FSs in speaking.

This study is further motivated by the need to investigate individual difference variables, specifically the role of WM, which has received attention in relation to the CAF of speech. Previous research has highlighted the impact of WM on both L2 development and L2 performance (Awaad & Tavakoli, 2022; Baddeley, 2015; Mitchell et al., 2013; Skehan, 2015). While there is existing research on the relationship between WM and the CAF of speech, there has been a limited focus on lower proficiency learners in the literature, making this a noteworthy gap in research.

Since this study focuses to examine two different types of CAFs—one for entire speech samples and the other for FSs—it is worth noting that, although many studies focus into CAF of L2 learners' output, there are still few in-depth studies that examine the CAF of the FSs themselves as used by L2 learners. Furthermore, among these measures, the complexity of FSs is relatively unexplored. To my knowledge, there is no research paper to date that have operationalised or measured the complexity of FSs. This study is motivated to fill this gap. The rationale behind investigating the measure of the complexity of FSs is twofold. Firstly, it allows us to gain insights into how the complexity of FSs can enhance the complexity of entire speech samples. Secondly, measuring the complexity of these sequences allows for a deeper understanding of a learner's proficiency by. Through this exploration, I hope to tap into the multidimensional aspect of complexity, operationalise it and develop a framework that could benefit other researchers in assessing the complexity of FSs.

To reiterate, the relationship between WM and CAF of speech has garnered attention from researchers. However, exploring the relationship between WM and the CAF of FSs elicited from a live performance of the task has been absent. This gap in research represents an important aspect of language production, especially for learners at lower proficiency levels, constituting an important gap in the literature.

Thus, this study aims to firstly address an important gap in the literature by identifying the best variable that can help beginner undergraduate college L2 learners to improve their speaking via the use of FSs. Specific attention will also be drawn to one of the cognitive factors that can have significant effects on L2 speaking–namely WM. This study focuses on measuring a number of variables, including general proficiency, vocabulary knowledge, knowledge of FS, WM and variables from the LHQ. Through these measurements, the study seeks to explain and predict the variance of dependent variables, including both CAF scores of entire speech and CAF scores of FSs. Furthermore, this study aims to investigate an unexplored facet of language production by examining the complexity of FSs, aiming to fill a critical gap in our understanding of FSs and their role in the overall proficiency of L2 learners' speech. Notably, while many studies focus on the CAF of L2 learners' output, there are still few in-depth studies that examine the CAF of FSs themselves, as used by L2 learners, and how this relates to the CAF of the entire speech sample. The current study aims to bridge this gap in our understanding of FSs in speech production.

Researching the relationship between FSs and speech production is of immense importance for both the research community as well as practitioners. Obtaining an understanding of how to integrate FSs into the English classroom, with a specific focus on enhancing speaking production, is vital. This research aims to provide more robust insights and identify specific variables that can help beginner learners in improving their speaking production through the use of FSs. By providing essential data, this study aims to help educators in developing effective teaching methods activities tailored to the needs of learners. Ultimately, this will advance the learning of FSs, thereby enhancing overall oral production among beginner learners.

1.2 Structure of the study

This thesis consists of six chapters. This chapter has already presented the motivations, rationale and the importance of this research. Chapter 2 reviews relevant literature on the relationship between the use of formulaic sequences and speaking ability. It defines FSs and justifies their selection, emphasising their importance in L2 learning, particularly in speaking production. The criteria for FS selection in L2 pedagogy are overviewed. The Literature Review Chapter then delves into linguistic knowledge processing during speaking and cognitive processing. Additionally, it clarifies the challenges faced by L2 speakers in speech production. This leads to a full description of issues related to L2 speaking, including the impact of factors like time constraints and individual differences including WM. After that, the chapter discusses the similarities and differences between L1 and L2 speakers, emphasising key aspects of speech production. It then focuses on qualities of speech production, namely complexity, accuracy and fluency and their typical measurement methods. The subsequent section explores additional factors influencing speech production before laying a theoretical foundation for teaching L2 speaking that this study advocates for. The chapter highlights a number of effective language learning strategies for the L2 classroom and discusses critical issues related to the learning and teaching of L2 speaking. Finally, the chapter concludes by presenting worldwide examples of challenges in teaching speaking before narrowing the focus to challenges faced by learners in Kuwait.

Chapter 3 provides an overview of the current study. It begins by outlining the aims, research questions and related hypotheses. Second, the chapter provides detailed accounts of the study design, participants, tasks, instruments, pilot study, and ethical procedure. Third, the data collection procedure is described in detail. Fourth, the steps undertaken in the data analysis, including the operationalisation and measurement of the complexity of FSs, are described and justified. Fifth, the procedures for transcribing, coding and analysing the data are presented and explained. The chapter concludes with a description of the analytical procedures aligned with the research questions, covering the measures adopted to address each research question in-depth.

Chapter 4 presents the results of the current study. First, the chapter details statistical assumptions, particularly focusing on the evaluation of linear regression analysis. Second, it presents and justifies the descriptive analyses obtained for language background tests, the cognitive task—WM tasks and the duration time of participants' speech samples before reporting the results for the five research questions. Finally, a summary of the key findings will be presented.

Chapter 5 discusses the key findings of this study. First, it provides a summary of the main results. Second, the results of the five research questions are discussed in relation to the hypotheses, previous studies and the qualitative data obtained through post-task interviews. Finally, the chapter explores pedagogical implications for the learning of FSs in general and the Kuwaiti context in particular.

Chapter 6 draws conclusions from the key findings of the current study and emphasises their significance. The contributions of this study are then highlighted before discussing their potential implications. The limitations of the study are then acknowledged. Recommendations and suggestions for potential areas for future research are finally made.

Chapter 2. Literature Review

2.1 Introduction

Both everyday speech and written materials contain numerous fixed expressions, such as *bite the bullet*, *piece of cake* and *kick the bucket*. These expressions find their place not only in dictionaries but also in the mental lexicon of speakers. Consequently, second language learners must learn these phrases if they want to become proficient L2 users. Despite the fact that formulaic sequences (FSs) are widespread in language use (Erman & Warren, 2000; Liang, 2017; Martinez & Schmitt, 2012; Pellicer-Sánchez & Boers, 2019; Schmitt, 2004; Wang, 2018) and a number of studies have been developed during recent years on the effectiveness of FSs on increasing speaking production (Goncharov, 2019; Nergis, 2021; Tavakoli & Uchihara, 2020), until today FSs have not been systematically integrated into teaching materials (e.g., L2 textbooks or vocabulary knowledge test) (Dillon, 2015; Mirzaei et al., 2016; Martinez & Schmitt, 2012; Wood, 2009). Instead, there remains a substantial reliance on teaching individual words (Vu & Peters, 2022). Traditional methods, such as wordlists (Granger & Meunier, 2008); have been the predominant way of teaching vocabulary for decades and which only give information about individual lexical forms (Martinez & Schmitt, 2012). Therefore, in the interest of the research community as well as practitioners, it is important to obtain a more profound understanding on how to integrate FSs into the English classroom, focusing specifically on enhancing speaking production and on methods for utilising activities and tasks that will advance learning of FSs and learners' oral production.

This chapter provides a review of the literature that has discussed the relationship between the use of formulaic sequences and speaking ability. The purpose of this paper is to show how FSs can be used by learners to enhance their speaking ability in order to achieve their personal communicative goals. First, a definition will be given of FSs and clarify the rationale behind selecting this terminology. After which the importance of FSs in L2 learning will be discussed, with a specific focus on speaking production. I then argue what the criteria for FSs selection in L2 pedagogy should be based on. The next section explains in detail how linguistic knowledge is processed through speaking and cognitive processing. In addition, it clarifies the challenges L2 speakers face in the four processes of speech production. It concludes with how L2 speakers' speech production can also be affected by other factors including time constraints. After that, I present cognitive variables that can have impacts on L2 learning—namely working memory, attention and noticing ability. I then provide a summary of similarities and differences between the L1 and L2 speaker, emphasising the aspects that are crucial for comprehending speech production. The specific focus of the analyses will be on complexity, accuracy and fluency, as explained in section of 2.3.5, followed by a thorough discussion to understand how these speech qualities are typically measured. Section 2.3.7 introduces measures of FSs, specifically focusing on accuracy and fluency. Section 2 ends with additional factors influencing speech production, extending beyond the standard scope of complexity, accuracy and fluency of entire speech, and FSs.

Section 2.4 aims to lay a theoretical foundation for teaching L2 speaking by advocating a pedagogic approach guided by essential theoretical principles for this study to follow. The recommended approach emphasises utterance-based learning, particularly highlighting Automatisation in Communicative Contexts of Essential Speech Segments (ACCESS) (Gatbonton and Segalowitz, 2005). In the next sub-section, I introduce the reader to a number of effective language learning strategies that could be employed in the L2 classroom, drawing from well-established theoretical and empirical research. I argue that using strategies like pre-task planning, content familiarity (i.e., repetition) and memorisation and performance can be especially advantageous for lower-proficiency learners. The following section (2.5) delves into critical issues related to the learning and teaching of L2 speaking. I outline various factors compelling reasons why speaking skills are frequently overlooked or receive insufficient attention in L2 classrooms. Here, I aim to shed light on the challenges impacting L2 learning, particularly those preventing speaking development in the classroom setting. It concludes by presenting world-wide examples of similar challenges faced by different contexts in teaching speaking within the L2 classroom. Subsequently, the scope of the discussion narrows to examine this phenomenon within the context relevant to this study – specifically, the situation in Kuwait (2.5.1). Finally, this chapter concludes with a brief summary of the key issues discussed throughout.

2.2 Formulaic sequences

According to Martinez and Schmitt (2012), one of the essential findings from corpus linguistics is that the mental lexicon does not only consist of single words, but also a large number of formulaic expressions/sequences. Wray (2002) identified more than 40 terms used in the literature to describe facets of formulaicity, such as *fixed expressions*, *formulaic language*, *conventional forms*, *lexical phrases*, *prefabricated patterns*, and various others. While the terms multiword expressions, and formulaic language or formulaic sequences are often assumed to refer to the same linguistic phenomenon (Hennecke, 2023; Wray 2002; Schmitt 2004; Thomson et al., 2019), formulaic sequences are considered a “broader” and “more inclusive” term than multiword expressions (Siyanova-Chanturia & Pellicer-Sánchez, 2018, p. 2). For example, while multiword expressions encompass semantically non-compositional (i.e., idiomatic) and highly fixed word equivalents, including compounds (type-token-ratio, *real estate*) and phrasal verbs (*bump into*) (Hennecke, 2023, p. 4), formulaic language extends to include semantically compositional sequences and various fixed sequences of a formulaic nature. These include greetings (e.g., How are you?) (Hennecke, 2023, p. 4), conversational routines (e.g., I don’t know) and common expressions (e.g., by the way) (Martinez & Schmitt, 2012). This distinction is particularly important in some research contexts, such as the investigation of speech samples from L2 learners, especially those with lower language proficiency, aiming to elicit a wide range of sequences for analysis. Additionally, with such proficiency levels (lower proficiency learners), it is more likely that learners will produce semantically compositional sequences than semantically non-compositional ones (these terms will be further discussed in Section 2.2.1).

Taking these considerations into account and following Schmitt (2004), Siyanova-Chanturia & Pellicer-Sánchez (2018), Wood (2010, 2015, 2020), Wray (2002, 2008) and additional scholars, I have chosen *formulaic language* as an umbrella term, while using *formulaic sequences* referring to individual instances of formulaic language. In this paper, the term FSs is used as introduced and defined by Wray (2002, p. 9):

> "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 memory at the time of use, rather than being subject to generation or analysis by the language grammar".

However, it is important to note that not all prefabricated sequences are considered formulaic in this study, as Martinez and Schmitt (2012) have noted that some prefabricated units encode very little meaning, such as *is the* or *is of a,* despite their high frequency. To extend Wray's (2002) definition of FSs, this study also categorises FSs based on words that have a strong relationship with each other in creating meaning (Wray, 2008). Consider the two sequences *think highly of* versus *think of the.* While the former is considered a FS due to its status as a prefabricated sequence that carries a distinct meaning (regardless of whether the sequence is semantically transparent or deceptive; see the following discussion Section 2.2.1 for a detailed explanation of these terms), unlike the latter, which, although frequent in corpora, encodes little meaning in itself.

These words have “an especially strong relationship with each other in creating their meaning”; for example, *by and large*, *of course*, *on the other hand* (Wray, 2008, p. 9). Although these sequences contain several words, they function as a single unit as they aim to reflect a single meaning and are retained as a whole instead of as a group of individual words (Cutler, 2020; Hall, 2010; Khodadady & Shamsaee, 2012; Martinez & Schmitt, 2012; McGuire & Larson-Hall, 2018; Tavakoli & Uchihara, 2020; Wray, 2002; Wood, 2020). In other words, both single formulaic expressions and single words fulfil the same communicative need in a specific communicative context; individual meaning or function corresponds to a form, despite the fact that form consists of several written or phonological words. Martinez and Schmitt (2012) illustrate this by comparing the use of *increasingly* in (1) with the use of the FS *more and more.*

(1) Increasingly, extreme weather events indicate that climate change is upon us.

The single word *increasingly* refers to a condition becoming distinctly predominant. Likewise, the use of the FS *more and more* can correspond to the same meaning (Martinez & Schmitt, 2012, p. 299); however, its usage and placement in the sentence are not necessarily identical.

The contrast in usage between the FS *more and more* and the single word *increasingly* is exemplified in the following instances:

(1) He worked in a series of increasingly senior posts on motorway projects around the UK.

(2) Such a huge amount of information attracts more and more people to use the web.

These examples, drawn from Sketch Engine, highlights a nuanced difference in the use of *increasingly* and *more and more*. While *increasingly* suggests a gradual progression or rise, especially in a sequence, *more and more* emphasises a growing quantity or number in a cumulative manner.

Recent work shows that FSs are vital in language use (Goncharov, 2019; Liang, 2017; Martinez & Schmitt, 2012; Pulido, 2022; Tavakoli & Uchihara, 2020; Van Vu & Peters, 2022; Wood, 2020) and have many advantages for the L2 learner (Goncharov, 2019; Khodadady & Shamsaee, 2012; Pellicer-Sánchez & Boers, 2019; Tavakoli & Uchihara, 2020; Van Vu & Peters, 2022) for the following reasons. First, FSs of different types of discourse represent almost 60% of spoken language (Thornbury, 2019; Erman & Warren, 2000). They are used in all genres, including scientific and academic discourse (Biber et al., 2004; Hyland, 2008; Wulff et al., 2009). Second, meanings and communicative functions are often linked to particular formulas (Martinez & Schmitt, 2012). In other words, certain phrases often carry specific meanings or fulfil particular communication purposes. For example, they can be employed to express a concept (*take into account* implies the need to consider something); handle everyday phrases (*Tell me about it!* implies a statement of strong agreement); express a general believed fact or suggestion (*Money talks* suggests money is influential); indicate discourse markers (*on the other hand* suggests a contrasting point); and can be used for technical expressions (*blood pressure is 140 over 60* reports the value of blood pressure in the medical field) (Schmitt & Carter, 2004). In addition, FSs have processing advantages (Cutler, 2020; Martinez & Schmitt, 2012; Wood 2020) as the repetition of sequences in short-term memory allows their consolidation in long-term memory as a single unit (Cutler, 2020; Ellis, 1996), which means less demand is placed on working memory and processing resources (Ellis 2002, Siyanova-Chanturia & Van Lancker Sidtis, 2018; Wood, 2009). It is for this reason that fluent utterances appear to be formulaic as they are constructed largely of phrases that tend to be retrieved from the mental lexicon as a whole, as prefabricated units (Boers et al., 2006; Cutler, 2020; McGuire & Larson-Hall, 2018; Tavakoli and Uchihara, 2020; Wray, 2002). This means formulaic expressions facilitate effective interaction.

However, the majority of research in this area has mainly focused on learning FSs from written production (see Boers & Lindstromberg, 2012; Pellicer-Sánchez & Boers, 2019 for reviews). Limited attention has been given to the exploration of the influence of FSs on both the learning and usage of FSs in spoken language when compared to studies in written production. For example, an early study highlighting the strong impact of L2 learners' use of FSs in their oral proficiency is the study conducted by Boers et al. (2006) in Belgium, where two groups of college participants were instructed using the same syllabus. However, in the experimental group, learners' attention was drawn to FSs in the sense that learners were encouraged to notice recurrent expressions in the L2 language they were exposed to. For example, the teacher would ask learners to identify frequent sequences through reading. The control group were instructed in "a more traditional way", where a distinction was maintained between grammar and vocabulary (individual words, synonyms, antonyms, etc.) (Boers et al., 2006, p.249). At the end of the training (i.e., after 22 hours of teaching both groups), the participants were interviewed by two judges, who were uninformed about the division of students in an experimental and a control group. The interview included two parts. While the former involved a conversation related to one of the topics that was instructed in class, the latter involved a more unprepared speech, but familiar theme, such as a travelling experience. Both judges suggested that the experimental group were more proficient in terms of fluency, range of expressions (lexical richness and syntactic complexity) and accuracy than the control group. Two other judges who were uninformed about the group the students belonged to listened to the interviews and counted the number of correct FSs. Interestingly, the FS counts demonstrated that the experimental participants outperformed the control students mainly in the first part of the interview (i.e., the part which related to one of the topics they were prepared for). Statistically speaking, findings revealed that the experimental group significantly outperformed the control group with regards to fluency and range of expression, but not with respect to accuracy.

More recently, Tavakoli and Uchihara (2020) explored the correlation between multiword units and L2 oral fluency across 56 learners of proficiency levels ranging from Low B1 to C1 of the CEFR. They found that more advanced learners used multiword units more extensively compared to those with lower proficiency. Based on these findings, the researchers concluded that FSs contributed to oral fluency, where individuals with a large repertoire of FSs retrieved them holistically and used them as a whole unit. This allowed learners to speak more rapidly due to their efficient access to information. This finding shows not only a positive correlation between the use of multi-word expressions and speaking fluency but also with proficiency levels in L2 speakers. While FSs have been shown to assist L2 learners come across as proficient L2 speakers (Boers et al,. 2006; Tavakoli & Uchihara, 2020), they also appear to contribute to the success of L2 learners in language tests, such as the IELTS.

Goncharov (2019) added that L2 learners in Ukraine obtained better grades when using preselected targeted FSs in the IELTS speaking test. Similarly, Mirzaei et al. (2016) in a study among Iranian IELTS candidates, exposed the learners to an intensive two-month lexis-based instruction course. Findings revealed the use of FSs allowed candidates to perform better in their speaking test and to develop a good range of L2 vocabulary needed for everyday use (Mirzaei et al., 2016).

The findings of Boers et al. (2006), Goncharov’s (2019), Mirzaei et al. (2016) and Tavakoli and Uchihara (2020) demonstrate that incorporating FSs in one’s speech improves fluency. Furthermore, FSs can be used as a strategy to facilitate the process of speech for learners who memorise these expressions (e.g., via memorisation) and recycle them in a conversation. In fact, several authors have concluded that FSs facilitate processing, speaking production, the handling of discourse and conversation, and language learning (Bygate, 2006; Carter & McCarthy 2006; Cutler, 2020; Ellis, 2012; McGuire & Larson-Hall, 2018; Nergis, 2021; Tavakoli & Uchihara, 2020; Vu & Peters, 2022; Wray, 2002).

It is important to note that formulaic sequences are "a dynamic, not static, solution" (Wray, 2002, p. 101). This means that there is not a single collection of FSs that all speakers have to learn and make use of; rather, these sequences vary depending on the language, context and individual communication needs and preferences. For example, FSs can prepare learners for speaking exams like IELTS in order to achieve a better score, as such tests have fixed questions that rely heavily on the use of FSs (Khodadady & Shamsaee, 2012). In fact, Wray (2002) argued that L2 learners' use of FSs varies according to their different needs and context. This is why some FSs remain for a short period while others become a constant feature in a speaker's vocabulary (Wray, 2002).

Several studies have posited that there are many chunks that are as frequent as, or more frequent than, the most frequent single words (Thornbury, 2019). For example, Shin and Nation (2008) used the British National Corpus (BNC) and observed that 84 collocations were in the top 1,000-word band (e.g., *you know*, *I think*, *a bit*, *as well*, *in fact*). Similarly, Martinez and Schmitt (2012) developed a PHRASE list for pedagogical reasons with the aim of providing a list of lexical expressions that could be used in teaching, tests and textbooks. The researchers distinguished over 500 'phrasal expressions' (95 per cent of them in the top 1,000 frequent words and 2.88 per cent in the second) that would qualify for inclusion in a list of the top 5,000 lexical families (e.g., *after all, as soon as, make sure, once again*), both written and spoken, in the BNC (Martinez and Schmitt, 2012). These lists would encourage the explicit learning of multiword phrases as single units in pedagogy. Instead of teachers guessing what lexical items to teach and test, the PHRASE List provides them with a principled data set of expressions to teach and test, compiled for pedagogic purposes (Martinez and Schmitt, 2012). Thus, it is important to pay attention to FSs in L2 teaching (Pellicer-Sánchez & Boers, 2019) and to view formulaic vocabulary as "simply being vocabulary" (Martinez and Schmitt, 2012, p.317).

However, the identification of FSs in text is not a simple matter, and range of different criteria have been used for this purpose. The next section looks into this in more depth.

2.2.1 Identification of formulaic sequences

Trying to determine the most frequent formulaic sequence in English by intuition (Martinez & Schmitt, 2012) and personal preference alone is not enough (Koprowski, 2005). This is because it is improbable that the list would be complete. Thus, Wood (2020) emphasised that researchers must apply a range of different methods to identify formulaic language.

Frequency is a determining aspect in the recognition of formulaic sequences (Siyanova-Chanturia & Pellicer-Sánchez, 2018; Wray, 2002). According to Nation (2001), one criterion for deciding which general English words to learn is by their frequency of occurrence. This also clearly applies to formulaic language (Nation & Waring, 1997; Siyanova-Chanturia & Pellicer-Sánchez, 2018). In fact, Martinez and Schmitt (2012, p. 302) referred to the frequency-based approach as a "valid indicator of usefulness". However, it should be taken into account that frequency-based corpora searches fail to detect low-frequency strings (Durrant & Mathews-Aydınlı, 2011). This is because many well-known formulas (e.g., *long live the King, kick the bucket,* etc.) have low frequencies of occurrence, even in large corpora (Wray, 2002, p.30). Wood (2020) emphasised the challenges of a frequency-based approach, especially with smaller datasets, where determining appropriate minimum frequency cutoffs becomes challenging. For example, a sequence might need 40 occurrences per million words to qualify as formulaic in larger corpora, whereas the same sequence might only need four occurrences in smaller ones (see Wood, 2020 for an in-depth review on this topic). Thus, further research is needed on standardising frequency cutoffs to better understand their impact on identifying formulaic language. In other words, frequency alone is not enough to measure formulaicity; nevertheless, we need to know how often a specific expression was attempted and therefore could have occurred when speakers needed to convey a specific utterance at a time (Wray, 2002). Wray (2002, p. 30–31) argued, this knowledge helps us to understand how some utterances are favoured over others as some expressions are more used than others (e.g., *happy birthday* over *many happy returns*). Once we have this information, we can then truly understand the power of the formulaicity.

Furthermore, several authors (e.g., Mathews-Aydınlı, 2011; Martinez, 2013; Thornbury, 2019; Wray, 2002) argued that the limitation of using computer searches to detect frequent strings is that they provide only indirect evidence for both non-compositionality (i.e., idiomaticity) and holistic units. Compositionality or semantically transparent FSs, are phrasal expressions that a learner is likely to understand, for example, *at this time*, as their meaning remains the same even if each word is substituted with its own definition (at+this+time); equivalent to “now” (Martinez & Schmitt, 2012). In contrast, a non-compositional formula or idiomatic expressions are expressions that as Laufer (1989) terms as “deceptively transparent”; words learners assume they understand but they misunderstand. For instance, *for some time* may be misunderstood as *a short amount of time* (Martinez & Schmitt, 2012). Therefore, compositionality or transparency overlaps with the clarity of formula meanings which makes the selection of chunks to not just be based on frequency, but also on other kinds of information (e.g., semantic analysis). In this sense, relying solely on frequency counts cannot differentiate between the occurrence of formulaic sequences and the same sequence of words arranged as a new combination of smaller units (Wray, 2002). Grant and Nation (2006), however, argued that the number of deceptively transparent multiword expressions in English is low and, therefore, should not be a major focus in L2 learning. However, deceptively transparent FSs can depend on contexts. If L2 learners think only in their L1 frame and are not aware of the English FSs, many errors will develop in their oral English expressions. An example of the difficulties experienced by L2 learners can be found in Liang (2017) who investigated the use of formulaic sequences in the popular American comedy show *Friends* via Corpus of Contemporary American English (COCA). The researcher referred to a conversation with Rachel saying, "Oh, Joey, you know what, no-one is gonna be able to tell". In this dialogue, *you know what* is a formulaic sequence. According to Liang (2017), many Chinese students would consider this formulaic sequence as a question, thinking they would have to respond with a "Yes" or "No" answer. Therefore, L2 learners can be instructed about the formulaic sequence *you know what*, used to start an utterance to gain the attention of the listener in order to then convey a certain message. In return, the listener knows that when the speaker uses the formulaic sequence *you know what*, the speaker is about to begin a conversation and will not need an immediate answer; the listener will only respond after the whole utterance. According to the author, if these FSs can be understood and utilised by learners their speaking ability can be improved (Liang, 2017).

Recent research has advocated the identification of FS based on Mutual Information (MI) score. The MI measure is a metric not only of word pair encounter (frequency with which two words co-occur) but also a measure of how strong the co-occurrence between a two-word combination is (strength of association – appearing together in a prepared sequence rather than by chance; Carroll & Conklin, 2020; Webb, 2020). According to Simpson-Vlach and Ellis (2010), there is no minimum threshold value for the MI score. This is because the MI is a scale and not test of significance; thus the information the MI score presents is used to compare with other MI scores. The higher the MI score, the stronger the association between words, whereas lower scores suggest that the combination of words is more likely due to chance (Simpson-Vlach, 2010). However, Li & Schmitt (2010, p. 37) indicated that certain threshold within this scale provide an outline for the strength of collocations represented by MI values, where an MI score of 3<MI<5 suggests “moderate-strength” word combinations, an MI value of 5<MI<7 indicates “stronger” word combinations, and “extremely strong” word-combinations have an MI value over 7. This assertation of the minimum value of the MI score being 3.0 is in line with many researchers (e.g., Carroll & Conklin, 2020; Millar, 2011; Webb, 2020) who used an MI value of 3.0 as the threshold above which a word combination suggests a statistical strength of co-occurrence.

Moreover, Martinez and Schmitt (2012) stated that the compilation of their phrasal list was not totally frequency-driven, as they did not want it to include sequences such as “is the” or “is of a” (Martinez and Schmitt, 2012), which can easily fulfil the frequency criteria but may not truly qualify to be formulaic (Wood, 2020). Thus, FSs should be selected according to the high frequency of the function it expresses, meaningfulness and relative non-compositionality in order to increase their usefulness and for them to be similar to the individual words in a standard frequency-based wordlist (e.g., Durrant & Mathews- Aydınlı, 2011; Martinez and Schmitt, 2012).

The key focus of this study is on speaking. Therefore, the next sub-section will cover spoken language in great detail.

2.3 Spoken language

This section discusses spoken language in relation to the use of formulaic sequences. It starts by defining and discussing the features of speaking ability (2.3.1). After that, I explain how linguistic knowledge is processed through speaking and cognitive processing in detail. In addition, this section clarifies the challenges L2 speakers face in the four processes of speech production. It concludes on how L2 speakers’ speech production can also be affected by context/time (2.3.2). I then present one of the factors that can have impacts on L2 learning—namely working memory (2.3.3). After that, a summary of similarities and differences between L1 and L2 speakers will be discussed, aiming to gain an understanding of the underlying psycholinguistic processes experienced by both L1 and L2 speakers during online speech performance (2.3.4). I will then proceed to evaluate the key qualities of speaking—namely fluency, accuracy and complexity (2.3.5). I will explain how these speech qualities are measured (2.3.6), narrowing the scope to the measures of fluency and accuracy of FSs (2.3.7). Finally, I will conclude this section by discussing additional factors that influence speech production, as will be outlined in section 2.3.8.

2.3.1 What is speaking?

Speaking is a form of social interaction (Alonso, 2018), which suggests that our linguistic capabilities are linked with other people; the context in which we use and learn linguistic utterances is created by other people who are doing the same thing reciprocally (de Jong et al., 2020; de Jong, 2023; Eskildsen & Cadierno, 2015). One of the vital aims of interaction is to be comprehensible to listeners (Correia, 2016; Saito et al., 2016; Saito & Liu, 2022) by expressing meaning through different linguistic resources such as patterns of linguistic structures, i.e. FSs (Alonso, 2018; Saito, 2020; Saito & Liu, 2022). Thus, it is important to note that speaking does not exist in isolation, as successful speakers must also be listeners who can communicate and respond to different aspects of speech (Ellis, 2014; Thornbury, 2005). Clearly, most language learners spend more time receiving input (listening) than producing it themselves (speaking).

de Jong et al. (de Jong et al., 2012) explained that speaking proficiency is multi-faceted, containing different linguistic components and language-processing components. While the former refers to declarative knowledge (explicit knowledge, hence with awareness) that later refer to procedural knowledge (the knowledge how to do something without awareness) (de Jong et al., 2012; Ullman, 2016; Ullman & Lovelett, 2018). From a psycholinguistic perspective to SLA research, models of speech production, exemplified by Levelt (1989) and adapted for L2 oral production, play a pivotal role in comprehending the mechanisms and processes involved in learners' L2 performance. The following section will present and describe the characteristics of spontaneous speech of both monolingual and bilingual adult speakers.

2.3.2 Psycholinguistic models of speaking

Speaking is a highly dynamic psycholinguistic process consisting of complex cognitive and linguistic skills. To theorise speaking as a cognitive process, Levelt (1989) offered a model of L1 speech production that has been adapted for L2 oral production (e.g., de Bot, 1992, 2020; Kormos, 2006, Segalowitz, 2010). The model (see Figure 2.1) represents speaking as involving four major processes: conceptualisation, formulation, articulation and self-monitoring. There is a consensus that three of the major processes occur in sequence (i.e. conceptualisation followed by formulation and subsequently articulation) (de Jong, 2023; Kormos, 2006), while self-monitoring is present throughout.

Conceptualisation involves the selection of a topic and seeking background knowledge (i.e., knowledge about the subject, about the discourse situation and the pattern of speech). The output of the conceptualisation process is a preverbal message, a conceptual structure which has not yet been formulated in words (De Bot 2020; Segalowitz, 2010). This means the message at this stage does not yet consist of lexical items and grammatical structure—the surface structure. In order for the message to be implemented with surface structure, it needs first to be encoded into a grammatical structure so that knowledge from the mental lexicon can then be added (Segalowitz, 2010). Therefore, this preverbal message is generated through two stages of macroplanning and microplanning (in the conceptualiser).

In macroplanning, the speaker chooses and organises the knowledge to be conveyed to achieve a specific communicative intention (de Jong et al., 2018). In other words, macroplanning is planning what to say. This stage allows the L1 speakers to select different speech registers to represent the utterance to be conveyed (Levelt, 1989), for instance, to account for whether it will occur in the context of formal or casual discourse. De Bot (2020) and Paradis (2004) posited that choosing which language to use is similar to selecting a speech register (e.g., formal versus informal speech) and this preference will be governed according to the socio-pragmatic knowledge of the circumstance.

Subsequently, performing the selection takes place at a microplanning stage, a process that results in the formulation of a preverbal speech (De Bot, 2020; Ellis & Yuan, 2005; Segalowitz, 2010). Like macroplanning, microplanning is not a lexical process (putting ideas into words). Therefore, despite the success of speech that can occur in an interaction, the processes underlying speech productions are not always smooth and efficient. This is because the string of words speakers produce is interrupted by a variety of disfluencies, including short pauses, repetitions and repairs and filler words (e.g., *uh* and *um*) (de Jong et al., 2018). Such disfluencies sometimes occur when speakers have a problem deciding how exactly to formulate or articulate speech (de Jong et al., 2018). This can happen at the time of topic shift (planning what to convey for the following topic) and takes resources away from the microplanning needed to turn ideas into utterances (De Bot, 2020; Segalowitz, 2010).

A monitor is involved in the conceptualisation, ensuring a successful speech—it allows the speaker to self-correct for grammar, expression and pronunciation. Then follows the formulator, which puts topics/ideas/information into specific words and grammatical forms as well as phonological encoding to convey meanings. After formulation, articulation involves using the articulatory organs to produce the message for the listener. Finally, self-monitoring enables speakers to recognise and self-correct errors. Kormos (2006) explained that both L1 and L2 speakers often monitor their own messages. This enables them to detect errors and re-express their message. Accordingly, in self-monitoring there can be a trade-off between retaining accuracy versus fluency in speech (Segalowitz, 2010). That is to say that self-monitoring speakers interrupt themselves to ensure accuracy resulting in said trade-off with flow and fluidity (fluency) (Segalowitz, 2010). As posited by Segalowitz (2010, p. 16) “clearly”, such self-monitoring can be more cognitively challenging in the L2 than in the L1, depending on the context in which one is interacting and the speaker’s level of proficiency.

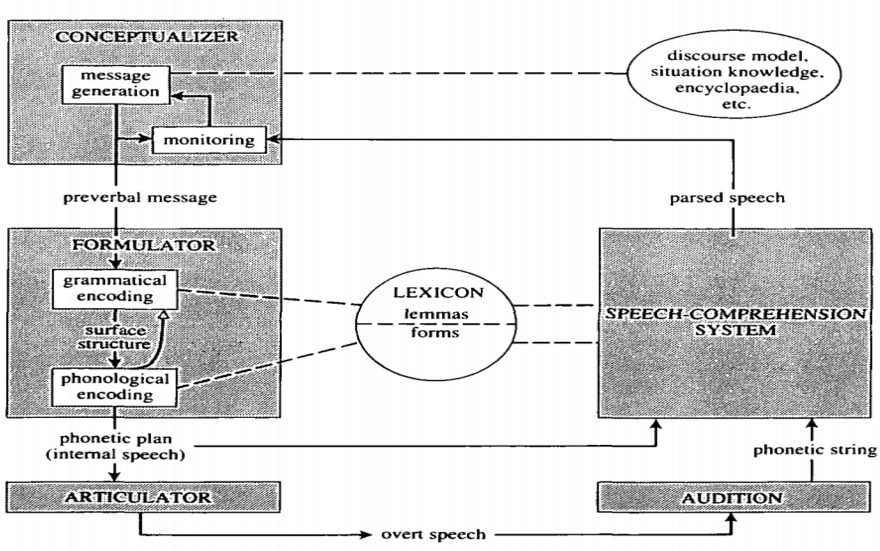


Figure 2.1 Levelt’s speech production model (Levelt, 1989)

This process happens quickly, and applying it effectively requires automatisation (Bygate, 2015, De Bot, 2020; Levelt, 1989). Automatisation is defined as fluent speech processing that results from extensive use and repetitive practice (Gatbonton & Segalowitz, 2005; Jeong & De Keyser, 2023; Mitchell et al., 2019), demanding minimal mental effort (Suzuki, 2023). This means that the development of automaticity is a slow process (Jeong & De Keyser, 2023). Thus, speaking can be difficult for L2 speakers as it requires various processes of speech production to be carried out simultaneously, including conceptualisation, formulation, articulation, self-monitoring and automation. Automatisation is important as humans have a limited and selective attention capacity to control these speaking processes consciously (Bygate, 2015). Automatisation reduces demands on WM, as processes that have been automatised no longer require conscious attention.

Unlike L1 speakers, L2 learners encounter a variety of challenges in the four processes (conceptualisation, formulation, articulation and self-monitoring) of speech production because of the increased burden on their cognitive resources, namely attention and working memory capacity (WMC) under time pressure (Bygate, 2018; Kormos, 2011). Because of their slower processing during formulation and articulation, and they also have to develop lexical and grammatical knowledge in the L2, L2 speakers are more likely to fall behind in formulating speech than L1 speakers (Bygate, 2018; De Bot, 1992, 2020; Kormos, 2006). L2 speakers are therefore more likely to lack fluency and accuracy as they are likely not to have sufficient WMC to develop all the components of speech in real time (Bygate, 2015, 2018). Therefore, the degree of efficiency of speech production (including the automaticity of encoding, planning the content of the topic, and managing speech breakdowns) performed by an L2 speaker determines his/her fluency (Lambert et al., 2017).

In addition to limited cognitive resources, L2 speakers are also influenced by temporal constraints and pressure to produce an utterance (Yuan & Ellis, 2003). Bachman and Palmer (2010) and Bygate (2018) stated that speaking is reciprocal—a speaker is involved in an oral interaction and responding instantly to the other interlocutor, making speaking a more challenging skill than written interaction (Bygate, 2018; Lowie et al., 2018). For example, speakers have to interact without taking the time to review and correct their speech as the listener will not wait long to re-communicate (unlike writing). Therefore, time pressure influences the process of conceptualisation, formulation and articulation, as speakers then do not have time for detailed planning of their utterances, which leads to pauses and corrections (Bygate, 2001, 2018). For this reason, having an expression ready to use helps the speaker avoid the effort of putting together the components of the expression every time they want to use it, as easier access to FSs allows for automaticity in response (Wray, 2002). Indeed, there is a correlation between an increase in processing pressure, subject to the amount of information to be processed and conveyed within a specific time, and an increase in the number of FSs the speaker knows (Wray, 2002). This is because retrieving formulaic sequences is an easier choice than deciding to express something and creating an utterance from scratch (Wray, 2002).

Of course, there are other cognitive variables that can have an impact on L2 speaking, namely working memory, attention and noticing ability.

2.3.3 Additional variables that can influence speaking in an L2

For a long time, linguists and language educators have been concerned with investigating factors that impact second language learning, particularly aspects that influence student talk time in class. Some complex and multidimensional factors involve cognitive variables, particularly WM, attention and noticing ability (Chan et al., 2012). WM is a system in the mind that chooses what to focus attention on (Mitchell et al., 2019). Attention selects and directs information to WM, which, in turn, supports cognitive functioning, including the enhancement of attentional control. Attention is therefore defined as a limited set of mental resources that determines L2 development, including the construction of new knowledge, retrieval of existing knowledge, and variation (Schmidt, 2001). On the other hand, noticing refers to a phenomenon that occurs when learners direct their attention to specific aspects of language. Noticing refers to the experience of drawing learners' selective attention to specific linguistic aspects (Mitchell et al., 2019). Sawyer and Ranta (2001, p. 342) discussed how these variables are thoroughly interconnected in determining a learner's capacity for learning, positing: if we presume that noticing is essential for learning and that attention is a prerequisite for noticing, and considering that one's attention capacity at any moment is constrained by their WMC, it logically follows that there must be a strong relationship between amount of learning and the capacity of one's WM.

WM is a determining individual difference variable in SLA (Baddeley, 2015; Conway et al., 2007; Mitchell et al., 2019, Skehan 2015). In fact, Skehan (2002, 2015) argued that L2 learning attainment depends on individual differences such as differences in WM. More specifically, WM refers to operations included in the temporary storage, management and maintenance of related information in the task while dealing with online cognitive processes involving language comprehension and production and learning in general (Miyake & Shah, 1999). WM is crucial to all higher cognitive tasks, such as learning abilities, maths skills, reasoning, speaking and language comprehension (e.g., Baddeley, 2001, 2003, 2015; Conway et al., 2007; Mitchell et al., 2019). The main task of WM is to provide "an interface between perception, long term memory, and action" (Baddeley, 2003, p. 829). For instance, when a teacher asks learners to describe a picture, L2 learners usually attend to the message that the picture expresses, and while holding this whole picture in their mind they attempt to search for the suitable form that best corresponds to the message (Ahmadian, 2013). This entire process of holding a piece of information and performing cognitive work upon it is carried out by WM.

The model of WM most widely recognised and considered a classic is the one developed by Baddeley and Hitch in 1974. A later version from Baddeley (2000) is presented in Figure 2.2. The original model had three main components: (a) *phonological loop*, which is responsible for maintaining phonological and verbal data and is assumed to have a significant role in L2 vocabulary learning (Baddeley, 2003, 2015), (b) *visuospatial sketchpad*, which is responsible for holding visual and spatial coding, and (c) *the central executive* (the core component of the WM system), responsible for the attentional control of WM, dividing attention between two vital stimuli and transferring attention from one task to another. Afterwards a fourth component, the *episodic buffe*r, was added (Baddeley, 2000), where different information is temporarily bound and held. The three short-term storage sub-components are called *slave systems* to represent their passive roles as sources of information controlled by the central executive. The shaded area in the Figure 2.2 indicates long-term storage allowing the components in the WM to access it. The WM can be best thought of as a channel through which knowledge has to pass in order to be permanently stored in long-term memory (Juffs & Harrington, 2011).

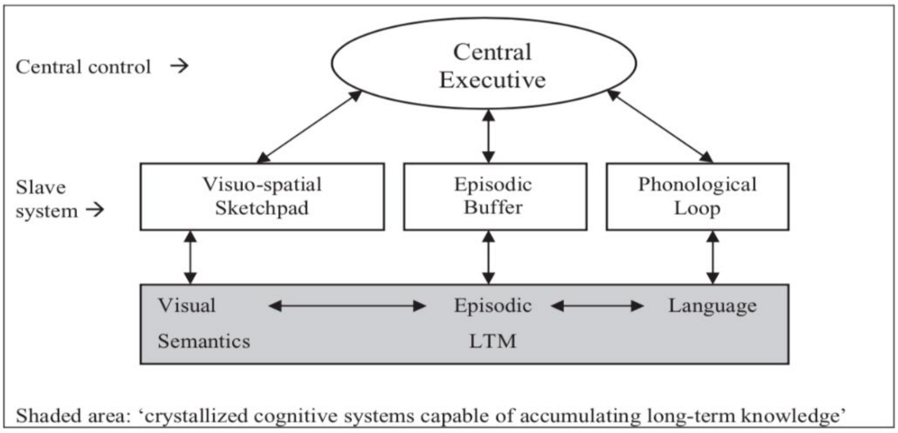


Figure 2.2 Processes of the working memory model (Baddeley, 2000)

Given the important role of WMC in human cognitive tasks, variations between high and low WMC may predictably result in variations in operating everyday higher-level tasks. These tasks include speaking, reading, mathematics, etc. (Conway, et al., 2007), as well as controlling thought and action, prioritising cognitive processes and switching between tasks (Mitchell et al., 2019). There is also evidence that L2 learners with higher WMC can gain more benefits from interactional feedback than learners with lower WMC (Mackey et al. 2010; Révész, 2012). This is because individual differences in WMC indicate differing ability to pay attention to the information, which is essential to both finishing a task and to prevent distraction (Ahmadian, 2013). Similarly, Rosen and Engle (1997) reported that only participants with high WMC were capable of producing oral tasks fluently while monitoring their speech. They also posited that participants with low WMC made more errors (i.e., lower accuracy) and did not have enough WMC to commit to the processing components that needed controlled attention. Indeed, these findings are corroborated in the studies of Gilabert and Munoz (2010) regarding fluency and are consistent with the studies conducted by Kormos and Trebits (2011) as well as Awaad and Tavakoli (2022) regarding accuracy.

In their correlational study, Gilabert and Munoz (2010) examined 59 participants (with L1 Catalan/Spanish) who were categorised into low and high language proficiency groups (high-intermediate/advanced learners of English). The researchers aimed to investigate whether differences in WM could account for variations in language proficiency and L2 performance. Among the tests employed in their study, the reading span test was utilised to gauge the participants' WM. The participants were asked to perform a single video-based narrative task (i.e., a film retelling task). The results revealed a significant correlation between WM and both fluency and lexical complexity in L2 speech performance.

Along the same lines, Kormos and Trebits (2011) conducted a study involving 44 secondary school students in their second academic year of an English-Hungarian bilingual program in Hungary. These students were rated as slightly above intermediate proficiency level (B1/B2 in the CEFR). Their research aimed to investigate the relationship between WMC and two narrative tasks in L2 oral performance. The tasks included a simple one, which involved telling a story, and a more complex one, which required inventing a story. Their findings indicated that a high WM had a positive impact on syntactic complexity (which was measured through various measures such as subordination, length of clause and the use of different verb forms) of narratives in the simple task. In the complex task, a high WM was associated with increased accuracy but less lexical complexity. Furthermore, no significant effect was observed on fluency. However, it is worth noting that this study relied on teachers' judgement to assess participants' language proficiency instead of using a standardised test, which could potentially have affected the results.

In a more recent study conducted by Awaad and Tavakoli (2022) examined 48 learners of English with a range of language proficiency at a secondary school in Jordan.

They investigated students' performance in two video-based narrative tasks that required varying levels of intentional reasoning. In line with the findings of Gilabert and Munoz (2010), Awaad and Tavakoli (2022) also found a correlation between WM and lexical complexity but not with fluency. Furthermore, Awaad and Tavakoli (2022) established that WM was associated with accuracy in both complex and less complex task types on second-language speech performance. However, both Mizera's (2006) and Gilabert and Munoz (2010) argued that differences in WM were significantly correlated with the performance of participants in the high proficiency group. This is because in the early phases of language learning, learners tend to allocate more cognitive resources to store immediate knowledge between conceptualisation and formulation than advanced learners would. During this initial phase, learners tend to rely heavily on attentional and memory resources to access and retrieve lexical items and construct their messages with appropriate syntax and morphology, thereby making working memory less significant (Gilabert & Muoz, 2010). Furthermore, the inconsistent findings regarding the association between high WMC and fluency in Gilabert and Muoz's study (2010) stand in contrast to the results of Kormos and Trebits (2011) and Awaad and Tavakoli (2022). This discrepancy, where the latter studies did not establish such a relationship, could be attributed to several factors. One of these factors is methodological, Gilabert and Munoz (2010) measured WM using the reading span test, whereas both Kormos and Trebits (2011) and Awaad and Tavakoli (2022) employed the backward-digit span. The latter is considered language-independent and, as a result, helps to minimise any influence of language proficiency on WM scores (Harrington and Sawyer 1992; Wright, 2010).

Révész (2012) explained that different subcomponents of WM are connected with different language sub-skills. For example, Baddeley (2015) pointed out that a particular component of WM which is likely to influence L2 learning is the central executive component due to its impact on comprehension, which in turn hinges on the ability to connect ideas in a coherent manner. Furthermore, it has been observed that learners with high phonological short-term memory were better at speaking (Révész, 2012). This is because the processing and temporary retention of both familiar and novel phonological knowledge depends on the phonological short-term memory or "phonological loop" (Baddeley, 2015; Juffs & Harrington, 2011), which corresponds to higher WM and guides more effective speaking. As FSs are treated as full chunks this can facilitate retention, because less WM space is needed (Cutler, 2020; Ellis 2002, Jolsvai et al. 2020; Wood, 2009, Wray, 2002). Therefore, the more WM one has, the more attentional resources can be processed, allowing for more learning, as enough memory will be available to process form simultaneously with meaning (Juffs & Harrington, 2011; Skehan, 2015). The significance of phonological short-term storage capacity to vocabulary learning is evidenced in Speciale, Ellis & Bywater (2004). The researchers conducted a 10-week study and experimental course in a university Spanish programme, showing that learners' ability to learn phonological strings predicted their ability to learn novel L2 words. They argued that vocabulary learning is controlled by both phonological short-term storage capacity and the learner's development of the phonological reoccurrences of language. However, Hamada and Koda (2011) and Juffs and Harrington (2011) stated that non-alphabetic language in learners' L1 may decrease the reliance on the phonological loop in vocabulary learning.

The role of WM in interaction has also been investigated. Mackey et al. (2002), the first to introduce WM into interaction research, found that learners with higher WMC spent longer analysing their output and compared it with existing knowledge, unlike learners with lower WM, who did not. In a more recent study, Kim et al.'s (2015) investigation included the relationship between WM and its influence on L2 question formation in English. Their study involved 81 intermediate-level English learners, revealing that the executive WM was a significant predictor of recast effects (i.e., influence of corrective feedback). Interaction is as much necessary as processing for learning, as it guides to the type of communication breakdown that allows learners to notice the gap between what they desire to convey and their ability to state what they want to convey using suitable words and forms (Juffs & Harrington, 2011). In fact, Mitchell et al. (Mitchell et al., 2019) suggested high WMC may be most beneficial when a certain level of noticing about language use is present. This means when learners notice language use, it can improve the efficiency and effectiveness of WM in language-related tasks, as these two cognitive processes complement each other in language learning and usage. Noticing via output and negotiation activate learning through (self) correction and enhanced input (Juffs & Harrington, 2011). Thus, large WMC may be most beneficial when some level of awareness about the L2 is involved (Williams, 2012).

In addition to investigating the role of WM in language learning, numerous studies have delved into the significance of attention and awareness in L2 learning. Attention is important for all aspects of L2 learning (DeKeyser, 2015; Schmidt, 2001; Suzuki, 2023). This is because it allows learners to become aware of the gaps between “what they can produce and what they need to produce” in addition to those between what they produce and what skilled target language speakers produce (Schmidt, 2001, p. 6). Thus, it focuses on “directing awareness” to the target language while learning it (Segalowitz & Hulstijn, 2005, p. 378; Schmidt, 2001), for example, the noticing of word order regularities (Wen & Skehan, 2021). Furthermore, beginner learners are cognitively overloaded and unable to pay attention to all meaningful differences at the same time (Schmidt, 2001). Schmidt (2001) therefore argued that if learners do not learn the basics, they cannot learn the complex details in a language. In fact, Doughty (2001) and Wen et al. (2021) proposed that noticing linguistic aspects such as lexical items, grammatical structure and phonology is the first phase in L2 processing phases, which steadily develop comprehension and integration. Therefore, if simpler processing routines are over-learned, learners will have better capacity to pay attention to details, eventually being capable of paying attention to aspects that L1 speakers attend to (Schmidt, 2001). This, therefore, allows growth in L2 knowledge as learners gradually develop their L2 knowledge by analysing the input they noticed.

Schmidt (1990, 1994, 2001) played a pivotal role in the development of what he referred to as the “Noticing Hypothesis”— that L2 learners are incapable of learning a language feature until they become conscious of that feature in the input. That is, people learn about the items that they attend to. The importance of noticing is demonstrated through Schmidt’s (1990) personal experience while learning Portuguese. He mentioned some long forms of question words that he tried to process; however, the researcher only started using these forms when they were eventually noticed (after five months), even though they had been present in the environment for the whole time (Schmidt, 1990). This was because these aspects were brought to his attention in class or made salient by other practices (Lightbown & Spada, 2021). Thus, the process of producing the L2 may enable learners to notice the gaps and problems in their current L2 system; it provides them with the opportunity to reflect on, discuss and consider these problems explicitly and offers them the chances to test new structures and forms (Swain, 1985). However, Schmidt (2001) made it clear that only the initial explanation of a feature requires conscious awareness. After this, explicit awareness is no longer needed and associative strengthening of the representation of the feature continues implicitly (Schmidt, 2001). However, this assumption that learning can only take place when consciously noticed is not necessarily true. Another form of learning, known as incidental learning, occurs when individuals acquire knowledge or skills without deliberate awareness of the learning process (de Jong et al., 2012; Ullman, 2016; Ullman & Lovelett, 2018). Also, the key question should not concern whether any learning can occur without attention, but instead whether more conscious involvement and attention leads to better learning.

However, the Noticing Hypothesis and the significance of attention and awareness—even for the initial explanation of a feature—have also been strongly criticised. Tomlin and Villa (1994) used the term "detection" to illustrate attention that happens without awareness, and emphasised its importance in SLA. Carroll (2006a) stated that language learning input, encompassing phonemes, syllables, morphemes, nouns, verbs, etc., involves structures that exist within the mind (related to metalinguistic knowledge) rather than in the environment (linguistic competence). However, it has also been mentioned in the literature that L2 classrooms that promote noticing tend to yield more significant benefits than those that do not (Hama and Leow, 2010; Mitchell et al., 2019). William (2005) investigated whether a form–meaning relation could be comprehended when participants' attention was drawn to it, and found that all learners who were aware of the relationship performed significantly better than the unaware learners. Hama and Leow (2010) investigated the role of awareness and found that none of their learners who were unaware of certain elements managed to learn, emphasising the essential role of noticing.

Fluency seems to also benefits from noticing and explicit teaching of multi-word expressions, as highlighted in the intervention study conducted by McGuire and Larson-Hall (2018). This study involved 19 mid-intermediate to advanced L2 students at a U.S. university, encompassing diverse native languages including Chinese, Thai and Japanese. The control group (N = 8) received instruction primarily centred on individual words and grammar rules, representing a traditional method of learning. In contrast, the treatment group was explicitly guided to notice and use FSs during their instructional sessions. Both groups engaged in communicative activities, with the experimental group specifically encouraged to integrate FSs into their conversations. Findings included a significant improvement in fluency, particularly in speech rate and mean length of runs, within the treatment group. This finding implies that noticing and using FSs played a facilitative role in learners' spoken interactions.

However, the relatively low number or participants raises concerns about the generalisability of the findings. Additionally, the use of dialogic format for the pre-and post-tests, might have influenced the scores in various ways. For example, participants' selections or application of FSs during these interactions could have affected the results. Furthermore, these results should be interpreted with caution, considering that the study took place in an immersion context where ample exposure is provided, contrasting with non-immersion context. Thus, the improvements identified in this study between fluency and FS use may not be applicable to situations where students do not receive substantial language input.

The usefulness of both noticing and explicit instruction has also showed gains in retrieving knowledge from learners' memory. Damanhouri (2018) interviewed 10 instructors of Arabic for adult non-native speakers (from Pakistan, India and Bangladesh) at King Abdulaziz University, and 15 students learning the modern form of Arabic in the same institution after having formerly memorised several chapters from the Qur'an. Findings revealed the usefulness of the noticing hypothesis in retrieving knowledge, such as phonological, lexical and structural data from the learners' memories. The researcher also stressed the effectiveness of explicit instructions, as such methods help learners relate to their inactive knowledge (Damanhouri, 2018). However, the researcher argued that the language is not learned because of noticing, but that it is the first stage in L2 processing development, and it gradually leads to comprehension and integration (Damanhouri, 2018). Similarly, Izumi (2002) compared the outcomes of output and modified input on noticing and development. Findings revealed that participants showed more noticing and more learning than did controls, and that modified input demonstrated more noticing but not more learning, suggesting that noticing is only the first stage of L2 learning.

One significant role of explicit instruction is that it helps focus attention on forms and meanings in the input, a requirement for the processes that follow (Schmidt, 1990).

Damanhouri (2018) provided an example regarding the learners in his study when they were learning Arabic vocabulary. They were supposed to learn the word and the way it is inflected for number, gender, and case, as well as knowing its synonyms and antonyms in order to notice and employ that word in varied contexts, given that Arabic is a highly inflectional language. According to Damanhouri (2018), explicit instruction develops the learners' inactive knowledge, making it available for use as they become aware of it. Indeed, Schmidt (2012) explicitly stated that the number of corrections learners receive from their teacher has no effect on the intake for learning as it is not a matter of correction— rather one of directing awareness to the correction. Thus, in order for learners to avoid errors, they should make conscious evaluations between their own output and input (Schmidt, 2012). Attention and noticing are necessary for understanding L2 learning as they play a major part in determining L2 development, including the construction of new knowledge, access to that knowledge, and variation (Schmidt, 2001; Schmidt, 2012). This is especially true for individuals who are learning English as a second language without immersion. In such cases, it is unlikely that they will encounter enough opportunities for L2 learning, such as FSs, to occur without some form of explicit instruction (Thomson, 2020).

Of course, learning can also occur implicitly, that is, learning without instructions. However, even if incidental learning occurs, this does not mean that learning happens without awareness at the time of learning (Schmidt, 2012).

2.3.4 Speaking in L1 and L2: a summary of similarities and differences

As previously mentioned, Levelt's monolingual speech production model has been expanded to be used for L2 speakers, as many researchers assume that both L1 and L2 speech production have the same underlying psycholinguistic mechanisms (e.g., De Bot, 1992, 2020; Kormos, 2006; Segalowitz, 2010). As this has been described in the previous section, I will restrict myself to a summary of similarities and differences between the L1 and L2 speaker, highlighting those aspects that we view as important for understanding speech production.

As mentioned in Section 2.3.2, during the macroplanning, speakers make decisions regarding what to convey–specifically, they choose and structure the information to fulfil a particular communicative purpose, such as choosing the appropriate speech register. Therefore, one of the similarities between L1 and L2 speakers in the speaking process is their similarity at this stage as information regarding language choice or register choice is part of the information package, all the other information (external world knowledge, the interlocutor's internal state of mind, discourse knowledge, etc.), that is available in the macroplanning stage influences the eventual formulation of the message (Segalowitz, 2010).

Both L1 and L2 speakers hesitate in the macroplanning stage, as they try to figure out what they would want to express in the first place (de Jong et al., 2018). Therefore, in this situation, the difficulty does not involve linguistic encoding, instead, it pertains to the conceptualisation or the creation of the messages' content to be conveyed (Levelt, 1989). This is because increased processing load resulting from conceptualisation difficulty probably has a negative impact on fluency (de Jong et al., 2018). In fact, this is evidenced in Roberts and Kirsner's (2000) study on nine native speakers of English, in which they assess how they speak spontaneously about themselves for about five minutes. The speech was examined with regard to two measures—fluency and topic shifts in the participants' speech. Results showed that fluency decreased during topic shifts. The researchers argued that speech does not become fluent until the macroplanning process is accomplished and "the system's resources are available solely to speech preparation and production processes" (Roberts & Kirsner, 2000, p. 153). Similarly, L2 speakers may encounter dysfluencies in communicative situations that require a higher level of macroplanning because of the change of processing resources (Segalowitz, 2010). This means that novice L2 learners who have difficulty executing a microplanning will automatically need additional time to ensure that macroplanning has been accomplished effectively before they try to speak. It is for this reason that the level of language mastery cannot affect the performance of effective macroplanning processes (Segalowitz, 2010). However, teaching learners FSs can take some of the effort out of planning as learners will be able to express fixed, memorised chunks rather than spending more time thinking about how to put a sentence together, which makes their speech more effortful and conscious.

L2 speech is often considered to have more disfluencies than L1 speech (de Jong et al., 2013; de Jong, 2023; Kormos, 2006; Liu, 2020; Segalowitz, 2010; Uchihara & Tavakoli, 2020; Viera, 2017). This can be explained by L2-specific difficulties in formulation and articulation (Segalowitz, 2010) along with having to inhibit L1 use (de Jong et al., 2018; Kormos, 2006). Indeed, L2 speakers lacks fluency because their lexical and grammatical L2 knowledge and abilities are still developing (De Bot, 1992; Viera, 2017). This makes it more likely for the L2 speaker to fall behind in speech formulation than the L1 speaker, as L2 speakers have a reduced processing speed during formulation and articulation and are therefore likely to show more disfluencies in speech (de Jong et al., 2018). In fact, the L2 speaker may lack or have partial knowledge of specific lexical items and the way they associate to each other (grammatical structure) in order to express the speaker's intention (i.e., content of the message). Consider the following example provided by Levelt (1999, p. 91–92), *there is a house with a tree to the left of it.* According to Levelt (1999, p. 91–92), there are many possibilities when expressing a message. At the microplanning stage, speakers have to choose which perspective to consider in order to reflect the way they interpret the event by using the appropriate preverbal message. In the previous example, the preverbal message goes beyond deciding which object is to the right or the left of the other, as it includes other factors such as expressing information that considers the position of the speaker, or of the listener, or of the house, or of the speaker's specific intention in expressing (that is a message regarding the house, and not the tree) (Segalowitz, 2010). Thus, L2 speakers often have to make an effort when formulating the preverbal message in order to overcome communicational problems; that is, by using strategies. Such strategies include formulating the preverbal messages (i.e., which includes transforming concepts/ideas/information into specific words and grammatical structures) to bypass their linguistic shortcoming (Segalowitz, 2010). It is of great importance to investigate to what extent relying more on fixed expressions may help L2 learners with their speech production; as they have minimal exposure to the target language, and producing novel utterances can be difficult. The use of formulaic sequences may reduce or avoid disfluencies by allowing the user to carry on with the formulation of the preverbal message.

Another major source of difference between monolingual and bilingual speech processing is the speed at which L2 speakers speak (Kormos, 2006; Viera, 2017). Whereas L1 speech processing is mainly automatic in both the formulation and the articulation, and can therefore run in parallel, L2 speech processing needs attention in both the grammatical and articulatory encoding levels, and as a result, part of the output can only be processed sequentially (Kormos, 2006). In others words, in L1 production, lexical, syntactic, morphological, and phonological encoding is often automatic, whereas these processes are only somewhat automatic in the L2, which makes L2 production slower. This is different to L1 speech production, in which conscious attention and control are normally only needed for conceptualising (Ellis & Yuan, 2005; de Jong et al., 2018), usually making L1 speech smooth and efficient (Kormos, 2006). Thus, L2 speakers have less automatised resources than native speakers. In Levelt's (1999) model, the phonological/phonetic system is located in the second major component (see Figure 2.1, Page 24). Each lemma stored in the mental lexicon entails morpho-phonological codes related to it, allowing the creation of explicit speech (Segalowitz, 2010). For L1 users, this whole process takes place in a highly automatic fashion as the articulatory programmes are stored as whole chunks in the mental lexicon (Levelt, 1989). L2 learners are therefore likely to experience challenges while formulating and articulating, as these processes place high demands on WM. De Bot (1992, 2020), however, argued that L2 beginners, rely heavily on L1 syllable programmes, while advanced L2 speakers are normally able to develop separate chunks for L2 syllables. This is attributed to the fact that less proficient learners' lexical retrieval procedures are still in the process of development and may be incomplete (Viera, 2017). Therefore, L2 speakers can face problems with fluency if they lack automatic access to the L2 syllable programmes.

Another significant difference between L1 and L2 speech processing is that in L2 speech production the effect or influence of L1 on the L2 cannot be excluded (Kormos, 2006). Findings from L2 speech production study suggest that data stores – including conceptual memory, the lexicon and the phonemes store – exhibit similarities between both L1 and L2, which makes L1 and L2 items compete for selection (La Heij, 2005; Liu, 2020; Viera, 2017). L1 influence can play a part in substituting L1 phonemes for similar L2 sounds as speakers often employ L1 rules while phonologically encoding words or phrases (Kormos, 2006). Furthermore, series of articulatory units used to generate the syllables of a given language (called gestural scores) are automatised for uttering L1 phonemes. However, L2 speakers find it demanding to acquire new gestural scores for L2 phonemes (Kormos, 2006). Therefore, learners transfer L1 pronunciation to L2 words as it is easier for articulation; this may also be the result of a limited competence or knowledge of the target language.

Another critical issue is the role of self-monitoring. Both L1 and L2 speakers are almost always monitoring their own speech (Kormos, 2006) as it helps them to detect errors and reformulate the utterance. This is evidenced in Seyfeddinipur et al. (2008), who showed L1 speakers interrupting themselves when noticing a slip of the tongue in their speech. The speakers interrupted themselves as they wanted to ensure accuracy in their speech. It is important to note that the act of interrupting oneself to repair speech happens simultaneously while planning to make the correction (Segalowitz, 2010). That is, speakers plan their correction while continuing to formulate their speech until they are ready to produce it. This reduces any hesitation time that would be required for re-planning if an interruption had taken place (Segalowitz, 2010). However, this is different from interrupting utterances immediately on detecting the error, as this may involve longer hesitation time while the speaker tries to re-plan a correction (Segalowitz, 2010). Kormos (2006) posited that monitoring in L1 and L2 is different, most importantly in the fact that monitoring needs attention and attentional resources are limited. Unlike in L1, L2 speech processing often requires attention at the phase of lexical, syntactic and phonological processing, taking into account that L2 speakers have minimal attention available for monitoring (Kormos, 2006). For this reason, self-monitoring may be more likely to occur and/or be more cognitively challenging in the L2 than in the L1, depending on the speaker's level of proficiency and the situation. It might be for this reason that Lightbown and Spada (2021) argued why fluent speakers do not construct new sentences word by word but, instead, employ strings of words that commonly co-occur (i.e., FSs). An essential aspect of automaticity in language processing involves the rapid retrieval of word meanings (Lightbown & Spada, 2021).

Automaticity plays a vital role in L2 production, and the way it can be developed in the course of L2 learning is an important topic in SLA research. Most L2 learners' ultimate goal is to achieve a high level of fluency.

When studying the psycholinguistic process and cognitive variables of L2 speaking as well as the similarities and differences between L1 and L2 speakers, it is important to clarify the qualities of speech– namely the terms 'fluency', 'accuracy', and 'complexity'.

2.3.5 Fluency, accuracy and complexity

The main aim for speakers is to interact (de Jong et al., 2020). Achieving this aim can, however, be very challenging for L2 learners. According to Bygate (2009), the interaction of linguistic knowledge and skills, cognitive processing and affective factors can pressurize L2 learners, and can significantly affect the quality of speaking, which is often measured in terms of complexity, accuracy and fluency, which is often referred to as the CAF triad. Generally, fluency of speech is defined as a fast and smooth speech that is conveyed coherently without many pauses, repetition and hesitations during real-time oral interaction (de Jong et al., 2020; Lennon, 2000; Segalowitz, 2010, Wood, 2009). Fluency consists of three dimensions: speed, breakdown and repair (Tavakoli & Skehan, 2005). Speed involves the rate of articulation, amount of speech and mean length of runs (number of syllables uttered in the longest stretch with no pauses); breakdown refers to the number and length of silent and non-lexical filled pauses. Repair involves the numbers of speech reformulations, repetitions, self-corrections and false starts (de Jong et al., 2020; Tavakoli & Skehan, 2005; Tavakoli & Uchihara, 2020). Koizumi and In'nami (2013) posited that vocabulary knowledge and fluency may be associated with one another as L2 learners with greater lexical knowledge (e.g., FSs), and faster access to it, can execute vocabulary searches in an easier and faster way. Therefore, learners with reduced vocabulary knowledge may be incapable of finding suitable words, or may spend more time searching for words at the formulation stage, resulting in less fluency (Koizumi & In'nami, 2013). Indeed, Foster and Skehan (2013) argued that fluency benefits when information is familiar and tasks require less planning. One characteristic of the use of FSs is that, because they are used and retrieved as a whole unit, speakers who utilise them seem to speak more fluently (e.g., Boers et al., 2006; Cutler, 2020; McGuire & Larson-Hall, 2018; Tavakoli & Uchihara, 2020; Wood, 2010; Wray, 2002). This is evidenced in the findings of a series of studies by Wood (2009; 2010; 2016), which have consistently demonstrated that practising use of FSs increases fluency. According to Wood (2009), increased use of FSs enabled a participant in his case study to show clear fluency gains in the measures of speech rate and mean length of runs.

In the SLA literature, fluency is generally distinguished from accuracy (Schmidt, 1992). For example, a learner may produce fluent utterances but with grammatical errors or may speak properly but not fluently (Lennon, 1990). Bui and Wong (2021) defined accuracy as the degree to which a learner's performance aligns with the grammatical structure of the second language they are learning. Thus, accuracy is the extent to which the uttered language corresponds to target language norms (Yuan & Ellis, 2003) in terms of form, meaning, function etc. Accuracy is problematic as there are many international varieties of English. Thus, assessors should decide which norms (i.e., as expressed by native speaker of the target language or to non-native practices acceptable in some social contexts/communities) to measure it against (Ellis, 2008). Housen and Kuiken (2009) added accuracy is the ability to produce error-free speech. According to Koizumi and In'nami (2013), the influence of vocabulary knowledge on accuracy may be associated to the ease of lexical searches. This is because learners with larger and deeper vocabulary knowledge can access words and expressions in the lexicon more easily, allowing their limited WMC to focus on other aspects, such as processes in conceptualisation and formulation. Allocating WM processing space to the formulator enables speakers to produce more accurate utterances (Koizumi & In'nami, 2013).

A language learner is also involved in creating messages that consist of lexical items, form and appropriate pronunciation which can increase complexity. Complexity is related to both form and meaning; it is the degree to which speech is 'elaborate' (Ellis, 2003, p.340). According to Housen and Kuiken (2009), it is the most complex, unclear and least understood quality of speaking among the triad (i.e., complexity, accuracy and fluency). Increased language complexity may indicate the progress of L2 interlanguage development (Norris & Ortega, 2009). However, it can also suggest learners' tendency to explore and test new linguistic structures (Bui & Skehan, 2018). Given the limited processing capacity of L2 speakers, L2 users are unable to entirely perform parallel processing and to attend to CAF all at the same time (Bui & Huang, 2018). As such, they must decide which performance area to prioritise and direct their attention to during L2 speaking; however, this is likely to result in diminished accuracy and complexity, creating a trade-off. Several researchers have taken up this matter, arguing that this highlights the benefits of FSs as they alleviate the time pressure associated with planning a novel utterance (e.g., Christiansen & Chater, 2016; Khodadady & Shamsaee, 2012; Wray, 2002). For beginner learners, the formulation and production of L2 speech involves tackling a whole range of phonological, lexical and syntactic problems, and they are likely to have limited WM resources to deal with them all in real time.

Achieving these qualities of speech is a fundamental aim for L2 learners. Therefore, it is important to understand how the CAF triad is typically measured, as outlined in the literature. By exploring these measurements, L2 learners can better work toward achieving proficiency in speech.

2.3.6 Measurement of the CAF triad in speaking

CAF scores have been assessed across different language domains through a variety of tools, ranging from holistic and subjective ratings by a lay or professional judge, to quantifiable evaluations (frequencies, ratios, formulas) of general or particular linguistic features of L2 production (Housen & Kuiken, 2009). Ellis and Barkhuizen (2005) recommended the use of the same measures used in previous research to increase the comparability of results. The most widely used measures in the literature, such as those used by Yuan and Ellis (2003), Ahmadian and Tavakoli (2011) and Ahmadian (2012) to evaluate the quality of the participants' CAF triad, include the following:

1) Fluency measures

a) *Rate A (number of syllables per minute of speech)*: the number of syllables within each narrative, divided by the number of seconds used to fulfil the task and multiplied by 60.
b) *Rate B (number of meaningful syllables per minutes of speech)*: Rate A's formula will be followed again, but excluding all syllables, words and phrases that were repeated, reformulated, or replaced.

It is important to note that several papers referenced in this study, such as Yuan and Ellis (2003), Ahmadian and Tavakoli (2011) and Ahmadian (2012), use the terms *Rate A* and *Rate B* to explain fluency measures. To maintain consistency throughout the paper, I will use *Rate A* and *Rate B*.

2) Accuracy measures:

While accuracy stands out as the most transparent and consistent construct of the CAF triad; however it remains challenging to measure accuracy and identify errors (Housen & Kuiken, 2009). As discussed in Section 2.3.5, accuracy poses challenges due to the existence of numerous international varieties of English. Thus, assessors must determine the norms against which to measure accuracy, whether expressed by native speakers of the target language or conforming to non-native practices acceptable in in some social contexts/communities (Ellis, 2008). Further, even when assessors decide on the norms for measuring accuracy, the identification of errors in L2 data remains challenging. Foster and Wigglesworth (2016, p. 103) illustrate this challenge with Example 1, extracted from an L2 language learner's written text:

a. *Nowadays, we have to accept that education is the best way of people who want to be the good situation in future.*

Corrected examples demonstrate the diversity in rendering this error-free (deleted words are crossed through, and supplied corrections are in bold):

*b. Nowadays, we have to accept that education is the best way ~~of~~ **for** people who want ~~to be in~~ a good situation in **the** future to achieve their goals.*

*c. Nowadays, we have to accept that education is the best way ~~of~~ **for** people ~~who want~~ to be ~~the~~ **in a** good situation in future.*

These examples highlight the various ways to correct a sentence. To address this, Foster and Wigglesworth (2016) recommended basing the analysis on clauses rather than other syntactic units, such as full AS-units (which might contain several clauses), as clauses are shorter and less likely to unfairly disadvantage more complex units that entail multiple clauses.

The most commonly used measure of accuracy in the field, used by researchers, is checking the proportion of error-free clauses (e.g., Ahmadian & Tavakoli, 2011; Awaad & Tavakoli, 2022; Nergis, 2021; Thai & Boers, 2016). This preference is attributed to its dual characteristics: firstly, it serves as a global and generalised measure, and secondly, it is relatively straightforward to use (Skehan et al., 2023). However, Foster and Wigglesworth (2016) argued that an error-free clause may not sufficiently distinguish between different impacts on communication, arguing that not all errors should be treated equally. They proposed a Weighted Clause Ratio (WCR) as a more appropriate measure, classifying errors at different levels based on their impact on communication. However, it is important to note that their criterion for evaluating the impact (or gravity, as they term it) of errors is not purely linguistic in nature, concentrating more on the difficulty of comprehending an utterance— an aspects that falls beyond the scope of this study. In fact, recent research, such as Awaad and

Tavakoli (2022), found that both global measures, namely error free clauses and WCR, not only show highly similar results but also demonstrate a positive and strong correlation. This suggests that the use of either measure could effectively represent the other (Awaad & Tavakoli, 2022), at least at this point of time. Nevertheless, the current group of participants, beginners, is also unlikely to produce errors of the subtlety and nuance that word require the framework of WCR. For these reasons, this study aims to measure accuracy of speech by focusing on error-free clauses, which is a standard measure known for its sensitivity in detecting accuracy across various language proficiency levels (Ellis and Barkhuizen 2005). Error-free clauses are measured in terms of:

Error-free clauses: the percentage of the clauses that are error-free. All syntactic, morphological and lexical errors in speech should be taken into account. Lexical errors are defined in terms of errors in lexical form or phrase, e.g. *I was waiting you* (Yuan & Ellis, 2003, p. 13).

- *Correct verb forms:* The percentage of all verbs that are uttered correctly with regard to tense, aspect, modality and subject–verb agreement.

3) Complexity measures:

Different complexity measures are suggested to be used for different levels of language proficiency. For example, subordination is regarded as a more advantageous metric for learners at the intermediate level (Norris & Ortega, 2009). On the other hand, Norris and Ortega (2009), suggested the use of a coordination measure that was proposed by Bardovi-Harlig (1992) as a useful indicator for learners' syntactic complexification at lower proficiency levels. In their meta-analysis review, Norris and Ortega (2006a) introduced various metrics that could tap into the construct of complexity in L2 production, including:

- *Syntactic complexity (amount of subordination)*: The ratio of clauses to the Analysis of Speech units (AS units) in participants' utterances. An AS unit refers to "a single speaker's utterance" which involves an independent clause or sub-clausal unit, along with any subordinate clause(s) related to it (Foster, Tonkyn & Wigglesworth, 2000, p. 365), which could make them useful units for evaluating spoken language in a study. An example of an AS unit is:

  I could not attend the meeting because I had a prior commitment.

  In this example, the AS unit consists of the main independent clause *I could not attend the meeting* and the subordinate clause introduced by *because: because I had a prior commitment.* Together, they form a single speakers' utterance, representing an AS unit.

- *Syntactic variety:* The total number of different grammatical verb forms uttered in speakers' performance. This study will adopt different kinds of tenses, such as simple present, simple past, present continuous, past continuous, future, etc.

Lexical complexity (variety of word types): The total number of diverse linguistic items that can be identified in the speakers' language performance (Housen et al., 2012). Since the first and second layers of the 1,000 most commonly used words (K1 and K2) are categorised as high-frequency items according to Nation (2013), the third set and beyond 1,000 (K3+) could be regarded as sophisticated/complex items, specifically referring to low-frequency words.

Around the same time, Schmitt and Schmitt (2014) proposed a frequency framework that includes high/mid/low-frequency categories, suggesting a shift in the boundary for high-frequency vocabulary to 3,000 word families. They argued that the first 1,000 word families are prevalent in English, contributing significantly to coverage. The second 1,000 adds a smaller but still valuable amount of coverage, followed by a diminishing contribution from the third 1,000. Beyond the fourth 1,000 families, coverage drops substantially, making high-frequency vocabulary the words occurring before the coverage percentages become so small that frequent occurrence across various texts is unlikely. While it may be true that lexical sophistication can extend beyond the first 3,000 most frequent words, it is important to consider proficiency levels when measuring lexical sophistication in L2 learners. Participants in the current study have not yet acquired the first 2,000 words of the language. Assessing them for proficiency beyond 3,000 words may not yield meaningful insights at this stage. In fact, Vilkaitė-Lozdienė and Schmitt (2019) noted that some studies suggested lowering the threshold for high-frequency words, acknowledging the challenge learners face in learning the first 2,000 most frequent words. Therefore, following Nation's (2013) perspective, words at the third or above 1,000 frequency (K3+) are rated as sophisticated/complex items. A similar decision was also made by Uchihara and Clenton (2018), who applied it to upper intermediate to advanced L2 learners of English at B2 to C1 levels.

2.3.7 Measurement of the CAF triad in FSs

Measuring oral proficiency is particularly complex because of the complexity of speaking ability itself. This is due to the difficulty in defining and assessing the qualities of speaking ability – fluency, accuracy and complexity (Boers et al., 2006). It is important to note that Boers et al. (2006) did not measure CAF of FSs; however they proposed measures to identify a FS in a speakers' real-time discourse concerning fluency and accuracy. To the best of my knowledge, these measures are considered the only ones related to the fluency and accuracy of FSs in the literature.

However, it is important to note that this review will not cover the measurement of the complexity of FSs. This is because there has been no prior examination of the measurement of complexity of FSs within the existing literature. For example, Boers et al.'s study (2006) proposed criteria to help recognising fluent and accurate FSs. However, the assessment of FS complexity, specifically evaluating some as more complex than others such as by using length-based and frequency-based values (Norris & Ortega, 2009), was not taken into consideration.

1) Fluency score of FS

According to McCarthy (2006), fluency vitally includes the ability to retrieve chunks (i.e., FSs). Thus, fluency is measured in terms of hesitations before or after the FS (i.e., breakdown fluency). In this sense, hesitations should appear only in parts of discourse that join the retrieved FS (Boers et al, 2006) and not within the FS itself. For example, in the FS *I am looking forward to meeting you*, hesitation and pauses should be before or after the FS, not within.

2) Accuracy score of FS

Accuracy is measured in terms of linguistic accuracy and appropriate use of FSs. Thus, accuracy of a FS score is obtained when a participant produces a full error-free chunk (Boers et al., 2006).

Sabbah (2015) highlighted the most challenging areas for Arab learners of English as including use of prepositions (since they are limited in Arabic), definite articles (misuse between *a* and *the*) and third person singular -s for verb agreement (as it does not exist in

Arabic, learners omit it). In fact, Qi & Ding (2011) posited that prepositions and articles seems to be challenging to many ESL learners and are likely to be the first to be forgotten under real-time pressure. Therefore, some exceptions can be accepted if subject-verb agreement (third person singular -s), prepositions and articles are uttered incorrectly or are missing. Accuracy is measured in terms of incorrect combinations of words that form the FS in order (e.g., *how you are* instead of *how are you*) or substitution (e.g*., I am frightened the meeting cannot continue* instead of *I am afraid*) of a contextually inappropriate word and an incorrect insertion in a sentence (e.g., *I am looking forward to meeting you* in the middle of the conversation instead of when ending it). In addition, all syntactic, morphological and lexical combinations of words that form the FS should be taken into consideration, for example, uttering *I'm afraid I had to go* to end a conversation will be marked as incorrect because of the tense. A morphological example can be found in Ellis's (1987) study in which he measured accuracy in terms of regular and irregular past tense.

The following section covers further topics that influence speech production, extending beyond the standard scope of CAF.

2.3.8 Factors influencing speech production

This section will provide an in-depth examination of the multifaceted factors that influence speech production, including their operationalisation and the methodologies used to measure their impact within the context of this study. Fluency, closely related to an individual's language proficiency, is typically assessed through various measures, including general language proficiency, vocabulary scores, FSs scores and digit span variables. The utilisation of standardised questionnaires has become important in exploring the mediation of speech production. Therefore, background variables will also be considered in the investigation of factors influencing speech production.

According to Ellis (2008), one of the essential factors in L2 learning is learners' ability to employ linguistic knowledge in varied tasks, known as language proficiency. In this respect, language proficiency includes not only the knowledge of the language but also the ability to use that knowledge effectively in real communication, which directly influences the fluency of speech. Thus, fluency is an important indicator of L2 proficiency (de Jong et al., 2020; Van Vue and Peters, 2020). Language proficiency includes organisational (i.e., grammatical and textual knowledge) and pragmatic competence (functional and sociolinguistic knowledge) (Bachman & Palmer, 2010). In this study, language proficiency is defined as organisational competence. This includes organisational knowledge that involves knowing how to arrange utterances, such as lexical and grammatical knowledge (Bachman & Palmer, 2010).

Moreover, among the numerous factors affecting speaking production, vocabulary knowledge holds a significant place (de Jong et al., 2012). A receptive vocabulary test functions as a simple and fast research tool that provides an indication of the learners' vocabulary size (Laufer & Goldstein, 2004) and an overall picture of the participants' level of proficiency (Read, 2000). However, it should be noted that Kremmel and Schmitt's (2016) study critiqued vocabulary scores elicited from test formats including multiple-choice. Their findings showed that such tests cannot indicate that the vocabulary items are comprehended to a greater degree than only the form-meaning link (see Kremmel & Schmitt, 2016 for a detailed critique), along with the potential impact of guessing. Despite this critique, a receptive vocabulary test utilising a multiple-choice format will be used in this study because it is more suited to the study participants' level of proficiency. Additionally, this type of test is also chosen because it is suitable for the target population, as they are accustomed to a multiple-choice format. Moreover, due to the Coronavirus context and the resultant online method, it will almost be impossible to collect data without using a method the participants were familiar with (i.e., a multiple-choice format test).

However, Webb and Sasao (2013) argued that target vocabulary should be sought in updated BNC/COCA lists that illustrate up-to-date English and provide a better representation of currently used vocabulary. Although the VST has been updated (see Paul Nation's website), according to Nation and Waring (2019), it is not suitable for assessing beginner/intermediate L2 learners, as there will be various unknown words in the test. This is because the updated VST measures up to 20,000 word families. It is for this reason that this study aims to use a test that is similar to the VST but that is more suitable for beginner learners while also based on the more recent BNC/COCA. McLean and Kramer (2015) designed the New Vocabulary Levels Test (NVLT) (recognition test), the aim of which is to measure receptive knowledge of English vocabulary from the first five 1,000-word frequency levels of the BNC and the Academic Word List (AWL) (Coxhead, 2000). In fact, Nation and Waring (2019) have suggested the use of the NVLT rather than the VST as a better means to assess beginner/intermediate learners. In this study, participants' vocabulary knowledge will be evaluated using both the NVLT and FS knowledge tests. This approach aims to determine whether vocabulary size or FS proficiency is a better predictor of complexity, accuracy, and fluency of speech. The rationale for combining these tests is supported by Nergis's findings, which showed that vocabulary instruction significantly increased oral complexity scores compared to FS instruction. As discussed previously in Section 2.4.3 findings included that vocabulary instruction significantly increased oral complexity scores when compared to FS instruction. However, it should be noted that while in Nergis (2021) study focus was given to academic vocabulary this study will use the NVLT- the first five 1,000-word frequency levels of the BNC.

Furthermore, standardised questionnaires have become indispensable tools for exploring the mediation of speech production. Anderson et al., (2018) emphasised the importance for language researchers to use an evidence-based instrument with high reliability and validity to detect relevant aspects of the participant's multifaceted language experiences, influenced by social, personal and contextual factors.

Closed questions, known for their ease of response compared to open-ended ones (Bryman, 2004), promote the occurrence of association of responses, as suggested by Bryman (2004) and Mackey & Gass (2005). However, Nunan (1992) argued that open-ended questions are also essential in the construction of a questionnaire, as they allow the researcher to obtain further information. Respondents may provide their own perceptions and unexpected answers may lead to different aspects of investigation (Bryman, 2004, Bryman, 2012). For this reason, the questionnaire this study will apply (as will be discussed further in the next chapter) includes some open-ended questions for participants to express their ideas freely without the restriction of selecting options, therefore providing "a far greater richness than fully quantitative data" (Dörnyei, 2003, p. 47).

2.4 Approaches to the teaching of L2 speaking

While the utilisation of FSs among L2 learners is associated with improved fluency (McGuire & Larson-Hall, 2018), effective teaching methods for integrating these sequences into one's language use have not been distinctly identified, or only a few studies have addressed this issue (Cutler, 2020; McGuire & Larson-Hall, 2018). To achieve success in speaking, it is essential that teachers direct learners thoroughly by introducing tasks that are related and sequenced in order to enable learners to raise their awareness of knowledge, skills and strategies (Burns, 2019; Hughes, 2010). Such classes can be very successful, as they target students' oral presentations, whole-class or pair discussions, and can also engage vocabulary, listening or reading tasks (Mystkowska-Wiertelak et al., 2017). In what follows, I begin by presenting a theoretical framework for teaching L2 speaking by reviewing key theoretical approaches, which in turn forms the pedagogical foundation recommended by this study—namely ACCESS (Gatbonton & Segalowitz, 2005). Drawing on the relevant literature, I also provide number of methodologies/strategies that have received theoretical and empirical attention, such as pre-task planning, repetition, memorisation and performance, all applicable to the L2 classroom. An in-depth exploration of each of these learning methodologies/strategies will be presented, highlighting their impact on language production, with a specific focus on complexity, accuracy and fluency. Additionally, I draw connections to cognitive variables, particularly WM. The aim in this section is to help the reader understand why these learning strategies are of particular importance in SLA research and how they play a pivotal role in facilitating the learning of L2 oral speech and how they interrelate with individual learners' variables, such as WMC and limited processing capacity.

2.4.1 Theoretical underpinning of teaching speaking

The Audiolingual approach is an oral-based method that was developed in the 20th century; speaking is the focus of this method (Ellis & Shintani, 2013). As pointed out in Larsen-Freeman (2000) and Richards and Rogers (2014), the approach has a strong theoretical basis in structuralism (a theory of language—Saussure, 1916) and behaviourism (Skinner, 1957); and was later supported by behaviourist learning theory (Ellis & Shintani, 2013).

In the behaviourist view, language is viewed as a system of habits, based on the notion of stimulus (input), response (output) and reinforcement (positive feedback) (Mitchell et al., 2019; Lightbown & Spada, 2021). The response learners produce to these stimuli in the classroom will be reinforced if correct (Mitchell et al., 2019). For example, in the context of meeting someone, if a learner greets successfully, the response (learner's output) will be reinforced by the teacher by providing positive feedback. Otherwise, if the response was unsuccessful, it would not be reinforced and the learner will hopefully avoid using it (Ellis &

Shintani, 2013). Habituation is the result of reinforcement leading to a stimulus and response pairing becoming strengthened through successful use and application (Mitchell et al., 2013). In other words, via repeated reinforcement, a response can be used successfully in a specific context, leading to habit building.

From a teaching perspective, the behaviourist theory is based on the view that effective learning could be achieved by the imitation, repetition and practice of the same utterance repeatedly (Mitchell et al., 2019; Richards, 2015) in addition to memorising dialogues so that the utterance can become a habit (Ellis & Shintani, 2013; Suzuki, 2023, Mayne, 2023). Thus, many drilling tasks will be deployed for learners to produce correct utterances (Mitchell et al., 2019). Drilling tasks typically involve repetitive exercises where learners practice specific language elements, such as vocabulary, grammar rules, pronunciation or sentence structures. The underlying idea behind drilling tasks in behaviourism is that repeated practice and reinforcement lead to the strengthening of the desired language behaviours, making them more automatic and accessible for learners.

Behaviourism was heavily criticised in the literature (e.g., see Chomsky, 1959) as it has also been claimed that basing language learning on a behaviourist learning theory is "highly limiting" (Ellis & Shintani, 2013, p.39). According to Bygate (2001,2016) and Ellis and Shintani (2013), behaviourism failed to show the relationship between language and meaning and neglected the importance of the learner's own contribution in real-life interactions. This is because stimulus is only given by the teacher to the learner, while in return the learner produces a response to the teacher. Thus, learners are left without the opportunity of using language in real-life interaction. Further, Ellis and Shintani (2013, p.39) argued that L2 learning is a "slow, organic process with errors". Thus, avoiding errors does not facilitate the process of learning as learners need to test their utterances (Ellis & Shintani, 2013).

Despite Ellis and Shintani's (2013) criticism of the behaviourist learning theory, the researchers highlight the effectiveness of its inclusion of memorisation and reinforcement; stating that such methods can contribute to learning and indeed, are still used. The researchers continued, stating that memorisation is viewed as an effective learning strategy that language learners can benefit from (Ellis & Shintani, 2013). Such benefits include the learning of L2 vocabulary, as memorisation enables learners to focus their attention on meaning (Ellis & Shintani, 2013). In fact, the behaviourist learning theory can be effective in facilitating the learning of formulaic expressions as it targets the teaching of patterns (i.e., the underlying structure of particular utterances) (Ellis & Shintani, 2013). On the other hand, second language acquisition (SLA) research advocates "reinforcement" (corrective feedback) especially in vocabulary learning (Lightbown, 2008). Indeed, in a study conducted by Lynch and Maclean (2000, 2001) in the context of English for specific purposes found that learners increased their speaking skill when they were given feedback from teachers; of course, improvements varied from one learner to another. Thus, corrective feedback is necessary to achieve successful interaction (Mystkowska-Wiertelak et al, 2017; Nation, 2007).

The inadequacy of behaviourism led to many developments in L2 learning. Communicative Language Teaching (CLT) marked the beginning of a major theoretical shift within language learning in the 20th century (Ellis & Shintani, 2013; Richards & Rogers 2014). CLT refers to tasks ranging from role-playing to games, where communication is taking place (Gatbonton & Segalowitz, 2005) rather than only practising correct usage (as in the behaviourist learning theory) (Ellis & Shintani, 2013).

A number of CLT classrooms follows a specific methodology, namely PPP (present-practice-produce) (Lindsay & Knight, 2006). PPP consists of three stages: (1) the presentation of target aspects through explicit instruction, (2) the practice of the learnt features through production tasks, and (3) producing and presenting what has been learnt through tasks (Ellis & Shintani, 2013). The overall goal is to allow learners to learn the L2 aspects deeply so that they will be able to produce them successfully without hesitation (to develop their implicit knowledge) (Ellis & Shintani, 2013). Goh and Burns (2012) and Burns (2019) put it another way—to enable learners to transfer their speaking ability developed via such communicative tasks to real life contexts.

PPP can be viewed as a development of Audiolingualism (Harmer, 2007). One of the reasons Audionlingualism was criticised was because controlled drills fail to achieve real-life communications (Ellis & Shintani, 2013). CLT, however, focused on the importance of communicative tasks (Ellis & Shintani, 2013). PPP is similar to Audiolingualism in terms of the types of controlled production tasks, but it also integrates the types of communicative tasks (discussions, role-plays, simulations, project work and information-gap tasks) supported in CLT at the production stage (Ellis & Shintani, 2013). PPP also includes explicit instruction, a feature that is restricted in behaviourism (Ellis & Shintani, 2013). In fact, Hinkel (2020) claimed that L2 classrooms that avoid the use of explicit instructions can lead to learners having major difficulties in reaching a high level of proficiency and developing syntactic and lexical accuracy in speaking. This is because meaning-focused instruction facilitates the development of skills including communication and oral fluency (Lightbown & Spada, 2021). Ellis & Shintani (2013, p. 44) claimed that theoretically, CLT is "far more successful" in its learning theory than are the principles in behaviourism. However, the researchers continued that, there is no clear evidence indicating the superiority of one method over the other (Ellis & Shintani, 2013). Nonetheless, within the domain of language teaching/learning, the implication goes beyond claiming which learning theory is more beneficial than the other as language learning involves more than one method and approach––including extensive practice, awareness and noticing. In fact, the success of the use of repetition and imitation is evidenced in Ghazi-Saidi and Ansaldo's (2017) neuroimaging study that was conducted on 12 Spanish-speaking (L1) participants and 12 Persian-speaking (L1) adults of French (L2). Participants were introduced to 130 new French vocabulary via audio-visual repetition and imitation database. Instructions were given to learners to examine each picture and imitate its name as closely as possible to the native model. Repetition was employed until the learners learnt the pronunciation of the word. Learners practised for 15 minutes, over 30 days, and were measured on accuracy and speed. Results indicated significant improvements in the effectiveness of repetition and imitation in L2 vocabulary learning. Repetition allowed learners to respond faster and more accurate, reflecting high level of automaticity (Bygate, 2018; Segalowitz & Hulstijn, 2005) which is indicative of successful learning, as noted by Segalowitz and Frenkiel-Fishman (2005). Indeed, Marinis and Armon-Lotem (2015, p.4) argued that the imitation of behaviour, including language, is an innate ability of humans and other species, and the repetition of sounds, words and sentences involves both uninstructed and instructed teaching and learning methods. As a result, repetitive practice (originates from behaviourism), which is viewed as unfavourable in CLT (Gatbonton & Segalowitz, 2005); is still employed and advocated by SLA research in the field (Ansaldo et al., 2017; Bygate et al., 2013; Bygate, 2018; de Jong and Perfetti, 2011; Gatbonton & Segalowitz, 2005; Horst, 2013; Jeong & De Keyser, 2023;). Nevertheless, behaviourism and CLT approaches seem to complement each other rather than stand in opposition (Richards, 2015); while the former concentrates on developing learners' communicative ability, the latter concentrates on putting this ability into practice. Thus, integrating aspects of both approaches is more advantageous than focusing on one approach.

For this reason, this study will follow an utterance-based approach that combines some of the features of both behaviourism and CLT—namely ACCESS (Gatbonton & Segalowitz, 2005).

2.4.2 The ACCESS methodology

Gatbonton and Segalowitz (2005) proposed an L2 learning methodology targeting speaking, initiated from a background in CLT, namely Automatisation in Communicative Contexts of Essential Speech Segments (ACCESS). This approach advocates automatisation through the employment of communicative activities. *Essential Speech Segments* refers to the target set of utterances that learners can leave class with (Gatbonton & Segalowitz, 2005). ACCESS aims that these essential speech segments are produced and practised in genuinely interactive communication. "Genuinely communicative" is defined as a task that consists of two or more participants working together to complete it by exchanging knowledge, in order to fill gaps in knowledge possessed by one and not the other (Gatbonton & Segalowitz, 2005). This means new information must pass from one speaker to another, filling the knowledge gap, and the requested information must be essential for the assigned task to be continued. Exchanging knowledge in an interactive communication is important because learners may learn additional information (declarative knowledge) from a teacher or peer, then put this knowledge into practice (procedural) for automatization to develop (Jeong and DeKeyser, 2023; O'Malley & Chamot, 1990). Such tasks give learners the opportunity to interact so that they can negotiate meaning, ask for explanation, give suggestions and most importantly use a wide range of language skills, facilitating L2 (Bygate, 2016, 2018; DiGiovanni et al., 2001). In other words, group work generates a communicative environment allowing peers to negotiate meaning, which promotes comprehension and therefore acquisition. In their approach, they do not only suggest that automatisation can work through communicative tasks, but also explain how other methods of supporting learning (e.g., explicit instructions of and repetition/practice of forms and formulaic utterances) can be embedded within communicative tasks.

Figure 2.3 demonstrates that ACCESS lessons consist of three stages—a Creative Automatisation Phase, a Language Consolidation Phase, and a Free Communication Phase. Each stage leads to the other, starting from The Creative Automatisation Phase; however, with the Creative Automatisation and Language Consolidation phases are able to be swapped around.

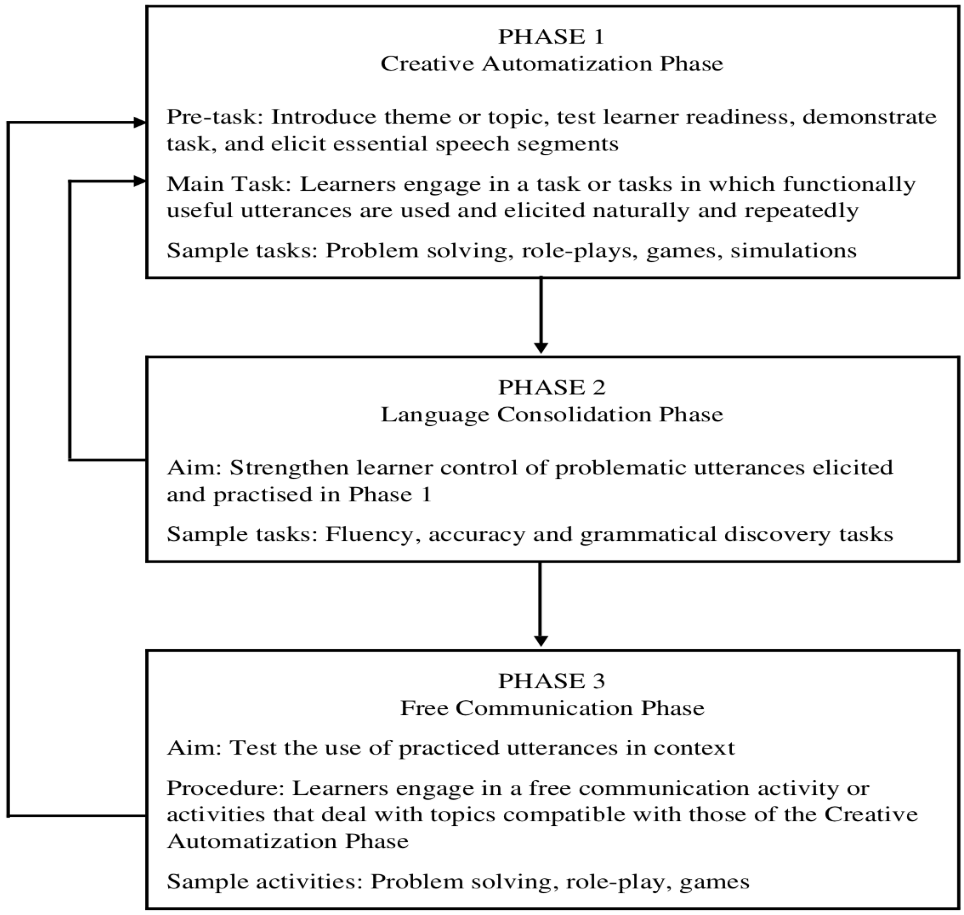


Figure 2.3 A plan of the ACCESS methodology, demonstrating the three stages in order

The first phase, The Creative Automatisation Phase, is aimed to be both communicative and capable of facilitating automaticity of utterances. It involves two parts – a Pre-Task and a Main Task. The Pre-Task is diagnostic to ensure that learners have enough linguistic knowledge (e.g., FSs) to start the task; the second is pedagogic and helps students learn the necessary expressions to continue with the task. The Main Task involves communicative activities where learners work collaboratively in pairs or in groups to achieve a goal; that is, to complete three stages in this task—arrange themselves into roles (role-play); gather information from the other group about how they assigned themselves into roles (interview) and finally, the students present the results of their interviews to the class. In return, the interviewed group gives feedback on whether the information is correct or not (presentation) (Gatbonton & Segalowitz, 2005). Therefore, the Creative Automatisation Phase encourages learning, use and practice of the important speech utterances through genuine communication, repetition and use of formulaic sequences. Learners are constantly utilising and hearing the same essential speech segments during the three stages, a process (repetition) that supports automatisation by achieving cognitively faster ways of processing information (DeKeyser, 2018); resulting in more efficient and more accurate production of speech (Gatbonton & Segalowitz, 2005). Dao and colleagues (2017) investigated the effectiveness of the ACCESS methodology in enhancing oral production efficiency and accuracy among thirty-three Chinese learners over a four-week classroom study. The study analysed various aspects, including focus, initiation, response and turn length during ACCESS task interactions. Findings revealed that learners consistently emphasised the form, specifically the past tense verb (as it was the target linguistic item), through repeated use of certain phrases, resulting in significant improvements in accuracy. The researchers argued that ACCESS tasks, designed for repetition and target achievement, may facilitate the development of fluent and accurate oral communication. However, it is worth noting that the study did not conduct post-tests to asses learners' form acquisition, highlighting the fact that tasks followed by ACCESS principles may not necessarily result in the acquisition of the target linguistic item and form.

In fact, the repetition of sentences (or phrases) in a task is known for its effectiveness in L2 learning (de Jong and Perfetti, 2011; Marinis and Armon-Lotem, 2015). A sentence repetition task (SRep) refers to listening to sentences and repeating them accurately (Marinis and Armon-Lotem, 2015). SRep tasks involve language processing at all levels (lexical, phonological, morphosyntactic, semantic) associated with comprehension and production, in addition to the ability to store and retrieve the L2 from memory (Marinis and Armon-Lotem, 2015). Gatbonton and Segalowitz (2005) provided an example of a commonly used communicative task that has one of its major components as the use of repetition; namely, "Find someone who….". Students have to interact with each other according to instructions that require the employment of a lexical expression followed by a gap, for example, "Have you ever……?" and record the information on paper. In this case, learners are involving themselves in a task in order to put their declarative knowledge (explicit knowledge, hence with awareness) into practice (Mitchell et al., 2019). Formulaic utterances are repeated via speaking tasks with different partners and/or to a time-limit, with the aim of encouraging 'chunking', rehearsing and memorising before performing dialogues that involves formulaic expressions.

However, in order for proceduralised knowledge to be used correctly and efficiently, extensive practice has to take place (Gatbonton & Segalowitz, 2005) in order to reduce the time needed to perform a skill, the error rate, the amount of attention needed and the degree of interference between other tasks and the skill itself (DeKeyser, 2015a). In fact, both Wray (2002) and Van Vu & Peters (2022) argued that unless the learned FSs are encountered and used frequently, the learner will not be able to recall them.

Repetition and practice is essential as the utterances conveyed and heard become integrated in the learner's memory, and eventually will lead to automaticity in both reception and production (DeKeyser, 2015, 2018; Gatbonton & Segalowitz, 2005; Jeong & De Keyser, 2023). In Gatbonton's and Segalowitz's (2005) approach, ACCESS, automatisation is promoted by the practice of required speech utterances through interaction, repetition and use of formulaic expressions in a given task. In fact, in a study conducted by de Jong and Perfetti (2011) found speaking about the same topic three times resulted in a better speech compared to a control group that discussed three different topics. The researchers illustrated the advantages were transferable to the next topic (de Jong & Perfetti, 2011). In addition, the study revealed that repeated use of specific vocabulary and sentence structures resulted to the automatisation of knowledge (de Jong & Perfetti, 2011). In other words, increased repetition paves the way to automaticity, which enables faster and more accurate processing. This reduces cognitive load (Leonard et al., 2011; Mitchell et al., 2019). However, according to Mitchell and colleagues (2019), automatised knowledge is context-specific, thus learners may not be able to use this knowledge in different contexts to those in which the skill was learnt. This means that if learners already have specific linguistic knowledge they may not be able to utilise it in a new context unless it has been practised in that context. Therefore, learners should be taught to utilise FSs in different contexts, fulfilling the demands of the genre of which they occur in. Given that FSs are useful in different contexts, such as giving commands, making requests, a range of politeness features and hedges (Wray, 2002), such phrases can be employed for daily life and can be transferred to match any context (e.g. business).

The second phase, Language Consolidation, allows the teacher to focus attention on certain utterances deployed in the former phase in order to improve fluency (to encourage smooth and rapid production) and grammatical knowledge (e.g., to notice the utterances with possession features used earlier, such as, "Elly is the wife of John vs Elly is John's wife") (Gatbonton & Segalowitz, 2005, p.334). Long (1996) argued that feedback involving the reformulation of a learner's own utterance can help with noticing gaps between the learner's own L2 output and L2 target forms, in addition to overcoming features of L1 influence. In fact, a consensus is found in the literature regarding L2 classrooms; those that promote noticing benefit consolidation more than those that do not (Lyster & Saito, 2013).

Finally, the Free Communication Phase aims to encourage learners to use the essential speech segments (talk about the topic) freely and to express further ideas, such as through discussing and comparing. Learners are given the opportunity to use large number of formulaic sequences productively, which therefore will help them to improve their memories and enrich their FS learning successfully (Craik & Lockhart, 1972; Laufer & Hulstijn, 2001; Folse & Briggs, 2004; Thornbury, 2019). In fact, Nation (2007, p. 310) stated that a new item should be covered many times so that learners would recycle their lexicon knowledge. Otherwise, the learning would act "as a burden". More recently, Pellicer-Sánchez & Boers (2019) considered that repeated encounters with a lexical item initially learned not only strengthen retention but also mediate familiarity with its phraseological behaviour, such as its typical use in context. Repetition or repeated encounters is particularly beneficial for learners with very little or no prior English instruction, as it can facilitate in automatising lower-level language aspects including lexical access and phonological processing, in order to allow memory resources to focus on processing and comprehending multi-morphemic units as a whole. Therefore, Hall (2010) posited the need for an abundant experience with language input to create a basic store of language forms, while Van Vu and Peters (2023) specifically advocated for the development of formulaic competence**.** This phase allows the learner to use input productively in order to practise and build up automatised knowledge.

In short, ACCESS aims to introduce learners to short FSs in order to practise using them before participating in an interactive task that involves the repeated employment of these FSs with the aim of achieving a communicative purpose.

The next section investigates effective strategies that could facilitate and enhance L2 speaking production.

2.4.3 Pre-task planning

L2 users practise their language use in different types of communications as they interact in L2 tasks or contexts (Bachman & Palmer, 2010). An important aspect in the employment of tasks that enhance language use and one which has obtained significant theoretical and empirical consideration is the planning processes that students engage in, together with the development of these tasks (Lambert, Kormos & Minn, 2017; see Pang & Skehan, 2014). The process of planning involves exposing learners to new language, activating language, recycling language, facilitating processing load and encouraging learners to decode the task in more complex ways (Skehan, 1998). Ellis (2009) identified three kinds of planning: rehearsal, within-task planning and pre-task planning. Rehearsal provides learners chances for pre-task practice, enabling them to perform the whole task before the actual performance, thereby enhancing their performance (Ellis, 2009; Stroud, 2021). Within-task planning, or online planning, occurs when enough time is given during speech production (Yuan & Ellis, 2003). Within-task planning may take two forms: it can be time pressured, where students are asked to present the task within a specific time limit, or unpressured, with no time restrictions set for task performance (Ellis, 2009). Lastly, pre-task planning (or strategic planning) can be employed in the pedagogic process in the sense that before real task performance, learners may be given time to plan the language (Wang et al., 2019) either in the L1 or L2 (Ellis, 2005). According to Ellis (2005), even the most messy and hurried speech or writing entails a certain an amount of planning.

Planning is essential to language production (Ahmadian & Tavakoli, 2011; Garbati & Mady, 2015; Stroud, 2021). Research have revealed positive effects of planning on different dimensions of speech performance–fluency, accuracy and complexity in different L2 contexts, such as in Brazil (Guará-Tavares, 2013), China (Yuan & Ellis, 2003) and Japan

(Stroud, 2021). It seems reasonable to argue that since L1 speech production is greatly automatised and this high degree of automatisation does not apply to the L2 (see Section 2.3.2), L2 learners may crucially need to structure plans for communicative situations. L2 learners may use planning time to access and activate prefabricated chunks which requires high degrees of cognitive processing (Mehnert, 1998). This process can help learners to alleviate time constraints during both the stages of speech conceptualisation or linguistic formulation (Wang et al., 2019) and the speech articulation stage (Stroud, 2021). This is clearly evident in Ahmadian and Tavakoli's (2011) study on 60 intermediate-level (according to placement tests including Oxford Placement Test) female Iranian learners. Their aim was to investigate the influences of simultaneous practice of careful online planning (given time to reflect how to formulate the next part of one's utterance in a speaking task) and repetition of tasks (i.e., careful online planning without task repetition, pressured online planning including task repetition, careful online planning with task repetition and pressured online planning without task repetition) on complexity, accuracy and fluency in the speaking production of L2 English learners. Findings showed that the opportunity to engage simultaneously in careful online planning and task repetition improved CAF significantly. In this respect, planning seems to provide a particular benefit to L2 speakers in the sense that they can plan their lexical and syntactic selection. The pressure on the limited processing capacity in either speech conceptualisation or linguistic formulation will therefore be considerably relieved and the requirement for L2 learners to pause will consequently decrease (Bui & Huang, 2018; Wang et al., 2019). Therefore, speakers engage in what Kormos (2006, p. 123) referred to as "covert repair" and cautiously plan their speech online. The findings of Ahmadian and Tavakoli's (2011) study reiterate Yuan and Ellis's (2003) and Ellis and Yuan's (2005) results in that, accordingly, when speakers are under time pressure, they are mainly engaged with how to express their message and attempting sophisticated language is not such a priority for them, as it overloads their WM (Ahmadian & Tavakoli, 2011). On the other hand, when plenty of time is provided, they can also access grammatical information (Yuan & Ellis, 2003). This interpretation suggests that supportive online planning provides enhanced opportunities for both monitoring and speech formulation (Wang et al., 2019). Indeed, studies on online planning consistently report that unpressured time conditions are usually associated with higher accuracy in learners' speech performance, with occasional increases in complexity (Ahmadian and Tavakoli 2011; Ellis and Yuan 2005; Hsu 2017; Wang et al., 2019; Yuan and Ellis 2003). In other words, giving learners ample time allows them to thoroughly pay attention to their performance and henceforth formulate grammatically accurate and perhaps complex utterances. Otherwise, they would sacrifice form (i.e., accuracy and/or complexity) over meaning (i.e., fluency); trade-off effects between accomplishing fluency, accuracy and complexity (see Section 2.3.2).

In fact, learners will also have the time to correct their inaccurate utterances in the formulation phase preceding articulation (Levelt, 1989). Indeed, some studies indicated that attentional resources for the Formulator and Articulator can be increased by planning and task repetition (Bygate, 2001; Lynch & Maclean, 2001; Mehnert, 1998). Tavakoli and Foster (2011) noted that this is possibly because the preliminary performance of the task, or planning time prior to the task, delve into the Conceptualiser's attentional needs which therefore provides more capacity for the articulation and encoding of linguistic material. Therefore, taking care of the Conceptualiser's attentional needs via planning gives more space (i.e., larger attentional capacity) for the encoding of L2 forms (i.e., grammatical accuracy). In other words, pre-task planning is likely to improve CAF because L2 speakers can conceptualise their performance beforehand. In fact, Ellis (2009) and Stroud (2021) stressed that planning may also have some effect on the Formulator and Articulator because learners are likely to have accessed relevant linguistic knowledge through planning, and will find it simpler to access them again during the formulation and articulation phase. Similarly, some scholars have emphasised that learners should be given sufficient time to plan prior to completing a task that targets improving fluency in oral production (Bui & Huang, 2018; Lambert, Kormos & Minn, 2017; Nation, 2011; Segalowitz & Hulstijn, 2005) on the grounds that there is significant consensus that complexity and fluency (Skehan & Foster, 2001) and accuracy are progressed by pre-task planning (Hsu, 2017; Yuan & Ellis, 2003).

Researchers have been exploring what constitutes a sufficient amount of planning time prior to the task. Wang et al (2019) noted that various durations of planning time are used, with the most common being ten minutes. For example, Hsu (2017) provided intermediate to advanced L2 learners a ten-minute pre-writing planning time, while Bui and Huang's (2018) also employed a ten-minute planning prior to presentations for participants at the upper-intermediate proficiency level. However, the determination of what constitutes a sufficient amount of planning time prior to the task may also vary depending on aspects such as research aims and the proficiency levels of the learners. For example, Skehan (2018) argued that at least for intermediate level, increased time does not seem to result in a higher level of performance. Previous research, such as Mochizuki & Ortega's (2008) investigation with beginning level participants and Tavakoli and Skehan's (2005) study involving both high beginner and intermediate levels participants were given five minutes of planning time. Given that learners have a limited and selective attention capacity (e.g., Kormos, 2006; Bygate, 2018; Bui & Skehan, 2018; Mitchell et al., 2019; Schmidt, 2001; Viera, 2017), Skehan and Foster (2001) emphasised when learners are provided time to plan their speaking performance they transfer into WM components from long-term memory identified to be relevant to the given task. In this way, planning assists learners in overcoming limitations in WM and enhancing performance (Ellis, 2005). However, Ortega (2005) argued the advantages of planning for performance may also rely on the capability to use what was planned and apply it to online performance. That is to say, WM plays a vital role in retrieving this information (see Section 2.3.3). Learners with better capacity seem to retrieve information (from planning) more successfully during the performance of demanding cognitive tasks, achieving greater fluency (Guará-Tavares, 2013; Stroud, 2021) and accuracy (de Paula, 2021). It might be for this reason that Guará-Tavares (2013) aimed to investigate whether there is a relationship between WMC and L2 speech performance (in terms of CAF) in unplanned and planned (i.e., 10 minutes to plan the speaking task prior to performance) conditions in 50 undergraduate intermediate-level learners. Findings demonstrated that there is a significant correlation between WMC, fluency and complexity in planned speech, but not with accuracy. In contrast, Ahmadian and Tavakoli's (2011) findings provided further evidence confirming the limited and selective facets of attentional capacity, such as participants who have spent more time for task completion have developed more accurate speech than those who have presented the task under time pressure (i.e., pressured online planning). More recently, de Paula et al. (2021) conducted a study involving twenty-eight intermediate undergraduate university students in Brazil. They wanted to investigate which group, individuals with lower or higher WMC, incorporated more new information during task repetition. In this study, task repetition served as a form of planning (rehearsal) that participants could recalled and apply in the second encounter of the same task. The findings revealed that participants in the higher WMC group demonstrated significantly higher percentages of new lexical items compared to those in the lower WMC group. The researchers proposed that these findings showed gains in both fluency and accuracy. Stroud (2021) examined 24 Japanese university students to investigate the impact of planning on oral performance. Findings revealed significant effects on performance for both rehearsal and strategic planning, especially among students who spoke less frequently than others during discussion, implying a positive influence on speech fluency and participation. Kawauchi's

(2005) findings included that planning had enhanced lower-proficiency learners' use of regular past tense (i.e., accuracy). Such finding indicates that planning is of considerable importance when learners are still encountering difficulty in accessing and operating their linguistic resources.

Interestingly, different studies seem to have mixed results regarding the effects of pre-task planning on accuracy. However, following Skehan (1996), Michel (2017) argued that trade-off effects are likely to be expected; for instance, accuracy in speech increases because of monitoring during online planning time, which may trade off with fluency. This trade-off effect is evidenced in Yuan and Ellis (2003). They compared Chinese university learners (in a non-immersion context, i.e., in China) carrying out a task under no-planning conditions, where they were given limited time to fulfil the task, with both those who were given planning conditions (strategic planning; 10 minutes of planning time was given) and the online planners (who were given unlimited time to formulate and monitor their speaking production as they carried out the task) (Yuan & Ellis, 2003). Findings suggested that for complexity (in terms of syntactical complexity but not syntactical variety or lexical range) and accuracy (in terms of more correct clauses and correct verbs), the planning-group showed greater effects. Interestingly, for fluency, findings indicated no statistically significant difference in all groups, although the pre-task planning group was more fluent than the online planners. This trade-off effects was also evident in Nergis's (2021) study, which showed that learners in the FS group outperformed those in the academic vocabulary (AcaVoc) group in terms of L2 oral fluency. However, FS instruction did not prove more effective than AcaVoc instruction in enhancing accuracy or complexity of speech. In contrast, AcaVoc instruction significantly increased oral complexity scores when compared to FS instruction. The researcher justified this by explaining that the academic instruction in the study introduced a substantial number of new verbs, which led to the AcaVoc group's superior performance in terms of syntactic variety. Thus the acquisition of an extensive range of AcaVoc facilitated the academic vocabulary group to excel in adding lexical variety (i.e., complexity) to their speech. Consequently, accuracy was the only aspect that was not affected by the type of instruction which can be attributed to the trade-off effect. Thus, fluency seems to contrasts with accuracy and complexity as demonstrated in both the studies of Yuan & Ellis (2003) and Nergis (2021).

Moreover, the impact of planning on fluency was demonstrated in Bui and Haung's (2018) study, which involved fifty-eight undergraduate students from a nursing school and computer science department, all assessed at an upper-intermediate proficiency level (equivalent to CEFR B2). Specifically, their research aimed to explore the influence of two key factors, pre-task planning and content familiarity, on L2 fluency. The results of Bui and Huang's (2018) study showed that planning helped to overcome the negative effects generated by unfamiliar topics and significantly reduced the dysfluency caused by an unfamiliar topic. According to the researchers, planning decreased the number of pauses, the average length of pauses and the total silence time in the middle of clauses as well as at the ends of independent clauses. It also influenced repairs such as decreasing the number of false starts, reformulations and repetitions (Bui & Huang, 2018).

However, it is important to note that the participants in Bui and Huang's (2018) study were instructed to produce a minimum of 10 sentences, with no specified upper limit on speech length. The researchers provided this rationale for not specifying a specific time limit on participants speech sample, with the goal of facilitating meaningful analyses, although they did not elaborate further. This differs from the methodology taken by Uchihara and Clenton (2018), who investigated intermediate and advanced learners. Their study featured speech samples ranging from one to four minutes and a total word production range of 81 to

319, which is comparable to other L2 speech studies measuring lexical performance (e.g., *Range* = 61 to 268 words in Saito, 2018).

However, it remains to be seen if these results can be generalised, as findings have been more varied in this respect, particularly regarding accuracy. One possible explanation could be methodological, such as variations in planning time or speech length. However, the inconsistency of findings could be attributed to different factors.

In short, while L1 speaking production is highly automatised, this is not the case for L2 speaking production as learners'

(a) L2 knowledge is still developing,

(b) L2 is more hesitant, consists of shorter utterances and contains more errors,

(c) L1 may alter with L2 (Kormos, 2006).

Thus, task planning can encourage learners to develop readiness for the actual task via explicit preparations, which therefore result in useful effects on the fluency, complexity and accuracy of L2 performance as shown in previous studies (e.g., Ahmadian and Tavakoli, 2011; Bui and Huang, 2018; Stroud, 2021). Planning lessens the load on WM and on the limited processing capacity as it allows learners to thoroughly pay attention to their performance rather than having to attend to all aspects of oral performance at the same time (Yuan & Ellis, 2003).

However, the opportunity to plan before speaking is not always available in everyday life. Another pedagogical methodology to facilitate L2 learners' performance of oral tasks (e.g., use of FSs) is through the familiarity with the content, the structure, the kind of task (or having formerly performed the task), which also prepares students for performance (Bui, 2014). Learners might be given the chance to practise their performance of the task (Bygate, 2018; de Paula et al., 2021; Ellis, 2005) by repeating the exact same task or perform a similar version of the task with somewhat different content (Bygate, 2001, 2018) – a point that will be followed up in the next section.

2.4.4 Repetition and practice

The learning and retention of vocabulary (i.e., FSs) can be facilitated through the implementation of strategies. Mitchell and colleagues (2019) defined learning strategies as methods used by the learner to make their L2 learning as successful as possible. One successful strategy/method could be through repetition, which is exactly the process of the various years of L1 acquisition, turning most adults into fluent speakers of their mother tongue (Segalowitz & Hulstijn, 2005). Nation (2007, p. 310) mentioned that a new word should be covered many times in a textbook so that learners would recycle their lexicon knowledge, otherwise the learning would act "as a burden". Put simply, as the learner practises components of the new ability by repetition, it becomes habitual or chunked, making these components to be employed extremely rapidly, and effectively while being unavailable to conscious awareness (Segalowitz & Hulstijn, 2005). Similar to behaviourism, Thomson (2020) noted that learning is associated with repetition.

Recent literature has demonstrated that rehearsal or repetition of tasks (task repetition) can progress different performances including L2 fluency (Garbati and Mady, 2015; Thai & Boers, 2016), complexity and accuracy (Thai & Boers, 2016). Task repetition is a type of planning (de Paula et al., 2021; Ellis, 2005, 2008). It refers to repeating the same or slightly modified task, "whether the whole tasks, or parts of a task" (Bygate & Samuda, 2005, p. 43; Bygate 2015, 2018). To begin with, repetition enables learners to involve themselves in tasks in order to put their declarative knowledge into practice (Mitchell et al., 2019). Indeed, Galian-Lopez (2018) stressed the use of repetition and practice, as automatisation can then be achieved. Repeating specific utterances to convey a certain function facilitates the retention of these utterances in memory and enables them to be easy retrievable for future employment (de Jong and Perfetti, 2011; Gatbonton & Segalowitz, 2005). Also, since individuals have a limited capacity for processing information, Segalowitz and Hulstijn (2009) argued it is clear that language users are unable to attend to all information at all linguistic levels all together to the same high level. Paying attention to information regarding the content and the development of the interaction uses much of this limited capacity. This concern of trade-off in order to enhance all qualities of speech has encouraged many SLA researchers to examine the impacts of various procedural options, including repeating tasks with regard to facets of learners' oral production. Therefore, automatising vocabulary (i.e., through repetition) should free up attentional resources needed for generating information at higher levels (Burns, 2019; Bygate, 2018; Segalowitz & Hulstijn, 2005). In this respect, the fundamental notion for such tasks is practice and repetition (Bygate, 2018; N. C. Ellis, 2002; Gatbonton & Segalowitz, 2005; Thomson, 2020). This follows the common belief that explicit knowledge can play a causal role in the learning of implicit knowledge via practice (e.g., DeKeyser, 2018; N. Ellis, 2005; Hulstijn, 2007; Jeong & De Keyser, 2023; Paradis, 2009). While explicit knowledge is knowledge that can be available to conscious awareness; implicit knowledge is defined as the non-conscious knowledge (Ullman & Lovelett, 2018). For the purpose of the present discussion it is worth noting that the declarative memory system (a part of the mind's system) appears to underly explicit knowledge, whereas the procedural memory system (also part of the mind's system) underlies only implicit knowledge (Ullman, 2016; Ullman & Lovelett, 2018). According to Ullman and Lovelett (2018), learning in the procedural memory system occurs progressively through repetitive exposure which is therefore slower than learning in declarative memory, yet what is finally learned appears to be processed more quickly and automatically than knowledge in declarative memory. Ullman and Lovelett (2018, p. 45) claimed that as exposure to the conscious declarative knowledge continues (e.g., FSs), learning in procedural memory and automatization will increase and this progress may reach to a point where the procedural knowledge becomes stronger than using explicit knowledge (i.e., declarative knowledge), and therefore be "relied on instead". In other words, frequent exposures (e.g., through practice and repeated rehearsals) to the input may progress declarative knowledge to procedural knowledge and then to automatic performance. This is why the activation of procedural knowledge is "more suitable" for fluency in speech (de Jong & Perfetti, 2011, p. 537; see Ullman & Lovelett, 2018 for more declarative/procedural distinction and relevance for L2 teaching and learning) – that is why there is a need to focus on practice and repetition. Combining repetition with FS that already alleviates demands on WM and may benefit CAF of speech may indeed progress automatization. Since repetition progress procedural knowledge surrounding FSs (part of the declarative input) thereby increasing their automaticity and subsequently the CAF of FS and thereby increase the CAF of overall speaking.

The success of the use of repetition and imitation is evidenced in Ghazi-Saidi and Ansaldo's (2017) neuroimaging study conducted on Spanish-speaking (L1) participants and Persian-speaking (L1) adults of French (L2). Note that the former group's L1 (i.e., Spanish) was close to L2 (i.e., French), while the latter group's L1 (i.e., Persian) was distant from L2 (i.e., French). Participants were introduced to 130 new pieces of French vocabulary via audio-visual repetition and imitation (originates from behaviourism) database. Instructions were given to learners to examine each picture and imitate its name as closely as possible to the native model. Repetition was employed until the learners learnt the pronunciation of the word. Learners practised for 15 minutes, over 30 days, and were measured on accuracy and speed. Results revealed significant improvements of the efficacy of repetition and imitation in L2 vocabulary learning; learners were able to respond more quickly and accurately. This indicates high automaticity following repetition (Segalowitz & Hulstijn, 2005), a sign of effective learning (Segalowitz and Frenkiel-Fishman, 2005). Increased repetition paves the way to automaticity, which enables faster and more accurate processing and reduces cognitive load (Leonard et al., 2011; Mitchell et al., 2019; Parker-Jones et al., 2011; Suzuki, 2023). It is for this reason that Gatbonton (1994) supported the selection of utterances (e.g. FSs) to be automatised in order to be more easily expressed in a variety of communicative interactions. Gatbonton's (1994) analysis of L2 learning that concentrates on repetition leading to automatisation has been updated in the so-called ACCESS methodology (mentioned in Section 2.4.2). Several authors have followed up on this – for instance, in Wood's (2009) case study of a Japanese learner of English involved a high extent of repetition and practice of FSs related to specific kinds of narrative expressions. According to the researcher, it is likely that this enabled an increased proficiency (i.e., specifically showing clear fluency gains) with the sequences (in the measures of speech rate and mean length of runs) as they alleviated load on WM and cognitive processing, and access to them became easy as they became part of the learner's linguistic knowledge (Wood, 2009). In other words, automatisation reduces demands on WM, as the data becomes restored as a chunk, facilitating access to it (Mitchell et al., 2019). Indeed, Wray (2002) stated that if the learned FSs are encountered and used frequently, the learner will be able to recall them. However, Ghazi-Saidi and Ansaldo's (2017) study made it clear that with distant languages, the amount of repetition or amount of practice should increase to improve performance (in terms of faster response and accuracy). On the other hand, it is important to consider that basic simple repetition without creating form-meaning connection is not enough (Segalowitz & Hulstijn, 2009; VanPatten et al., 2004). If learners are simply repeating forms, without accessing the meaning (e.g., *see you Saturday* instead of see you *on Saturday*), the link between a specific form and its meaning will not be strengthened (i.e., using words without knowing how to match meaning with form).

Second, repetition allows familiarity with the domain knowledge as it enables L2 speakers to lessen the need to collect information for generating an idea (i.e., processing the preverbal message). Since Levelt's (1989) model of speech production relies on the workings of three components – Conceptualiser, Formulator and Articulator (2.3.2) – task repetition enables L2 learners to depend on formerly conceptualised task content and to recycle recently used linguistic knowledge to produce their utterance (Lambert, Kormos & Minn, 2017). In other words, the user will spend less time in the conceptualiser deciding what to express as the speech is repeated via task repetition, minimising the necessity for planning and therefore allowing for faster delivery. In this case, task repetition enables the speaker to give more processing resources to the accurate formulation (wording) of the planned content of speech, which may improve accuracy and complexity (Thai & Boers, 2016). Bui and Huang's (2018) study yielded many fruitful results regarding this matter (i.e., benefitting from ready-made knowledge that could speed up idea retrieval). Participants were requested to present two speaking topics: natural viruses and computer viruses. While nursing major students were more familiar with the former topic, the opposite applied to the computer majors. Participants were divided into pairs (i.e., one from each major) in the planning time (10 minutes before presenting), but no particular instruction was provided. The participants were allowed to use a pen and paper, but they were not allowed to refer to their paper during performance. As might be expected, findings showed that participants produced longer utterances on familiar topics than on unfamiliar ones, suggesting that prior knowledge in a particular domain facilitates the production of ideas and access. The researchers have made an intriguing connection between their findings and Levelt's (1989, 1999) speaking model, such as the benefit of prior knowledge (viruses in their study) allowed learners to overcome the challenges in the conceptualisation and formulation, resulting in a reduced breakdowns in speech (Bui & Huang, 2018). Indeed, Bygate (2018) considered repeating tasks a strong indicator for L2 fluency. He argued that being familiar with the topic of a speaking task (i.e., content) was more helpful than asking learners to just perform more speaking tasks, as a quantity of the work of conceptualisation, formulation and articulation operationalised on the first performance is accumulated in the learners' memory store and can be recycled on the second opportunity. This is reiterated in Bui and Huang's (2018) study when participants reduced mid-clause pauses and total silence as compared with the unfamiliar tasks. Indeed, the automatised ability of accessing FSs will decrease the necessity of pausing in order to select the right words and syntactic encoding. In this sense, repetition facilitates task familiarity allowing learners to access their memory for the language that is most relevant to the task; and this is how familiarity with the speech content is formed (Ahmadian & Tavakoli, 2011).

Thus far, this review has examined the importance of repeating oral performance; however, the question that arises is the number of repetitions needed for oral communication to be effective. Some studies asked their participants to repeat their performance once (e.g., Bygate, 1996, 1999, 2001; Wang, 2014), or a number of times ranging from three or four (Gass et al., 1999) to six (Lambert, Kormos & Minn, 2017) to 11 (Ahmadian, 2011). For some studies repetition effects have been identified immediately (e.g., Lynch & Maclean, 2000, 2001; Wang, 2014) or after days (Gass et al., 1999) or weeks (Ahmadian & Tavakoli, 2011; Bygate, 1996, 1999, 2001). Put simply, there is unlikely to be a golden rule about the number of repetitions required for learners to enhance the varied aspects of their L2 fluency in the short term and how this may differ with task type and learner proficiency level. However, it seems reasonable to argue that the effects of repetition (i.e., in speaking task) on speaking production may vary according to the speaker's level of proficiency. This is because the automatisation of linguistic encoding is strongly correlated with L2 proficiency (Segalowitz, 2010; Suzuki, 2023). Advanced L2 speakers may have more attentional resources available for conceptualising task content. Cleary, this is because from a usage-based perceptive, higher-proficiency learners are engaged with the language to a greater extent (Kyle, 2016). As learners progress and accumulate more language experience, their working memory becomes more efficient at handling the language's intricacies by producing more phonetically and syntactically complex, accurate as well as fluent utterances. Lambert, Kormos & Minn (2017) argued this is why higher proficiency L2 learners may benefit less from task repetition, as one of its key functions is to enable learners to activate familiar content and linguistic formulation processes. This, however, does not indicate that task repetition is not suitable for advanced learners as there are also tasks that for example involve high demands during the conceptualisation phase.

While there are many advantages to task repetition, there are also some disadvantages. It is important to note that learners may get bored and fatigued by repeating tasks (Bygate, 2001). Some evidence indicated that while learners consider repeating tasks to be a beneficial classroom task (e.g., Pinter, 2010; Lambert et al, 2017), L2 teachers were hesitant to develop it in their practice, fearing that learners may find task repetition boring (Ahmadian et al.; 2017). Lambert, Kormos & Minn (2017) investigated Japanese university learners' (N = 32) reactions to tasks repetition (six times) as speaker and listener with different interlocutors each time and their opinion regarding the benefits of such repetition. Participants' perceptions were elicited through a questionnaire. Results showed that between 47 per cent and 72 per cent of the participants considered that three or four performances were enough. This observation aligns with the researchers' conclusion, as it triangulates the findings of the fluency analysis, particularly reflecting progress during the first four performances of each task (Lambert, Kormos & Minn, 2017). Moreover, it showed that learners did not lose interest or become fatigued, but valued the opportunity to carry out the same tasks with different interlocutors across the time period of a standard lesson (Lambert, Kormos & Minn, 2017). This was evidenced through the aural-oral task repetition analysis, which showed a significant effect and a great effect size for oral task repetition with different interlocutors (as speaker and listener) on L2 speech fluency (i.e., speed, breakdown and repair fluency) for participants at different levels of proficiency (high-proficiency group equivalent to B2 and C1 levels; intermediate-level group to the B1 level; and the low-proficiency learners to the A2 level on the CEFR). However, these results cannot be generalised, as the sample size used in this study is small, and learners may vary in terms of individual differences, such as motivation, WM and planning skills.

In fact, this requirement of alternating interlocuters (speaker and listener) each time the task is repeated is described by Bygate (1996, p. 145) as a task that can "provide different learning opportunities" on the grounds that it prepares each learner for a real-life situation. More recently, Lambert, Kormos & Minn (2017) indicated such a methodology allows learners to be exposed to both input-based and output-based forms of the task. As such, task repetition may assist learners in enhancing their performances by allowing them to "activate", "refine" and "optimise" their linguistic knowledge with the aim to effectively complete a given task consistent with the particular interactive demands that it involves (Lambert, Kormos & Minn, 2017, p. 169).

While it appears that the learning of FSs via repetition as a learning strategy improves L2 learners' oral proficiency including complexity, accuracy and fluency and helps automatisation, other studies also encourage different strategies such as memorisation (e.g., Dai & Ding, 2010; Ellis & Shintani, 2013; Suzuki, 2023; Wray & Fitzpatrick, 2008). The role of memorisation will clearly be discussed in the next sub-section.

2.4.5 Memorisation and performance

Several scholars consider the learning of FSs is vital in L2 learning as its use makes speech sound natural, idiomatic and native-like (Conklin & Schmitt, 2008; Ding & Qi, 2005; McGuire & Larson-Hall, 2018; Wood, 2010). However, few studies focus on how to develop complexity, accuracy and fluency of L2 learners' FSs (Qi & Ding, 2011). For many years, the issue of whether to include memorisation of FSs in the second language classroom has been controversial. Research has resulted in two different approaches: the first group of researchers and educators advocate memorisation as an effective learning strategy that language learners can benefit from (e.g., Ellis & Shintani, 2013, Dai & Ding, 2010; Suzuki, 2023; Thomson, 2020; Wood, 2009; Wray & Fitzpatrick, 2008). This is because the skill of memorisation begins with conscious and deliberate efforts to store information in short-term memory. In contrast, the second group is more critical about its use. Cook (1994, p. 133) posited that repetition and memorisation are considered as two "outlaws". Granger (1998) claimed that there is no clear association between memorisation of FSs and the production of creative language. Similarly, many common views in the West consider memorisation as a "superficial" and meaningless language learning method (Wray & Fitzpatrick, 2008, p. 125).

While the value of memorisation of texts in L2 learning is debatable (Dai & Ding, 2010), some studies, specifically those targeting Chinese learners, have shown that memorisation is a successful learning strategy that facilitates the learning of formulaic expressions, which therefore helps to enhance the quality and fluency of learners' language use (e.g., Dai & Ding, 2010; Ding, 2007; Qi & Ding, 2011; Wray & Fitzpatrick, 2008). Suzuki (2023) claimed that L2 learners benefit from contextualised practice, such as memorisation—engaging repeatedly with context-rich texts in various ways. Indeed, Ding (2007) reported how a group of three successful Chinese L2 learners (of English) won the annual nationwide English-speaking competitions for college students. It was found that the memorisation and use of collocations, phrases and sentence patterns were the reasons for their success. For example, one of the learners stated that many students indicated “Family is important”; however, she used a sentence pattern from a textbook: “Nothing can be compared with the importance of family” (Ding, 2007, p. 277). Ding (2007) strongly encouraged planning and rehearsing by text memorisation and imitation, as such tasks allow learners to apply what they have recited and noticed so that it develops to be their own language. The role of consciousness in second language learning (i.e., noticing) has been thoroughly discussed in Section 2.2.3. Schmidt (1990, 2001) stressed the importance of noticing in the sense that improvement in language ability occurs when the learner becomes aware of a certain language component. On the other hand, N. Ellis (2003), Van Vu & Peters (2022) and Wulff (2019) accentuates the role of frequency of exposure and practice of learned items. Learners require multiple encounters with lexical items for effective learning to take place. That is, the employment of text memorisation does not only focus a learner’s attention on language form but also makes use of the form more likely. Indeed, the usefulness of memorisation depends on what the learner pays attention to (Wray & Fitzpatrick, 2008). It is in this sense that Cooper (2004) argued that memorisation is beneficial when it enhances and/or assists comprehension. Thus, learners in the classroom can be given the opportunity to memorise texts or dialogues that are related to their needs and rehearse them in front of the class; this may allow them to notice the gap between their own performance and their classmates. Such practice could encourage them to ask for clarification, which may facilitate understanding (Thornbury, 2005, p. 43). However, some students may become disinterested and disengaged when faced with repetitive memorisation tasks, leading to decreased motivation to learn the language. For this reason, memorisation of texts should be interesting and relevant to the students’ interest and needs. For example, learners can memorise texts or phrases that are relevant to future job interactions, subsequently practicing them with peers or in group activities, or even through role-play tasks. It is also important for students to recognise the benefits and outcomes of employing such a learning strategy, which can be demonstrated through activities like before-and-after oral speaking tasks, such as recording their progress. Indeed, Nation (2007) suggests one of the best fluency activities is recording, where the learner speaks about a given topic (e.g., arranging a meeting) to the recorder, listens to it to check whether corrections could be made, and then re-records the speech. Knowing that they are going to be recorded, learners might find greater motivation and encouragement to employ the strategy of memorising utterances or FSs to effectively convey meaningful messages. Despite the benefits of memorisation in language learning, as discussed earlier, this paper also recommends that L2 classrooms take into account diverse learning preferences and strategies to accommodate various learning styles. Although memorisation of utterances might prove useful in theory, in practice however students may be reluctant to carry out such difficult tasks and may be too shy to perform in front of their peers.

Additionally, rehearsing the task allows repetition of input as well as output, thus engaging and improving listening skills as well, as students will be listening to the reading of these dialogues by a native speaker on a recorder. This is demonstrated in Wray and Fitzpatrick's (2008) study as they measured the effectiveness of memorising and rehearsing conversational turns for L2 learners. One of the participants reported that the rehearsal of patterns had largely enhanced her listening skills, an aspect that made the researchers consider that memorisation has several potential benefits in L2 learning (Wray & Fitzpatrick, 2008). FSs not only help the user in an interaction to produce messages by using memorised fixed strings, but they also help the listener to decode the message easily (as they are formulaic). Thus, the utilisation of FSs improves the comprehensibility of L2 speakers' productions (Saito, 2020; Saito & Liu, 2022), resulting in a more effective conversation

(Wray, 2002). This is because if the speaker has non-native phonology, the listener will need to make more effort to process the phonological decoding, which is not the case if the sequences are familiar to the hearer (Wray, 2002). In other words, the speaker's use of novel utterances requires the listener to engage in greater processing than if the speaker uses prefabricated chunks (that express the same idea) as these chunks will be in the listener's repertoire, reducing the processing load. This is evidenced in a study conducted by Shams and Shamsaee (2010), who found that speech with higher rates of formulaic language seemed more interesting to listeners. Thus, FSs are beneficial for L2 learners, especially for beginners, as their non-native phonology might get in the way of their communicative success, while the familiarity of the FSs may benefit their interlocutors in overcoming this hurdle. Note that this is not to downplay the vital role that human creative linguistic capacity plays, but formulaic expressions do indeed facilitate the interlocutor's decoding and L2 learners' speaking production. If L2 learners have minimal exposure to the target language, producing novel utterances can be difficult (Wray, 2002). In fact, learners may not be prepared to produce spontaneous communication at a time when they are not yet ready (e.g., due to level of L2 proficiency).

Despite these positive findings, one may turn to the expectation that learners who memorise texts would consume extra time on learning, which may account for learners' better performance. It is for this reason that Dai and Ding (2010) investigated whether text memorisation has any outcomes on L2 learners' proficiency and writing ability and whether text recitation generates the same or different outcomes in advanced and beginner learners in the use of FSs in their writing. The study was conducted on the Chinese undergraduate English course of a military academy and lasted for around three months. The amount of time spent on learning outside the class was controlled as the military programme set a specific time for self-study (no one is allowed not to study). After the self-study period learners were not allowed to study further and were sent to their dormitories. Therefore, it was possible for the researchers to compare a group requested to memorise text (experimental group; 26 participants) with a group not given the task (control group; 29 students). Both groups were instructed by the same teacher and were significantly comparable in terms of English proficiency (as revealed in their pre-test scores). Therefore, five advanced learners and five beginner learners were selected from the experimental group in order to investigate the differences in their performance of text memorisation. Results showed that the experimental group outperformed the control group significantly, which could indicate that text memorisation assisted these learners in enhancing their English proficiency. However, the findings of Dai and Ding's (2010) study should be used with caution for two reasons: first, the sample size featured in the study is considerably small without the researchers providing convincing justification for such a small sample size and how the data can be considered representative of language learners in this particular context. Second, the study targeted the effectiveness of memorising FSs via written production as opposed to the demand of on-line production, as written interaction may involve a larger variety of lexicons, including FSs, because it is not produced under real-time pressure (as noted in Chapter 2, Literature review, Section 2.2.2). However, in a later study Qi and Ding (2011) investigated the way Chinese L2 learners of English improve their FS use in a speaking production. Their investigation included the use of FSs by 56 advanced Chinese university English majors in their memorised monologues at the start and end of a three-year period. The two monologues involved the same topic: *Comment on college students renting apartments off campus*, and each monologue lasted for around three minutes. The use of FSs was assessed with regard to three measures: frequency, accuracy and variation (i.e., range). Findings demonstrated that Chinese EFL learners had improved in frequency, accuracy and variation of the FSs developed in speaking production; however, despite over three years of college study, they still do not achieve a highly advanced level in the use of FSs in terms of frequency and accuracy. An example of the accuracy errors experienced by L2 learners, and the most difficult part for them to deal with, was the use of FSs including prepositions (e.g., *in campus* instead of *on campus)* and articles (e.g., *in the society* instead of *in society*), which when combined, accounted for about 60 per cent of all the errors of FSs in speech. However, learners showed considerable improvement in overall speaking proficiency as they progressed from Year One to Year Four (not in error types, e.g., prepositions). Along the same lines, Sabbah (2015) highlighted the most challenging areas for Arab learners of English as including use of prepositions (since they are limited in Arabic), definite articles (misuse between *a* and *the*) and third person singular -s for verb agreement (learners omit it). In fact, Qi & Ding (2011) posited that prepositions and articles seem to be challenging to many ESL learners and are likely to be the first to be forgotten under real-time pressure. Such findings are valuable to language educators in order to notify them of where to target their focus and when to improve the idiomaticity and fluency of their learners' speaking production.

In other investigations, two of the few studies examining the usefulness of learning specifically targeted FSs (i.e., selected on the basis of the needs of the learner) via memorisation are Wray's (2004) and Wray and Fitzpatrick's (2008). Wray (2004) examined the performance of a novice learner of Welsh using the L2 after four days of training to perform a cooking show. Findings revealed that the memorisation of long texts assisted the learner's performance. Wray and Fitzpatrick (2008) found that memorisation and practice of L2 interactional turns significantly facilitated real-life communication among adult learners. They explicitly expressed that memorising utterances progresses learners' fluency, and suggested that memorisation can be used as a means of determining the strengths and weaknesses of learners with regard to morphological, lexical and phraseological knowledge

(Wray and Fitzpatrick, 2008). Indeed, the memorisation of chunks facilitates faster processing, not only for production but also for reception, as it is easier to access linguistic material from long-term memory than produce a novel one (Ellis et al., 2008). Interestingly, findings also showed that the memorisation of utterances boosted participants' confidence – it gave them the chance to sound like native-speakers, decreased the stress of on line production and gave them self-assurance about being understood (Wray and Fitzpatrick, 2008). Wray (2002) explained that even if L2 learners do not fully understand the formulaic sequences, they will still be confident to use them as formulating utterances is extremely demanding. This is evident in Hall's (2010) study, which sought to investigate whether formulaic chunks were perceived in the oral production of three adult beginner learners (speeches limited to three or four words at a time with little grammatical inflections, and slow word production) of English for more than a nine-week period in a language course. These learners came from diverse backgrounds with L1 Korean, L1 Ukrainian and L1 Arabic. Language use targeted for teaching in the programme involved simple vocabulary such as names of clothing, colours, family, money and basic communicative purposes (e.g., asking for prices, comparing, expressing daily habits using the simple present and present progressive). Results revealed that formulae were evident in all learners' oral production to a varying degree. This is because formulaic sequences make grammatical rules and lexical entries accessible (Wray, 2002).

Given that FSs are stored, retrieved and used as a whole, it is the memory system that supports the use of FSs. However, it is not possible to memorise all utterances. Thus, in order to increase the usefulness of memorisation, a balance must be struck where large linguistic units can be called upon when needed and modified based on the requirements of the context (Wray & Fitzpatrick, 2008). A humorous example is found in the famous American comedy show *Friends* (Season 7, Episode 1), with Monica using the FS *steal my thunder* (i.e., reappropriating my moment of celebration and taking it for yourself) in multiple ways including, "you kind of stole my thunder", "thunder being stolen", "it's not the first time you've stolen my thunder". In order for learners to use language creatively and flexibly, Hall (2010) stressed the need for an abundant experience with language input to create a basic store of language forms.

As was mentioned in Section 2.4.2, the Free Communication Phase (i.e., the final stage in the ACCESS methodology) can be used as a teaching approach that allows the learner to use input productively in order to practise and build up automatised knowledge. Learners with very little or no instruction in English are required to have automatised some lower-level language aspects including lexical access and phonological processing in order to allow memory resources to process and absorb multi-morphemic units as a whole (Ellis, 2001). For instance, beginner L2 learners such as the participants in Hall's (2010) study have a limited processing span and minimal L2 knowledge, thus a form–meaning mapping task is demanding in even the friendliest context of language learning. Therefore, beginner learners may only use formulae whose forms are salient, either via frequency or by teaching, or whose meaning lies within the memory span of the learner, and are obtainable from the context (Hall, 2010). As proficiency level increases with noticing and explicit instructions, L2 speakers can gradually be able to pay attention to morphological details (i.e., make some minor changes to the lexical item) and if a certain piece of vocabulary is forgotten, find an alternative one to replace it (Wray & Fitzpatrick, 2008). Indeed, recognising and strengthening patterns that were not fully understood at the beginning may lead to discovery of new meanings (Dahlin & Watkins, 2000).

It is argued that despite the criticism of using strategies including text memorisation these strategies/methodologies used by Asian students should be considered in a more positive light based on their successful results in L2 learning – a point that Pennycook (1996) had argued for. Of course, this paper does not suggest that such methodologies (e.g., repetition, FS memorisation and rehearsal) fit all L2 contexts. Instead, these strategies/methodologies are to be applied with cautious consideration of learners' needs and context. It is advantageous to note that the teaching approach that instructors use has an effect on the quality and usefulness of the learning context for second language learners (Gibbons, 2007). Therefore, when teachers and language educators have a deep understanding of both L2 theory and how it informs practice, they are more capable of fulfilling learners' needs in the L2 classroom.

2.5 Key issues in L2 learning and teaching of speaking

Despite the importance of speaking skills in second language learning, paradoxically, little consideration is given to them in the world of linguistics and language teaching, where the focus remains firmly on written skills (Hughes, 2010). In fact, it was believed that effective speaking skill developed automatically from the teaching of writing skills, grammar and vocabulary (Nation & Newton, 2009; Thornbury, 2005). Goh (2007, p. ii) described oral skills in the classroom as the "less frequently taught" (Goh, 2007). It seems that the teaching of speaking in language classrooms has long been challenging for many language educators, as oral interaction can be problematic (Burns, 2019; Bygate, 2001, 2009; Thornbury, 2005) and researchers agreed that principled methodologies on how to teach speaking are still lacking (Bygate, 2001; Segalowitz, 2010). However, a consensus on a unified speaking pedagogy is unlikely to be reached as learners have different needs—including the level of L2 proficiency, individual differences, etc. Thus, the goal is not to unify the methodology but to better inform the development of future methodology and learning principles through advancements in the understanding of language ability in general, and facilitating L2 oral production specifically.

In fact, for a long time, linguists and language educators have been concerned with investigating factors that impact second language learning, particularly aspects that hinder speaking in the classroom. There are several reasons why speaking skills are often not instructed or rarely being given attention in the L2 classroom. As previously discussed in the literature review (i.e., the background and rationale for the study chapter), if speaking production is not assessed, teachers may be hesitant to allocate classroom time to teaching speaking production. Instead, they may prioritise test-driven aspects such as grammar, writing or teaching list of vocabulary (Pakula, 2019). Another contributing factor to this issue may be that, even if textbooks incorporate oral activities, it may be that the kind of these oral tasks may not necessarily progress speaking proficiency or promote learning. For example, a speaking task centered around a specific context might be introduced only once in the textbook without being repeated in further units throughout the textbook or receiving further elaboration. This comes in sharp contrast with the so-called ACCESS methodology (as discussed in Section 2.3.2) and several authors who stressed that the repetition and consistent practice of tasks leads to automatization (e.g., Mitchell et al., 2019; Wood, 2009; Van Vu & Peters, 2022; Wray, 2002 –see Chapter 2, Literature review, Section 2.3.4 for more details). Indeed, the relatedness of topics is a matter that should be continued throughout the following sections of the textbook in order to familiarise learners with both the content and language use, thereby minimising unfamiliar occurrences (Nation, 2007; Mystkowska-Wiertelak, et al., 2017). It is also possible that teachers may be uncertain about how to teach oral English (Chen & Goh, 2014), possibly due to insufficient pedagogical knowledge concerning English speaking proficiency (Chen & Goh, 2014) or inadequate understanding of their own spoken language proficiency (Liao, 2009; Pakula, 2019). For example, in a study conducted by Tavakoli and Hunter (2018) among 84 L2 teachers in England, investigating teachers' perceptions on fluency (a fast and smooth speech) and their practices on employing it to learners, found that teachers only concentrated on error-corrections during speaking tasks. On the other hand, cultural issues may also play a role in student reluctance to participate in classroom speaking activities, as they may fear making mistakes or simply shy (due to their low oral proficiency skills) (Alharbi, 2015; Gudu, 2015; Stupar-Rutenfrans, Ketelaars, and van Gisbergen, 2017; Pakula, 2019). In this respect, findings from Littlewood's (2004) study of 567 students in Hong Kong showed that one of the main issues stopping students' participation is timidity and ability to express ideas. Mystkowska-Wiertelak et al. (2017) study on 48 Polish university students have yielded many fruitful results, it was revealed that providing L2 learners with helpful vocabulary assists them to express meanings, increases their confidence and thus develops their willingness to communicate. Similarly, Burns (2019) stated that students' confidence is a major factor that triggers L2 learning including speaking activities, which can therefore affect their willingness to communicate. Providing learners with sufficient linguistic knowledge related to the task can reduce unfamiliarity, thereby boosting learners' confidence in their ability to communicate. This is particularly important given the limited exposure and interaction that many L2 learners encounter, which makes it challenging to develop an accurate understanding if preferred lexical combinations (Van Vu & Peters, 2022). However, willingness to communicate in class is not a key focus of this paper, which concentrates on the use of linguistic knowledge (in terms of meaning) in assisting the improvement of L2 speech production. The absence of vocabulary can significantly prevent learners from stating ideas coherently to produce authentic speech (Hinkel, 2020; Van Vu & Peters, 2022; Webb & Nation, 2017).

In various regions around the world there is a consensus that majority of English learners face the same problem that is – they cannot produce desired English language even after having learned it for a long time – as in the following: Alonso (2013) stated that students in Spain struggle to produce proficient oral English even after completing English as a mandatory subject in Spanish secondary education. Similarly, Gudu (2015) observed a comparable situation among school graduates in Kenya. Ariyanti (2016) highlighted that Indonesian students are often regarded as low achievers in oral English, despite its inclusion as a compulsory subject from elementary to university levels. Furthermore, Alharbi (2015) claimed that learners in Saudia Arabia face challenges related to insufficient communication skills in the L2 (English). More recently, Zhang and Liu (2018, p.1341) expressed explicitly that oral English has consistently been the “largest obstacle” to Chinese learners. Of course, these are only some of the examples illustrating the global phenomenon of challenges in teaching speaking production within classroom context, a phenomenon that is also relevant to the context of this study.

The next section will expand further to the situation of oral production in Kuwait.

2.5.1 The situation in Kuwait

Generally, English is a compulsory subject in Kuwait from the time students start their primary education at the age of six. English as a second language is taught for around 40 minutes a session five times a week in schools, and students are required to learn different aspects of the English language, including a large amount of vocabulary. Similar to many contexts around the world, this traditional method of teaching English persists even at the university and college levels. Specifically, in the College of Business Studies, Kuwait, the main focus of teaching the L2 (English) is on teaching vocabulary lists and grammar instructions, or common verbs and prepositions with ‘fill in the blanks’ and multiple-choice exercises completed through text books, worksheets or quizzes. Clearly, greater emphasis is placed on linguistic competence than on linguistic performance, which subsequently hinders learners’ speaking proficiency and the assessment method itself (Correia, 2016). Unfortunately, there is often a lack of emphasis on practicing spoken language within the classroom. Consequently, L2 students have very little exposure to communicative activities that would enhance their speaking proficiency.

While in some contexts, such as Norway, English oral ability is assessed at upper secondary level (Bøhn, 2015) and in others, such as Brazil, at university level (in both public and private institutions) (Scaramucci & Boffi, 2013), this is not the case in Kuwait. In Kuwait, assessments are distributed during the middle and at the end of the instruction period and consist of a reading comprehension section, vocabulary, grammar and writing, with no attention given to oral exams. Since there are no clear criteria to measure speaking performance, this implies this assessment is subjective (Scaramucci & Boffi, 2013). This might be the reason why many students enter college unable to use the language productively (i.e., speaking), even after having learned it from the first grade onwards.

These assessments focus on items of reading comprehension, grammar, vocabulary and what to say in a given situation (language function). Assessment is constructed by the teacher, determined by the textbook or by materials prepared by the teacher, since there are no standards or criteria developed by the institution. The institution only controls the number of assessments during the semester and the date and time of final exams, in addition to requiring the four mentioned sections are part of the exam (reading comprehension, vocabulary, grammar and writing). Although, it is beyond the scope of the present study to delve into the assessment aspect further, clearly future research could explore the evaluation of oral performance and ways of assessing it in the classroom. This exploration should specifically focusing on lower-proficiency leaners who need to fulfil their needs, such as using the L2 for their future careers.

2.6 Summary of the chapter

In this chapter, I have provided an overview of the literature on the relationship between the use of FSs and second language oral performance and discussed the underlying rationale for this study. Starting with a definition of the term FSs, I discussed the importance of FSs in language use, particularly in L2 learning.

I further explored Levelt's L1 speech production model and its adaptation for L2 oral production, highlighting the distinctions between L1 and L2 speech. I have also presented an in-depth discussion of several critical factors, including WM, attention and noticing ability, emphasising how each of these individual learner factors can significantly influence L2 oral performance and potentially mediate L2 speech production. Additionally, I further discussed the challenges faced by L2 speakers while attending to CAF all at the same time. Then, I moved on to clarify the qualities of speech—namely 'fluency', 'accuracy', and 'complexity and discussed the way they are measured in the literature. The chapter also explored the measures of FSs, specifically focusing on accuracy and fluency. Section 2 ended with additional factors that influence speech production, extending beyond the standard scope of complexity, accuracy and fluency of entire speech, and FSs.

To foster success in speaking, I introduced Gatbonton & Segalowitz' (2005) ACCESS pedagogical framework as an approach to promote learners' speech production, highlighting its potential to facilitate L2 speaking performances. Next, I detailed various strategies applicable in the L2 classroom, such as pre-task planning, repetition, memorisation and performance, and their impact on L2 CAF production, including their role in reducing cognitive load, particularly on working memory.

The chapter concludes by addressing key issues in L2 speaking learning and teaching, contrasting them with the ACCESS methodology. I emphasised the importance of aligning pedagogical methods and pedagogical theory to enhance learning outcomes. Moreover, I provided examples of similar language learning challenges worldwide and narrowed the focus to the specific context of Kuwait, where this study is situated.

In the next chapter, I will present and discuss the methodology of this study.

Chapter 3. The Current Study

3.1 Introduction

This chapter starts by presenting the aims, research questions and corresponding hypotheses that guided this research. Next, the study design employed here will be explained, with a justification for the chosen methodology. The chapter proceeds to set a context for the study participants, discusses the tasks, instruments, as well as the issues encountered during the pilot study. Following this, it outlines the ethical procedures, followed by a detailed discussion of the data collection process. Next, I present the steps that were followed in the data analysis, narrowing the focus to provide a comprehensive explanation of the operationalisation and measurement of complexity of the FSs elicited from students' speech samples. The choice of the analysis measures is acknowledged and then justified. The chapter further discusses the procedures for transcribing and coding the data, along with addressing interrater reliability issues. Then, it outlines the analytical procedures aligned with the research questions, covering the measures that were adopted to address each research question in-depth. Finally, the chapter concludes with a summary.

3.2 Aims of the study

As observed in the literature review, speaking can be challenging for L2 speakers, requiring the simultaneous execution of various components such as fluency, accuracy, complexity and automatization. Unlike L1 speakers, L2 learners face challenges in speech production due to increased cognitive load on attention and WMC, particularly under time pressure. The literature also suggested that introducing L2 speakers to FSs can facilitate speaking production, discourse handling, conversation, and language learning and processing. However, there is a significant gap in understanding the usage of formulaic language, especially among beginner learners.

While FSs are frequently discussed, there is limited exploration of their complexity and the impact of complexity, accuracy and fluency on overall fluency of the speech samples.

Moreover, despite advocacy for the use of FSs in the literature, few studies delve into how L2 learners can effectively learn these sequences and the variables that facilitate their learning.

This study set out to explore the need to investigate the influence of individual differences variables, including linguistic variables such as vocabulary tests, cognitive variables like WM and background variables, on the learning and production of complexity, accuracy and fluency in overall speech of undergraduate students via a speaking task in a second language context. Additionally, it explores how these individual differences variables influence the complexity, accuracy and fluency of formulaic sequences produced by these learners. This study also aims to investigate the application of FSs with regard to the students' textbook. It specifically addresses L1 Arabic adult lower-proficiency learners (those who have not yet learned the 2,000 most frequent words of the language) who need to be prepared to use the necessary amount of L2 (English) for their future careers in business.

As discussed in the previous chapter, this study is motivated by existing research which suggests that FSs plays big role in facilitating speaking production (Goncharov, 2019; Liang, 2017; Mirzaei et al., 2016). This study expands our knowledge of the link between FSs and CAF of L2 speech performance in several directions. First, it aims to explore the extent to which FSs does indeed mediate CAF of speech among low-proficiency learners. Unlike most studies in this field, which has focused on the relationship between FSs and speech performance in intermediate to advanced language proficiency levels (Boers et al., 2006; Goncharov, 2019; Khodadady & Shamsaee, 2012; Mirzaei et al., 2016), this study shifts its focus to adult Arabic beginner learners. Furthermore, beginner learners, particularly those in non-immersion contexts, have limited exposure to L2 input. In such cases, their primary source of L2 input typically comes from the classroom. This exploration is vital for improving their conversational skills and future career prospects.

Secondly, this study aims to examine the relationship between various individual variables (general proficiency test, vocabulary knowledge, FS knowledge, WM test and variables from background questionnaire), in relation to speech production, particularly in the context of FSs examined through oral performance. Given the scarcity of correlational studies in SLA that examine these factors and their influence on speech production, this research aims to identify the variables that are most predictive of both CAF in speech and CAF of FSs, offering insights into enhancing speech production.

In addition, the present study aims to explore various dimensions of complexity and to provide further insights into the multidimensional aspects of complexity of FSs. As was discussed earlier in 2.3.7, while the measurement of complexity of entire speech is common and standardised in the literature, the measurement of complexity of FS remains rare. Examining the complexity of FSs in this study is important for two main purposes:

1) It allows us to gain insights into how the complexity of FSs can enhance the complexity of entire speech.

2) It facilitates a more comprehensive evaluation of a learner's proficiency by measuring the complexity of FSs sequences.

This is a methodological shortcoming of the research in this field, and it constitutes a novel aspect this study aims to address.

In addition to the lack of measurement of complexity of FSs, this study aims to investigate another individual difference, WM, in lower-proficiency learners and its potential influence on CAF at lower proficiency levels. As already discussed in Chapter 1, Introduction, Section1.1, the motivation for incorporating WMC to the current study as another independent variable stems from previous research indicating its significant impact on L2 development and L2 performance (Awaad & Tavakoli, 2022; Baddeley, 2015; Mitchell et al., 2013; Skehan, 2015). As mentioned before, research has shown that learners with high phonological short-term memory tend to excel in speaking (Révész, 2012). This is because the processing and temporary retention of both familiar and novel phonological knowledge rely on phonological short-term memory (Juffs & Harrington, 2011), which is linked to higher WMC and leads to more effective speaking. However, limited attention has been directed toward understanding how WMC influences speech production among beginners. Thus in relation to the positive findings between the relationship between WM and speech production, it can be hypothesised that WMC, as a cognitive ability factor, may mediate CAF of speech for lower proficiency learners, which is what the current study is interested in.

In addition to examining WM in lower-proficiency learners, this study also aims to explore a relatively understudied aspect in the literature: the relationship between WMC and the CAF of FSs elicited from a live performance of the task. As mentioned before, past studies have indicated the pivotal role of WM in both L2 production and L2 processing. More specifically, WMC has been associated with L2 vocabulary development. As discussed in the previous chapter, Baddeley (2003, 2015) highlighted that the phonological loop component plays a crucial role in the learning of L2 words. Given that FSs are an integral part of vocabulary knowledge, it is posited that a higher WMC enables the allocation of additional attentional resources, facilitating the intricate handling of lexical input and output. Therefore, it is suggested that WMC could enhance vocabulary learning (Baddeley, 2015; Juffs & Harrington, 2011), particularly for individuals at lower proficiency levels via L2 speech production. Consequently, a hypothesis can be formulated, suggesting that WMC, as a cognitive ability factor, may play a large role in mediating the number of FSs produced by participants in their speech samples.

3.3 Research questions

This study has been guided by the following research questions (RQs) and their related hypotheses (H):

RQ1) Do ESL beginner college learners use formulaic language included within the curriculum and textbook when performing an oral proficiency task? If so, how do they use these FSs?

H1. Drawing from the literature, when speakers are under time constraints, their primary focus is on conveying their message, while the use of complex language not being a priority for them, largely due to the potential working memory load (Ahmadian & Tavakoli, 2010). In the speaking task, learners are more likely to use simpler FSs rather than the more complex ones (both available in their textbook) that fulfil the same communicative function.

Simpler FSs include w*hat is your name* and *I am sorry I have to go now* as opposed to more complex ones, which include *let me introduce myself* and *I'm afraid I have to go now*, despite the fact that both types are found in the textbook.

It is important to note that FS complexity was evaluated using various criteria from the literature to capture its multidimensional nature. This included counting the number of syllables per word and per FS, as well as considering the semantic transparency of the FS. A FS was deemed less complex if its meaning could be easily derived from its parts, while deceptively transparent FSs, where the meaning could not be inferred from the components, were considered more complex and coded accordingly. Native speakers also provided judgements on the complexity of FSs using a six-point Likert scale. These judgements, along with the indicators of complexity (word length, semantic transparency and native speaker intuition), were combined to create a final complexity composite score (from 1 to any) for each FS. Further details on the exploration of FS complexity are provided in Section 3.10.3.

RQ2) To what extent is the CAF of students' entire speech sample related to their use of FSs in a speaking task?

H2. Drawing from the literature, fluency is enhanced when the content is familiar, and tasks demand less intricate planning (Foster & Skehan, 2013) Additionally, findings from Bui and Huang's (2018) study revealed that fluency was enhanced on familiar topics compared to unfamiliar ones, indicating that prior knowledge in a specific domain facilitates idea generation and accessibility. Given that the speaking task used in this study is taken from the students' textbook and is expected to be within the learners' prior experience, it is reasonable to assume that the task is familiar to the participants. As a result, of the three variables of complexity, accuracy and fluency, I expect fluency to be most closely related to participants' use of FSs, followed by accuracy and finally complexity.

FSs in this study were classified as FSs uniformly, following the criteria outlined by Carrol and Conklin (2020) and Wood (2020). Any word combination with a MI value of at least 3.0 was considered formulaic, based on words that have a strong relationship with each other in creating meaning (See Section 2.2 for more details). This identification of FSs from participants' speech samples and the textbook is detailed, and the software used to identify these FSs is discussed in Sections 3.10.1 and 3.10.2. After extracting FSs from participants' speech samples, the total number of FSs used in each participant's speech sample was counted. These FSs were then examined from three distinct perspectives: complexity, accuracy and fluency.

RQ3a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their personal background variables (Language History Questionnaire – LHQ) on the one hand, and CAF of the entire speech sample on the other?

H3a.

- FS tests, particularly the test of familiarity of FSs (FSs taken from students' textbook), will have a strong correlation specifically with the fluency of entire speech followed by the complexity and accuracy of speech. This hypothesis is grounded in the understanding that FSs play a big role in facilitating speech production (Boers et al., 2006; Wood, 2009; 2010; 2016). It is also aligned with the goals of the ACCESS methodology (Gatbonton & Segalowitz, 2005) and the crucial role of familiarity with the domain knowledge in enhancing speech production (Bui & Huang, 2016; Dai & Ding, 2010; Ellis, 2003; Thai and Boers, 2016).
- The vocabulary test will have a high correlation with the fluency of entire speech. This hypothesis is based on the idea that vocabulary knowledge and fluency are interconnected. L2 learners with broader lexical knowledge and quicker access to it may excel in conducting vocabulary searches with ease and speed (Koizumi & In'nami, 2013).

- The general proficiency test will have a moderate correlation with the C, A and F of entire speech.
- The higher the WM, the better C, A and F of entire speech is. This hypothesis was corroborated by the literature, suggesting that working memory plays a crucial role in various higher cognitive tasks, including speaking (e.g., Baddeley, 2001, 2003, 2015; Conway et al., 2007; Mitchell et al., 2019) (see Chapter 2, Literature review, Section 2.3.3).
- The more hours learners spend using everyday spoken English (gathered from the LHQ), the better C, A and F of entire speech is. Clearly, increased exposure and practice in the L2 result in improved speech production. This process mirrors the process of several years of L1 acquisition, which ultimately turns most adults into fluent speakers of their native language (Segalowitz & Hulstijn, 2005).

RQ3b) Which of the predictors (general language proficiency, vocabulary test scores, WM test scores, personal background variables) best predict speaking task performance as measured through CAF scores of the entire speech samples?

H3b. FS tests scores, followed by WM test scores, are hypothesised to be the best predictors of CAF of the entire speech samples elicited by participants. This hypothesis is based on a series of studies by Wood's (2009, 2010, 2016) which have consistently demonstrated that increased exposure to FSs helped a participant to show significant gains in fluency in particular (see Page 42). Regarding working memory, this hypothesis is grounded in the research by Gilabert and Munoz (2010), Kormos and Trebits (2011) as well as Awaad and Tavakoli (2022), who found a relationship between higher WMC and certain aspects of speech production (Chapter 2, Literature review, Section 2.3.3).

To summarise briefly, it is important to note that in the current study:

- General language proficiency was operationalised using the Quick Placement Test (QPT), aligning its proficiency levels with the CEFR. For example, students scoring 0-17 were classified as beginners (A1 level).
- Vocabulary knowledge was measured through the Vocabulary Size Test (VST), focusing on participants' knowledge of the five most frequent levels of word families.
- FSs knowledge was measured using the Phrasal Vocabulary Size Test (PVST), which also assessed participants' receptive knowledge of the 5,000 most frequent word families.
- The familiarity of FSs was operationalised using the Test of Familiarity of FSs (TF-FSs), a simple test gauging participants' knowledge of FSs and their associated meanings, derived from the same chapter that the task was adapted from.
- Participants' WM was measured using the WM task from the Wechsler Adult Intelligence Scale IV (Wechsler, 2008), consisting of digit span forwards, digit span backwards and digit span sequencing.
- The questionnaire utilised was the Language History Questionnaire (Li et al., 2013), adapted and administered via a link to gain

insights into L2 learners' context, language use practices, proficiency in multiple languages and dominance of the learned languages.

A detailed explanation of these predictor variables, their implementation and the rationale behind their selection will be provided in Section 3.6.

RQ4a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, WM and the results of their personal background variables on the one hand, and the use of FSs on the other, in speech?

H4a.

- A strong correlation between vocabulary knowledge and FSs is expected.
- A slightly weaker correlation between general language proficiency and FSs is expected.
- The greater the learners' WM, the more FSs they will use in speaking.
- The more hours learners spend using everyday spoken English (gathered from the LHQ), the better they will be at using complex FSs.

These hypotheses are based on the same reasoning as in hypothesis 3a.

RQ4b) Which of the predictors (vocabulary test scores, WM test scores, personal background variables) best predict speaking task performance as measured through CAF scores for FSs?

H4b. The hours of English used in daily activities, as well as vocabulary test scores, are hypothesised to be the highest predictors of CAF scores for FSs. Again, these hypotheses are founded on the same rationale as in hypothesis 3a.

RQ5) What are the students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency?

3.4 Study design

As noted, the current study was designed to investigate the impact of both linguistic and cognitive variables on two aspects: the learning and production of CAF of entire speech, as well as the CAF of FSs. In order to gain a deeper insight into the relationship between the variables under investigation, answer the research questions from different perspectives, and establish research validity (Dörnyei, 2007), a mixed-method design was employed.

Combining a multi-method design in research is referred to as the triangulation technique (Dörnyei, 2007). This design involves the use of different data-gathering methods, and the insights obtained from one method can support those from another (Creswell, 2015). This technique is valuable in conducting any research because, according to Small (2011) and Creswell (2015), triangulation of data enhances the depth of the research design, emphasising that no single method of data collection is devoid of limitations. Thus, employing multiple methods can help mitigate or offset some of these limitation (Creswell et al., 2018).

Additionally, Small (2011) posited that adopting a mixed-method design, which integrates both quantitative and qualitative data within the same study, can enhance the robustness of the investigation. Hence, each method can provide insights into different stages of analysis (Creswell, 2015; Dörnyei, 2007). By utilising this multifaceted approach, findings become more reliable and credible (Dörnyei, 2007; Small, 2011). It was for these reasons that this study applied five mixed methods of data collection: speaking task, tests, the students' textbook, post-task interviews and questionnaires.

Since the current study had multiple independent variables, a multiple regression design was employed to investigate the predictive power of different variables on participants' speech performance in L2. This design is typically employed to identify how one or a combination of independent variables can explain variations in a dependent variable. The current study involved three dependent variables for each stage of analysis. These variables included the evaluation of complexity, accuracy and fluency in two aspects: learners' overall speech and the FSs found in their speech samples. General proficiency test, vocabulary tests, working memory test and variables obtained from background history questionnaire served as independent variables and were operationalised through a range of standard measures. Table 3.1 illustrates the relationships between the RQs, the chosen methods and the corresponding data collection tools.

Table 3.1. Associations between RQs, methods and data collection tools

Research question	**Method**	**Data collection tool**
1. Do ESL beginner college learners use formulaic language included within the curriculum and textbook when performing an oral proficiency task? If so, how do they use these FSs?	Comparison with textbook content (first three units) Interview through Teams	Students' text-book Post-speaking task (Appendix 1)

	Test of familiarity of FSs (FSs selected from textbook)	Checklist of FSs selected from textbook and test (Appendix 2)
2. To what extent is the CAF of students' entire speech sample related to their use of FSs in a speaking task?	Speaking performance via Teams	Speaking task (Appendix 3)
3. a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their personal background variables (LHQ) on the one hand, and CAF of the entire speech sample on the other?	Language tests (general proficiency test, vocabulary test, FSs test and test of familiarity of FSs) via FlexiQuiz WM test via Teams. Questionnaire via FlexiQuiz Speaking performance via Teams	Tests: QTP, NVLT, PVST, TF-FSs (Appendices 4, 5, 6 and 7) Wechsler Adult Intelligence Scale IV (Appendix 8) (Appendices 9 and 10) Speaking task (Appendix 3)
b) Which variables (general proficiency, vocabulary test scores, working memory test scores, personal background	Language tests (general proficiency test, vocabulary test, FSs test	Tests: QTP, NVLT, PVST, TF-FSs (Appendices 4, 5, 6 and 7)

variables) best predict speaking task performance as measured through CAF scores of the entire speech samples?	and test of familiarity of FSs) via FlexiQuiz WM test via Teams. Questionnaire via FlexiQuiz Speaking performance via Teams	 Wechsler Adult Intelligence Scale IV (Appendix 8) (Appendices 9 and 10) Speaking task (Appendix 3)
4. a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their personal background variables on the one hand, and the use of FSs on the other, in speech?	Language tests (general proficiency test, vocabulary test, FSs test and test of familiarity of FSs) via FlexiQuiz WM test via Teams. Questionnaire via FlexiQuiz Speaking performance via Teams	Tests: QTP, NVLT, PVST, TF-FSs (Appendices 4, 5, 6 and 7) Wechsler Adult Intelligence Scale IV (Appendix 8) (Appendices 9 and 10) Speaking task (Appendix 3)

b) Which of the predictors (general proficiency, vocabulary test scores, working memory test scores, personal background variables) best predict speaking task performance as measured through CAF scores for FSs?	Language tests (general proficiency test, vocabulary test, FSs test and test of familiarity of FSs) via FlexiQuiz WM test via Teams. Questionnaire via FlexiQuiz Speaking performance via Teams	Tests: QTP, NVLT, PVST, TF-FSs (Appendices 4, 5, 6 and 7) Wechsler Adult Intelligence Scale IV (Appendix 8) (Appendices 9 and 10) Speaking task (Appendix 3)
5. What are the students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency?	Interview through Teams	Post-speaking task (Appendix 1)

The design was counterbalanced between participants in order to minimise practise and order effects among the tests. It is important to note that I have varied the position of some critical tests (e.g., FSs). Additionally, the non-language task (WM) was used between the language tasks to reduce the possibility of order effects and to also avoid participants' boredom and fatigue of language tests. Therefore, participants were divided into three different groups which took the tasks in three different orders. No effects were evident in tests' order effects and performance. Table 3.2 demonstrates the three task orders in the study.

Table 3.2. Counterbalanced order of tasks

Order 1	General proficiency test	Working memory test	Vocabulary test	FS test
Order 2	Vocabulary test	FS test	Working memory test	General proficiency test
Order 3	FS test	Working memory test	General proficiency test	Vocabulary test

3.5 Population and sampling

Fifty-one Kuwaiti students, all female, from the Public Authority for Applied Education and Training (PAAET), specifically enrolled in the College of Business Studies (CBS) participated in this study. These students were pursuing a two-year diploma at CBS. All participants shared the same L1, namely Kuwaiti-Arabic. They had been studying English since the first grade and had not resided in an English-speaking country for more than six months before. At the CBS, depending on the schedule, students received either three sessions of English language per week, totalling around three hours, or two sessions a week, which also amounted to roughly three hours. These students were enrolled in their last English course before graduation, which is specifically designed for English for specific purposes (ESP) contexts. While it seems reasonable to argue that an ESP class cannot be taught at this level, there does appear to be a mismatch between curriculum design, desired student outcomes and students' actual proficiency levels, with textbooks and other material selection and design.

In short, the process of selecting the participants was based on the following criteria:

1. Participants were all females as they were drawn from the women's college where I was previously based before pursuing my PhD studies.
2. Participants were selected according to their level of education – they would have finished general English classes and are in their final specialised English class.

This pool of participants was selected because they were both approachable and aligned with the study's objectives (Punch, 2013). As mentioned earlier, these participants were selected from the same college where I was previously based before pursuing my PhD studies, making access to them facilitated by the help of colleagues.

Selection was applied to this population using a random technique, which is one form of probability sampling (Dörnyei, 2007). All students (N = 520) in the CBS enrolled in the final ESP course should have an equal chance to participate in the study. Thus, I took a list of all learners and selected 51 participants from this group using a simple random sampling - method. This was accomplished by utilising software, such as Excel, to generate random samples. This kind of sampling is useful for researchers as it enables them to make generalisations about the larger population (Cohen et al., 2017), especially when small groups like the sample in this study (51 students) have been used. Selection from CBS ensured that participants were homogeneous in terms of gender, native language and ethnic background.

3.6 Tasks and Materials

3.6.1 Speaking task

The main data collection started with the productive speaking task. As discussed in the previous chapter, Uchihara and Clenton (2018) employed speech samples ranging from one to four minutes, with total word production ranging from 81 to 319, aligning with the typical lengths observed in other L2 speech studies focusing on lexical performance (e.g., *Range* = 61 to 268 words in Saito, 2018). However, Uchihara and Clenton's (2018) study involved intermediate and advanced learners, whereas the present study is focused mainly on novice learners (CEFR A1– A2), who will need more time to formulate their utterances. Thus, each participant was asked to plan and perform an approximate four-minute personal task in a monologic format revolving around the topic of presenting a self-introduction in a work–life context (see Appendix 3). This topic has been adapted from the participants' textbook, specifically from a chapter that they had already encountered at the time of data collection (Units One, Two and Three).

3.6.2 Tests

In this section I will present an overview of the five types of tests that were adopted for this study: language proficiency test, vocabulary test, FS test, test of familiarity of FSs and WM test.

a) Proficiency test

The Quick Placement Test (QPT; Oxford University Press & Cambridge Local Examinations Syndicate, 2004) was used in this study to measure participants' level of language proficiency. The test consisted of 60 multiple-choice questions including vocabulary, grammar and reading, taking around 30 minutes to complete. The test booklet provided the Key (answers) and participants' level of proficiency based on scores they obtained. Level of proficiency is based on the CEFR. The students who scored 0-17 were considered as beginners (A1 level); students who scored 18–29 were considered as elementary (A2 level) and students who scored 30–39 were considered as lower-intermediate (B1 level) (see Appendix 4).

b) Vocabulary test

While this study focuses on FSs (which is a specific subcategory of vocabulary), it is important to obtain a general overview of informants' vocabulary knowledge. The Vocabulary Size Test (VST) (Nation & Beglar, 2007), for example, assesses vocabulary knowledge based on word family frequency created from the spoken subset of the British National Corpus (BNC) (Nation, 2012). However, the VST measures written receptive knowledge and this study focuses on speaking, so an oral productive task might be assumed to be more suitable. Nevertheless, production tests can be more useful in assessing more advanced L2 learners' proficiency (Read, 2019). Since I aim to assess beginner ESL learners who have limited knowledge of the language, receptive tests are likely to be more suitable for their level of proficiency as productive tests may be too demanding (Laufer & Goldstein, 2004).

The test consisted of five parts with 24 items at each level. It aimed to assess knowledge of the 5,000 most frequent word families, as well as a 30-item section (in Part 6) that assessed knowledge of the AWL through a multiple-choice format. Meara (2010, p. 5), however, claimed that if the participant is less advanced, then it would be "a waste of time" to ask them to complete all levels of the test. Since the piloting yielded several insights, including that participants had a range of different proficiencies (ranging from A1 to B1), the main data collection allowed participants to complete the entire test within the recommended time limit (30 minutes) (McLean and Kramer, 2015). This allowed participants to complete the maximum level they could reach to within the time limit without omitting any of the levels. Put simply, measuring knowledge of the five most frequent levels of word families may indicate the maximum range in vocabulary learning at least for most L2 learners (Webb & Sasao, 2013).

In each question, the participant had to decide upon the correct meaning in the L2 that corresponded to the word written in bold in the given sentence (see Appendix 5).

c) Formulaic sequences test

This study also planned to assess participants' knowledge of FSs receptively. According to Martinez (2011), the Phrasal Vocabulary Size Test (PVST) (Martinez, 2011) measures knowledge of FSs on the PHRASE List and could be used together with the VST (as mentioned earlier the NVLT is very similar to the VST; however, has been updated to suit the latest BNC/COCA list). The PVST was chosen because, to the researcher's knowledge, there are no other tests that measure FSs. However, the advantages of utilising the PVST were twofold. First, it focused on the same FSs included in the PHRASE List. Therefore, assessing learners on such knowledge will help me gain an overall idea of which frequency band to start with during the speaking task. Second, like the NVLT, it assessed receptive knowledge of the 5,000 most frequent word families, but with 10 items for each level. However, it is important to note that the FSs chosen in the PVST are not only based on word frequency but also on meaningfulness. For example, sequences like *is the* or *is of a*, which encode very little meaning in themselves, were not included in the test. Third, it is highly consistent with the format of the NVLT. The PVST has the same multiple-choice format and five levels. In each question, the participant had to decide upon the correct meaning in the L2 that corresponded to the FS written in bold in the given sentence (see Appendix 6).

d) Test of familiarity of FSs

The test of familiarity of FSs (TF-FSs) was a simple test used to gauge participants' knowledge of FSs and their associated meanings, all of which were extracted from the same chapter that the task was adapted from (see Appendix 2 for the FSs checklist). These FSs pertained to self-introduction in a business context, for instance, phrases like *I'm afraid I have to go*. I developed this test to be more specific to the domain of knowledge (expressions targeted by the textbook) that participants in this study should possess. To the best of my knowledge, this approach is not commonly found in the literature. However, it serves as a valuable addition to the PVST because it focuses on the FSs that these participants should already be familiar with from their textbook. The aim of this test was to determine if participants actually knew these FSs and if they had successfully acquired or learned them from the textbook. The rationale behind using this test is that it helps interpret the results by determining if participants' lack of use of these phrases in the speaking task is due to unfamiliarity or perhaps lack of practice. This understanding is important for explaining the outcomes of the speaking task.

TF-FSs comprised 14 FSs, for which participants had to decide the correct meaning in the L1 that corresponded to the FS written in bold (Appendix 7). The choice to provide multiple-choice responses in the L1 stemmed from a specific research objective: to ascertain whether the participants had genuinely encountered these FSs in their textbook. Since the majority of participants were beginner learners, the test aimed to reduce any possible confusion and enhance comprehension, particularly for those at lower proficiency levels, such as A1 levels according to the CEFR. In other words, the main aim of this test was to assess participants' understanding and recognition of FS.

e) Digit span tasks

WM span tasks have been shown to measure the storage and processing function of WM (Conway et al., 2005). The WM span task is important because it could show that learners with higher WMC can store and retain more FSs than learners with lower WMC. The current study used the WM task from the Wechsler Adult Intelligence Scale IV (Wechsler, 2008) to test participants' WM. Participants' short-term memory capacity was measured by the number, or span, of unrelated digits that can be remembered. These sets of unrelated items were presented in three parts: *digit span forwards* (DSF), where the test-taker was requested to repeat increasing spans of digits in the order they were presented; *digit span backwards* (DSB), where the test-taker was requested to repeat increasing spans of digits in reverse order; and *digit span sequencing* (DSS), where the test-taker was asked to repeat a sequence of numbers in ascending order. It is important to note that the test was conducted in an auditory mode.

Each task was administered separately, not only to avoid confusion but also to obtain distinct results to reflect the underlying cognitive processes. The DSF will assess the phonological WM, which is responsible for maintaining phonological and verbal data and is assumed to have a significant role in L2 vocabulary learning (Baddeley, 2003). On the other hand, DSB is designed to go beyond measuring the phonological WM memory (Kormos & Sáfár, 2008). It is claimed to measure the capacity of complex WM such as the functioning of the central executive, responsible for regulating the attentional control of WM (Hale, Hoeppner and Fiorello, 2002). Finally, the DSS is used to increase the WM demands of the task (Wechsler, 2008). All three subtests were administered to the participant regardless of their performance in any of the tests. Each test consisted of eight items and each item comprised two trials (i.e., 16 items altogether in each subtest). Appendix 8 shows a snapshot of a DSF task (Wechsler, 2014, p. 90).

The test was conducted in the participants' L1 (Arabic) to enhance the validity of the results, as participants learned to calculate in Arabic. Logically speaking, adding L2 competence to the test adds additional variables that may influence the results of WM tests (e.g., a participant scores poorly due to lack of proficiency of L2 numbers, as they learned to count in their L1), thereby invalidating any interpretations that are made on the basis of this result. Indeed, as participants' L2 competence is low, this might influence performance on L2 item repetition, as Kormos & Sáfár (2008) have indicated.

e) Scoring

To ensure accuracy of scoring, each participant's response was recorded via a digital voice recorder and on a piece of paper. For each trial, one point was given to each item if the participant provided a correct response. If she did not give a correct response or did not answer within around 30 seconds for each trial, she scored 0 points. The Digit Span total raw score was calculated by adding the Forwards (16 points), Backwards (16 points) and Sequencing (16 points) item scores, making a possible total 48 points. The test was discontinued after the participant had scored 0 on both trials of an item, and the last number of digits recalled determined her Longest Digit Span (LDS). For example, if the participant repeated seven digits forwards once and missed both trials on an eight-digit span, the LDS Forwards would be seven (Wechsler, 2014, p. 89). This means participants should repeat the highest number of sets of digits in at least 50 per cent of the cases (at least one time) (Kormos & Sáfár, 2008).

3.6.3 Textbook content

The study investigated the textbook (Unit One, Two and Three) used by the CBS for graduate business students, namely *Business Result* (Hughes & Naunton, 2017). This is a commercial textbook published by Oxford University Press. The textbook is wide-ranging, including all skills (listening, speaking, reading and writing). There are 15 units in the textbook, and each unit revolves around business/work-related topics. For example, the first unit is about *Working life* (outcomes include talking about oneself and one's work, showing interest during conversation, etc.); the second is titled *Work–life balance* (outcomes include talking about work–life balance, and exchanging contact information) (see Appendix 2 for the FSs checklist that are used in these chapters). It is important to consider that this textbook is designed for an intermediate level learner (at B1–B2 level according to the CEFR) whereas most students (including participants in the current study) using this book at the CBS are only beginners (at A1–A2 level). As already noted, this suggests a mismatch between the curriculum design and desired outcomes and the students' actual levels of proficiency.

3.6.4 Post-task interviews

The post-task interview was conducted immediately after the productive speaking task involving all 51 participants. The objectives were twofold: firstly, to probe into students' attitudes towards speaking in the classroom; and secondly to explore their views on vocabulary learning, more specifically, the learning of FSs learning in the classroom. Utilising questionnaire or interviews for immediate stimulated recall of perceptions is generally considered a reliable and valid technique to gather retrospective data. This approach involves retrieving information from participants' memory after the event, thus reducing the influence of memory decline (Dörnyei, 2007).

The post-task interview utilised a semi-structured format, chosen for its alignment with the study's data needs for addressing research questions. Brinkmann and Kvale (2015) highlighted that semi-structured interviews mimic conversations people have in daily life, making them ideal for capturing emotional, reactive and routine aspects of individual's experiences. The main goal of using these interviews was to explore participants' thoughts and feelings regarding speaking and vocabulary learning, with the possibility of unexpected insights to emerge from the discussions. It is worth mentioning that examining learners' perceptions regarding their speaking ability and vocabulary learning has not been very widely utilised in CAF research. However, the combination of participants' views on speaking with the detailed linguistic performance analyses applied in this study was designed to attain a deeper understanding than has been previously demonstrated, thereby enhancing the credibility of existing assumptions.

The interview questions were selected and categorised based on the themes discussed in the literature such as whether speaking is practised in the classroom, the specific focus during task performance, vocabulary learning methods, approaches to learning vocabulary words and FSs and if they encounter the new lexical item/expression after learning it. These questions were designed to elicit participants' perceptions and attitudes, in order to address the research questions. For instance, examples of such questions included 'How often do you practice speaking activities in the classroom? and 'Could you share what specific aspects or areas you concentrated on while performing the task?' (see Appendix 1)

The interviews comprised four main items with a total of nine follow up questions. Question 1 focused on gathering information about the speaking task they had completed and whether they had practiced similar speaking tasks in the classroom. Question 2 asked about their experiences with speaking activities in the classroom and their perceived relevance to an ESP course. Item 3 provided insights into vocabulary learning in the classroom, including the learning of individual words compared to formulaic sequences, the methods used to introduce new vocabulary to learners, and participants' views on the requirements of their current ESP course in terms of vocabulary learning. Lastly, Question 4 delved into their perceptions of how to enhance the specific English learning experience for L2 learners.

Given the semi-structured format of these interviews, these questions were intended to serve as prompts to encourage and guide participants discussions, rather than a rigid script to follow throughout the interview. As a result, not all questions were asked to every participant; instead, the follow-up questions were tailored to each individual's responses during the interview. It is worth noting that many of the participants were motivated and enthusiastic to share their perceptions during the interviews. They expressed their satisfaction with the opportunity to contribute to the improvement of ESP courses through this interview.

3.6.5 Questionnaire

The language history questionnaire (LHQ 3.0) (Li et al., 2013) was adapted and administered via the following link to gain insights into the L2 learners' context and practices of language use, proficiency (listening, speaking, reading and writing abilities) in multiple languages, and dominance of the learned languages (see Appendix 9). According to Li et al. (2013, p. 673), results from such questionnaires have often been utilised to "predict or correlate with learners' linguistic performance in cognitive and behavioural tests". Despite the fact that many of the linguistic background questionnaires (including the LHQ) share some common aspects and have similar question items, the LHQ had been chosen because it integrated different measures of language background, proficiency, usage and dominance. Initially, the LHQ was chosen as it offered a web-based questionnaire that enables researchers to gather and save LHQ outcomes in the cloud, and it can facilitate data collection and analyses in a language background and language assessment setting (Li et al., 2013). However, during the piloting, some major adjustments had to be made, as I noticed that some of the open-ended results were omitted from the web-based Excel spreadsheet that the LHQ produces as its data output and a question had to be added. For this reason, I decided to recreate and re-administer the LHQ using FlexiQuiz instead – which was also applied in the main data collection.

Additionally, based on the results of the piloting, minor changes were made to the wording of the LHQ after receiving feedback from participants about not understanding some of the Arabic translations of the English questions used in the original questionnaire (a feature that the LHQ provides).

The LHQ (3.0) consisted of 27 questions; however, as three questions were omitted (as they were unnecessary for my context) and one question was added, the modified LHQ contains only 25 (Appendix 10). The questionnaire included mainly closed questions, as this type of question is easier for participants. The closed questions were presented as multiple choice and Likert scale items. Questions 1–9 concerned basic demographic information of the participants (e.g., age) and their linguistic history (e.g., age of acquisition, years of language use, etc.) for each language that the participants learned. Questions 10–22 included information about dominance of the language learned based on participants' self-rating. This included information about language usage in different contexts and habits. Question 23 was the question added to the LHQ through its re-creation in FlexiQuiz. It was an open-ended question used to gain more insights in the benefits of repetition (one of the study's aims; Question 23 *If you watch TV or listen to the radio in your second language or any of the languages you have learned, do you repeat watching or listening to the same program more than once?*). The remaining two questions were open-ended (for the LHQ) allowing participants to add any further information regarding language background or usage.

It is important to note the omission of three questions from the questionnaire, justified by their irrelevance to the study's context. Specifically, Items 16 and 27 were excluded due to their focus on accents, which falls outside the scope of this study. Additionally, given Kuwait's status as a small, monolingual country with a near ubiquitous unified cultural background, question 24 which involved respondents' connection to different cultures and languages, was also removed.

3.7 Statistical power of the study

While the sample size was small, it is important to note that a power analysis was conducted to ensure the robustness of the statistical analyses in this study. GPower (Faul et al., 2009) was used to perform a post-hoc power analysis to estimate the power of the analysis for the current study's sample size. The analysis was carried out for a linear regression fixed model to calculate the power of each individually significant regression model with an alpha level of 0.05 and a sample size of 51. The results, as shown in Table 3.3, indicate a power above .80 for all measures. The suggests that a reliable level of confidence can be maintained in the findings from linear regression analysis, as presented in Table 3.3.

Table 3.3. Power calculation for all dependent variables

Complexity of speech	Power
General Proficiency	97%
Test of familiarity of FSs	99.8%
Number of FSs used	83.6%
Accuracy of speech	Power
General Proficiency	99%
Test of familiarity of FSs	97.3%
Number of FSs used	98.6%
Rate A	Power
General Proficiency	98.4%
Test of familiarity of FSs	99.6%
Number of FSs used	90.2%
Rate B	Power
General Proficiency	95.6%

Test of familiarity of FSs	99.9%
Number of FSs used	77.78%
Complexity of FSs	Power
General Proficiency	89.2%
Test of familiarity of FSs	98.4%
Number of FSs used	65.5%
Accuracy of FSs	Power
General Proficiency	90.65%
Test of familiarity of FSs	96.7%
Number of FSs used	>99.9%
Fluency of FSs	Power
General Proficiency	82%
Test of familiarity of FSs	96.8%
Number of FSs used	>99.9%

3.8 Pilot study

A pilot study was carried out to evaluate the "feasibility" and "usefulness" of the data collection instruments and make any required changes before employing those instruments with the study participants (Mackey & Gass, 2005). I regarded the pilot study as an essential opportunity to trial and enhance my data collection tools. The pilot study was conducted between 16 August 2020 and 20 September 2020, online, via Microsoft Teams and FlexiQuiz software (because of the Covid-19 curfew restrictions). Contact with potential participants at the CBS in Kuwait was achieved through former colleagues. The piloting consisted of a sample number of 12 ESP college participants who were taking their last English course before graduation. All participants responded enthusiastically to the opportunity of learning more English by participating in the study. These students took three separate days to complete all data instruments in the study. The data collection methods used were as follows: language proficiency test (QPT – Appendix 4), vocabulary test (NVLT – Appendix 5), FS test (PVST – Appendix 6), WM test (the Wechsler Adult Intelligence Scale IV – Appendix 8), test of familiarity of FSs (Appendix 7), speaking (role play) task (Appendix 11) and Language History Questionnaire (Appendices 9 and 10). It is important to note that the order of the tests (e.g. general proficiency test, vocabulary test and FS test) was counterbalanced to control for any test order and practice effect. For example, if a vocabulary test precedes the speaking task, participants will know vocabulary is an area of interest for the speaking task.

All tests were piloted, firstly to elicit information about the time it takes participants to complete each test, the clarity of instructions, and the point at which I should stop them during a test (if they were unable to continue because of level of difficulty). I also wanted to gain a more in-depth understanding of the applications and implications of the software, FlexiQuiz, through which the tests will be administered and scored (i.e., language proficiency test, vocabulary test, FS test, test of familiarity).

Second, the LHQ was piloted to participants in order to gain insights about their understanding of the items and the time it took them to complete it. For instance, as mentioned earlier, there was feedback from participants about not understanding some of the Arabic translations of the English questions used in the original questionnaire, so I changed the wording ready for the full study. I also wanted to take the opportunity to explore the LHQ web in terms of the way it collects data, and analysing that data. In addition, a major adjustment was made to the formatting of the questionnaire, as I noticed that some of the open-ended results were omitted from the web-based Excel spreadsheet that the LHQ produces as its data output, so I re-administered the LHQ using FlexiQuiz instead. After these preliminary alterations, I re-issued the questionnaire to all participants including those who had completed the original version (in terms of wording and format of questionnaire) to ensure that it was understandable. Positive feedback was received about the development of the adapted LHQ, such as “the language is clearer and the formatting is easier than the previous version”.

The remaining data collection instruments of the pilot study (i.e., speaking task and WM test) were administered and recorded online via Teams software (because of Covid-19 curfew restrictions). Participants of relatively the same age, but not necessarily the same level of proficiency (N = 12), were divided randomly into pairs (total six pairs) to perform the dialogic speaking task. Thus, pairs were asked to plan to perform an approximately four-minute role-play task revolving around a dialogue between two colleagues in a work–life context (see Appendix 11). This topic was adapted from the participants’ textbook, specifically from a chapter that they would have come across prior to the data collection (Unit One).

The results of the piloting data indicated that changing the modality of the speaking task from a dialogic format to a monologic one was more effective with my targeted sample of participants (i.e., mostly beginner L2 learners at A1–A2 level). This is because the dialogic speaking tasks in the pilot revealed some challenges, as I did not obtain many long stretches of speech from many of the participants (e.g., some participants with higher proficiencies took and maintained control over the conversational floor). This resulted in an uneven distribution of talk time between participants, especially related to those with different proficiency levels.

After the piloting, the data methods were revised as described and made ready, and the study was conducted in July 2021.

3.9 Ethical considerations

Permission was obtained from the Ethics Committee in the Department of Applied Linguistics before conducting the study (see Appendix 12). Prior to commencing the data collection, all participants, including those involved in the pilot study, received both verbal and written information about the purpose of the research (see Appendix 13). The information sheet, which was also translated into the participants' L1 to ensure comprehension, explained the purpose of the research and explicitly informed them of their right to withdraw from the study at any time (see Appendix 14). Moreover, they were assured that their identities and identifying information would not be revealed, and the collected data would be securely stored and used for research purposes only. Consent was obtained from all participants (see Appendices 15 and 16), and before initiating the data collection, each participant was again asked individually to re-confirm their consent verbally (which was also recorded prior to the speaking task). Both Cohen et al. (2017) and Punch (2013) consider requesting permission as protecting the participants' human rights.

3.10 Data collection

Data collection was carried out in July 2021, which coincided with the end of the students' second semester, which was on-line in the given context. Given that this study required each participant to engage in multiple data collection, i.e., speaking task, linguistic and cognitive tests, post-task interview and questionnaire, it was not feasible to conduct all of them in a single session. Therefore, the data collection process was organised into four separate days for each participant. The data collection sequence began with the speaking task. Subsequently, all participants immediately proceeded to the post-task interview, followed by the completion of the test of familiarity of FSs. The remaining tests were gathered during two separate sessions, with each session occurring on a different day. Finally, the data collection culminated with the questionnaire, which was administered on the last day of data collection. This decision to extend the data collection across four days was made with the goal of mitigating participant fatigue, which could negatively affect their performance. The adjustment was in response to feedback from participants during the pilot study when data collection was conducted over three days, with participants finding it long and tiring.

The fifty-one participants received a brief about the study from their classroom teacher. All participants were interviewed and recorded individually online via Microsoft Teams. It is important to note that no major technological issues were encountered during conducting the study. However, there were minor glitches, such as occasional connection issues, leading to small recording errors. These technical difficulties were mitigated by the ability to contact participants outside of Teams and their extended availability after the completion of the study. Any confusion that arose from these issues was resolved through clarification from the participants. Overall, the use of Microsoft Teams for data collection was effective for my study's purposes.

As the session began, I greeted each participant, introduced myself and engaged in a two minute chat to help her relax. Following that, I confirmed the participant's willingness to take part in the study and made sure whether she was aware of her right to withdraw from the study at any time. The researcher guaranteed the confidentiality of the participant's involvement and the anonymity of her identity. Each participant read the information sheet and signed the consent form. Once both their written and verbal consent were obtained, they proceeded with the speaking task.

Each participant was shown the task prompt on screen; the prompt was available throughout until the actual performance. I read the task in L2 and then provided an L1 explanation in order to ensure comprehension (Storch & Wigglesworth, 2003). Before starting the task, the participant was permitted to ask any questions and was then given planning time. Based on the findings of my pilot study, 15 minutes of planning time seemed too long, as even with such an extended period, lower-proficiency participants (at A1 level according to CEFR) did not produce many long stretches of speech (whereas some participants with higher proficiencies would take and maintain control over the conversational floor). As Skehan (2018) noted that at least for intermediate level, increased time does not lead to higher performance (see Chapter 2, Literature review, Section 2.4.3), a pattern reflected in the pilot study's results with the current. Since this study is interested in investigating participants' maximum potential – that is, what they can do with a bit of help and time – it is relevant to ascertain whether learners use what they have learned in class and from the book, with a specific focus on their use of FSs, when they are planning for the given oral task. Therefore, participants were given five minutes of planning time. During planning time, I kept sound from my side on mute in order to avoid distractions.

The reason for starting with the productive task was not to alert participants to the aim of the study (use of vocabulary and specifically FS). The findings would then be an artefact of the fact that I have alerted them to the importance of vocabulary/FSs, and would not be a valid reflection of students' everyday spontaneous use of language/FSs. In the current study, it is important to allow participants to carry out the task as they would normally do.

Following the completion of the speaking task, I immediately conducted the post-task interview. Arabic was chosen as the language used in the interviews not only to accommodate their beginner-level proficiency but also to create a comfortable environment for them to converse in their mother tongue.

I started the interviews with a set of broad, pre-prepared questions to encourage participants to express their thoughts freely. These questions acted as catalysts for discussion, ensuring the interview remained focused. It is important to note that these questions did not rigidly dictate the flow of the interview; rather, they served to stimulate ideas and maintain focus if participants deviated from the main discussion points. This enabled me to delve deeper into issues they found particularly important, as it facilitated a more profound investigation (Dörnyei, 2007). In other words, semi-structured interviews provided both flexibility and control in the data collection process.

An expected challenge prior to condcuting the interviews was the possibility that participants' might be hesitant to provide open responses. This hesitancy could be attributed to their apprehension about sharing sensitive information with a colleague and may be further amplified by shnyess, potentially resulting in responses that reflect an idealised scenario that is not representative of the current situation (Dörnyei, 2007). To mitigate this concern, I strategically planned for an informal conversation at the beginning of the interview. Additionally, participants were reminded of data confidentiality and encouraged to express their opinions freely, as pseudonyms would be employed to safeguard their identities.

At the end of the post-task interview, I asked participants whether they were familiar with the fixed expressions and their meanings that were presented in the same chapter that the task was adapted from. To investigate students' knowledge of FSs I presented them with a test (i.e., TF-FSs). Participants were told not to guess, but choose the fourth option, "I don't know", if they did not know the correct meaning. The test was administered via FlexiQuiz and was graded automatically via the software.

Following that, all language tests, including the language proficiency test, vocabulary test and FS tests, were administered via FlexiQuiz and were graded automatically via the software. As for the WM test, participants were tested individually in a quiet room via Microsoft Teams software and test material was recorded via digital voice recorder. I followed the Wechsler (2014) guidance for administering the test.

After performing all the tasks, each participant completed a background questionnaire. The use of FlexiQuiz allowed for the omission of unnecessary questions and for changes of questions in the questionnaire (as mentioned previously with regards to wording only). After the completion of the questionnaire format, I sent the URL onto participants to complete the LHQ and instructions were given for completing it using the FlexiQuiz software. Findings from the LHQ were manually coded and transcribed in an Excel spreadsheet.

3.11 Data analysis

In this section, the data analysis procedures are described and discussed, starting with how FSs were elicited from participants' speech samples and textbook. This is followed by the exploration of the process for identifying FSs through the use of TAALES software, while also addressing the issues and problems encountered in this process. After that, a detailed discussion of the novel aspect in this study – the measures of complexity of FSs – will be provided.

3.11.1 Identification of FSs elicited via participants' speech samples and textbook

Following Carroll and Conklin (2020) as well as Wood (2020) any word combination that had an MI value of at least 3.0 was considered as formulaic (Chapter 2, Literature Review, Section 2.2.1). Additionally, a checklist was designed to ascertain whether these sequences are indeed FSs on the basis of some of the identification criteria found in the literature, such as frequency of occurrence (i.e., following Wood (2020), using frequency cut-off of a minimum of ten occurrences per million words (e.g. the FS *looking forward to* had 13.75 occurrences per million words in the Spoken COCA, which qualified the chunk as frequent and thus a FS) (Martinez & Schmitt, 2012; Nation, 2001; Nation & Waring, 1997); semantic analysis (i.e., idiomaticity – selection of FSs on the grounds that they are "deceptively transparent"; e.g., the FS *in charge of* may be misinterpreted with a price asked for goods/services); and utility – i.e., selection of chunks according to their usefulness to the needs of learners (Durrant & Mathews-Aydınlı, 2011; Thornbury, 2019), (e.g., the FS *nice to* *meet you* is a common phrase used when meeting someone for the first time, especially in a business context) (see Section 2.2.1). Note that not all criteria must be present in the sequence for it to be identified as formulaic (Wood, 2020).

The following subsection provides a detailed explanation of the data collection process for MI scores of FSs in both speech samples and the textbook.

3.11.2 Identifying FSs - using TAALES

It is worth noting that the MI scores of multiword units in this study were collected using TAALES (2.2) – Tool for the Automatic Analysis of Lexical Sophistication (Kyle & Crossley, 2015; Kyle et al., 2018). TAALES was chosen because it is an advanced piece of software that automatically computes an MI score (derived from COCA) for extended multi-word units (or trigrams as it is referred to in the software) in addition to providing several facets of lexical sophistication, including phrase frequency. While the textbook used in the current study follows British English norms, I have opted to use the spoken subsection of the Corpus of Contemporary American English (COCA: Davies, 2009) as a reference corpus in order to calculate FSs indices. This is because the COCA is incorporated by TAALES. However, the use of COCA should not be problematic as its corpus of the spoken genre is large, consisting of 79 million words (Tavakoli and Uchihara, 2020) taken from spontaneous conversations (from a wide range of TV and radio programs in the United States). A similar decision was also made by Tavakoli and Uchihara (2020) based on the size of the corpus.

Section 2.2 examined the benefits of the PHRASE list, which was advocated as a valid source of expressions to teach and test, compiled with pedagogic purposes in mind, and especially relevant to the targeted pool of participants (mostly at A1–A2 level according to the CEFR) as 95 per cent of the expression words occurred in the top 1,000 frequent words and 2.88 per cent in the second top 1,000 frequent words – both written and spoken – in the British National Corpus (BNC) (Martinez and Schmitt, 2012). For this reason, the use of the

PHRASE list for comparing and contrasting with expressions in the students' textbook was considered. This approach would facilitate the assessment of complex FSs, by knowing which FS is more or less complex than the other. This made the scoring of FSs elicited from speech samples and the textbook straightforward when they overlapped with Martinez and Schmit's PHRASE list, such as *in charge*. However, many FSs extracted from both the textbook and participants did not align with the PHRASE list. This discrepancy can be attributed to the fact that the FSs listed in the PHRASE list encompass general spoken language, including expressions like *take place* and *make sure* (see Appendix 17 for more examples), rather than the specific ESP business FSs targeted in this study, such as *nice to meet you*. Consequently, the PHRASE list became an unsuitable reference for assessing the complexity of FSs in this study.

Before carrying out the pilot study, checking whether any of the FSs in the textbook (e.g., in the "Key expressions" box, up to a chapter relevant for the learners, Unit 3) were listed in the PHRASE list did not seem challenging. However, once the work began, challenges started to appear. As a result, I had to manage the following criteria and procedures to overcome all the problems that were occurring. Firstly, the formulas in the current textbook are presented in a full expression/utterance form (including function words, e.g., the subject *I*), whereas the PHRASE list only gives the formulaic chunk, omitting the remainder of the utterance and function words (at least in most instances). For example, *look forward to* appeared as it is (formulaic chunk) in the list, but was inserted in full expression form in the textbook – *I look forward to hearing from you*. Clearly, the representation of the expressions in the textbook may not seem appropriate as the pronoun can vary, in addition to the remainder of the chunk. For example, the FS *look forward to* can be followed by several plausible verb options and not only the verb *hearing* (such as *look forward to seeing X*, *look forward to meeting X*, etc). The MI measure was needed to identify and extract multi-word unit (MWU) combinations embedded within the full expressions provided by the book – which is another reason why TAALES is used in this study (as TAALES extract multiword units automatically). It is important to note that in many instances the MI of subject + auxiliary of MWUs was not included, such as in instances like *you can depend on me* – where only the MI score of *depend* + *on* was measured. This is because the subject and auxiliary can be varied systematically (the auxiliary could be replaced with "could", "may", etc.) but the core is *depend on*.

Second, many of the FSs found in the textbook consisted of more than two words, such as *keep in contact*. This therefore created another challenge when searching for the MI score of the whole phrase, namely, the transitions of the whole multi-word unit (e.g., *keep* + *in* + *contact*). Again, this problem was tackled by the use of TAALES 2.2 as it was possible with this software to obtain a list of extended multi-word units that were found in the data set with their relevant MI scores.

Finally, all speech samples were checked manually for multi-word units uttered with errors (i.e., inaccurately) by participants. This is because erroneous FSs could not be extracted with TAALES, as this software can only find error-free examples. Of course, these erroneous multiword units (e.g., *thanks God*) were corrected and calculated again separately in TAALES to detect whether they qualified as an FS (based on a minimum MI score of 3).

After identifying FSs in both speech samples and the textbook, I was able to measure the CAF of FSs. It is important to note that the measurement of accuracy and fluency of FSs has already been discussed in the previous chapter, as they are established measures taken from the literature. However, as highlighted in the literature review, the measurement of complexity of FSs has not been explored before. The next section details the novel aspect of this study–the measurement of complexity of FSs.

3.11.3 Complexity score of FS

Complexity was measured in terms of lexical richness and syntactic complexity (Boers et al., 2006). Thus, FSs were counted in terms of types rather than tokens. To my knowledge, the measurement of complexity of entire speech is common and standardised in the literature but it is rare to look at the complexity of the FS itself. As was mentioned in the literature review chapter, Boers' et al.'s (2006) study focused on identifying FSs based on fluency and accuracy of FSs, without considering the measurement of complexity of FSs, such as assessing some as more complex than others, based on criteria like length and frequency (as demonstrated by Norris & Ortega, 2009). However, one of the aims of the current study is to measure the complexity of FSs found in a learners' textbook and speech samples by considering a range of lexical and syntactic indicators. As such, I had to design complexity-of-FS measures based on certain criteria in the literature. In doing so, I aim to tap into different dimensions of complexity and to provide further insights into the multidimensional aspects of complexity. The following details the novel aspects of my research on measures used to gauge the complexity of FSs demonstrated in the students' textbook and speech samples.

Firstly, counting the number of morphemes that comprise a FS is an important indicator as it reveals inflectional and derivational complexity, which therefore adds to the level of syntactic complexity (Lahmann, Steinkrauss & Schmid, 2019). However, counting the number of morphemes in words can be problematic for a number of reasons. Most notably, there is a lack of agreement within the field with regard to the methodology of counting morphemes and further dispute as to what constitutes a morpheme in and of itself and within the learner's mind (e.g., *ation*, *ism* and *ist)*, especially with low-proficiency learners who have limited knowledge of morphology and may have learned these morphemes as part of chunks and consider them as units rather than multimorphemic complex words. Furthermore, this lack of consensus within the field is reflected in the challenges of morphemic analysis. For example, dealing with French, Latin or Greek roots/affixes, such as the affix -*ition* in the word *nutrition*, can be difficult to analyse, as it may not be clear whether the affix is -*ition* or -*ion* (Bauer & Nation, 1993). Additionally, there appears to be lack of accurate tools that can be used for the purpose of calculating the number of morphemes within a word, unlike with syllables. In addition, As such this study relies on counting the number of syllables per word, and the FS as a whole, as part of the method for ascertaining the complexity of FS. This was decided based on the fact that the methodology for the calculation of syllables is generally agreed upon throughout the field of CAF, with available tools that can easily and accurately measure the number of syllables per word. Counting syllables provides a more reliable and accurate measure for length than counting morphemes. As a result, it contributes to the calculation for a complexity composite score for FS available to the study. However, Norris and Ortega (2009) proposed that measures of length are instead only basic indicators of complexity as they do not notify us of the type of complexity. Thus, a number of different measures were used to investigate the type of complexity, which is related to different linguistic aspects including idiomaticity (i.e., transparency).

Semantically transparent FSs are phrasal expressions that a learner is likely to understand; for example, *at this time*, as the meaning is compositional, that is, it can be derived from the component parts – (at+this+time), equivalent to "now" (Martinez & Schmitt, 2012). On the other hand, non-transparent FSs – or as Laufer (1989) terms them, "deceptively transparent" – are expressions where meaning is non-compositional – the meaning cannot be derived from the components; words learners assume they understand but they misunderstand. For instance, *for some time* may be misunderstood as *a short amount of time* (Martinez & Schmitt, 2012) whereas it means for an extended period of time. It follows that semantically transparent FSs will be regarded as less complex than deceptively transparent ones (see Chapter 2, Literature review, Section 2.2.1). If the FS is deceptively transparent, a score was earned and the FS coded with 1. If the FS is semantically transparent, no score was added and the FS therefore coded with 0. The same FSs were re-evaluated according to their transparency by a second rater.

Several authors regard native speaker intuition as a useful tool for judging FSs in spoken data, or in smaller or particular data sets (e.g., Martinez & Schmitt, 2012; Wood, 2020). For this reason, a monolingual English speaker who is also an Applied Linguist, provided judgements about the complexity of the FS on a six-point Likert scale. Such a scale allowing the possibility of high measurement precision (see Nemoto & Beglar, 2014, for a detailed analysis concerning the advantages of a six-point Likert scale).

All scores were added for the mentioned indicators of complexity (i.e., word length, semantic transparency and monolingual English-speaker judgement), which made up a final complexity composite score (from 1 to any) for each FS. To reiterate, the computation of complexity scores for FSs is rare or even absent in the literature, as to my knowledge considering how complex a FS has not been measured yet. As a result, the novel method of calculating FS complexity in this study was designed to fill this gap in the literature and enable this study to identify the complexity of FSs and its subsequent influence on the complexity of speech in general.

Another dimension of FSs in the textbook and speech samples is frequency. Frequency analysis was obtained by using the spoken part of the COCA as it is a large corpus (Kyle et al., 2018; Tavakoli & Uchihara, 2020) taken from spontaneous conversations (from wide range of TV and radio programs in the United States) (Tavakoli & Uchihara, 2020). According to Wood (2020), using a frequency cut-off of a minimum of ten occurrences per million words and with a MI score of no less than 3.0 detects the frequency of word strings in speech, and words with a high frequency of co-occurrence. The normalised frequencies of FSs (per million words) that specifically occur in a spoken discourse were taken and listed in the spreadsheet. Although the frequency of the FS itself is a good indicator of its complexity (i.e., the more frequent ones are less complex), one of the limitations of computer searches is that they may fail to detect frequent sequences (Durrant & Mathews-Aydınlı, 2011). For example, Wood (2020, p. 38) noted that some word strings such as *in spite of* or *how are you,* which seem formulaic with regards to both "saliency" and "unitary" meaning or function, might not occur in high frequencies in any given corpus. This may depend on what is in the corpus (cultural differences between the participants' background and register and what is available in the corpus), certain phrases being common in some conversations, but not others, or it might be that a certain spoken frequency corpus may not contain many conversations where people get to meet a person for the first time. Asking such direct questions may also not be common covered in some corpora, which means that the particular FS is not that frequent. However, this is not the case with the spoken COCA corpora as it reports multi-word units such as *how are you* as frequent, with a frequency of 52.86 (per million) and 116.67 in the TV/Movie subsection. Therefore, it may also seem reasonable to argue that the textbook does not reflect the frequency of patterns found in real life data. This lack of reflection of real life is likely to be one of the contributing factors to why I have found such low occurrences of strings including *where are you from* (1.49 in spoken and 5.50 in TV/M occurrences per million words) and *where do you live* (0. 35 in spoken and 1.92 in TV/M occurrences per million words), which appear to have a low frequency of occurrence in the corpus as a phrase yet would not be considered complex. Therefore, because of this frequency analysis complication it was detached from the complexity composite scores assembled in both the pilot study and the main data collection.

3.12 Data transcription and coding

In this section, I describe the methodologies used to analyse the data gathered from both the oral monologic task and post-task interview. This includes the transcription, coding and annotation of the data. Furthermore, I provide a detailed description of the data preparation process for statistical analysis.

3.12.1 Simple transcriptions

As previously mentioned, I obtained speech samples from the participants by recording their monologic task performances. In the initial stage, all recordings were transcribed manually into a Word document. The transcription process required frequent playback and rewinding of the recorded data. As the participants were mainly beginners, I did not require the option to slow down the recording, as their speech rate was generally manageable for transcription. For each participant, a separate Word document was created, containing simple transcriptions of their L2 speech and post-task interview samples. When conducting these simple transcriptions for the speaking task performance, the following guidelines were followed:

- Transcriptions began with the first utterance spoken by the participant, excluding any initial silences from the analysis.
- Punctuation marks were omitted, and all utterances, including filled pauses, were transcribed.
- Occasional use of fillers such as ‘I guess’ and ‘so’, were categorised as filled pauses, as these did not carry contextual meaning.
- Only two words, ‘ahh’ and ‘umm’, were used to represent all filled pauses.

After the simple transcriptions were finalised, the Word files were copied and transferred to a separate folder, where these duplicate copies were utilised for further analysis. Keeping the simple transcriptions in this way would benefit the researcher in later stages by facilitating syllable counts, CAF analysis, and the examination of post-task interview responses for further analysis.

3.12.2 Data coding

A qualitative approach was applied to code the FSs found in the first three chapters of the textbook, identified based on their MI score as discussed in Section 3.10.1. Additionally, this approach was used to explore responses provided by the participants during the post-task interviews in order to answer the research questions. Additionally, coding was utilised to examine further comments made by the respondents in the open-ended questions within the questionnaires.

Coding extends beyond assigning a digit to participants' responses; hence, I chose to employ thematic analysis for this study. This choice is highly applicable to the current study for several reasons. Thematic analysis includes processes such as transcribing the data, identifying and coding developing themes within the data, and then categorizing the answers into meaningful classifications (Cohen et al., 2017). These themes are then compared and contrasted with each other (Guest et al., 2012).

The key aspect of coding is to define each variable and list coding descriptions for every possible value that the specific variable can have (Dörnyei, 2007). The coding process, in this context, involves summarising data that were produced from participants' responses into numerical values in order to develop meaningful classifications (Cohen et al., 2017; Guest et al., 2012). Example of code aggregation to generate theme is provided in Figure 3.1. These classifications are then investigated via statistical analyses (Dörnyei, 2007).

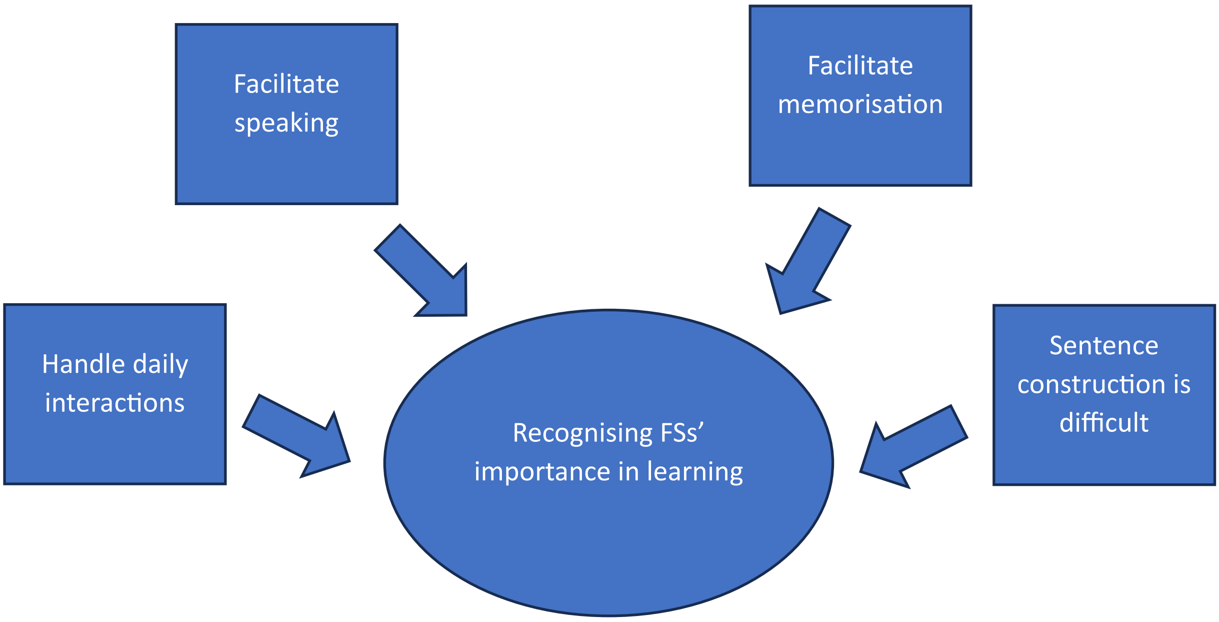


Figure 3.1 Example of code aggregation

3.12.3 Identification of speech aspects: using Audacity

The speaking task performances were analysed using Audacity software (Version 2.4.2.0; Audacity Team, 2020) and saved as MP3 files. The Audacity program was used to measure temporal aspects of fluency: phonation time (for each participant), length of pause, speech rate, and articulation rate. This temporal analysis, such as the duration of speech and pauses, supplemented the data found in the transcripts and allowed the calculation of speed and composite measures (Rate A and Rate B). This process was repeated three times to ensure the consistency of results.

3.12.4 Syllable counts

In this study, syllable counts were used as the basic unit of counting (part of the Rate A, Rate B and complexity of FS measures). In order to conduct the syllable counts, the transcriptions were reviewed and the texts were cleared of any filled pauses or punctuation marks, if present. Following Thai and Boers (2016), a syllable counter website, syllablecount.com, was utilised to do the counts in participants' speech samples. To ensure the accuracy of the counts, 10 per cent of the samples were double-checked through manual counting. The agreement between the manually counted syllables and the counts provided by the tool was found to be above .98 (Cohen's Kappa).

3.12.5 Combining all scores for statistical analysis

The final step of the data analysis involved compiling all the analysis and measures into a single spreadsheet to prepare the data for statistical analysis. Throughout this process, the majority of the information needed for the CAF measures, including both the entire speech and CAF of FSs, tests scores and background variables, was computed directly on the Excel spreadsheet. The Excel file was the most convenient platform for these calculations, with specific formulas used as needed. For instance, articulation rate was computed using the formula: (number of syllables x 60) / total speech time (excluding pauses) in seconds. Similarly, the transfer of data from the syllable count website to Excel was a straightforward process.

For other measures that required evaluation by a rater, such as the complexity of FSs assessed through native speaker judgments and the re-evaluation of FSs based on their transparency, which was another component of the FS complexity measure, the ratings provided by the rater were recorded by copying and pasting the scores onto an Excel spreadsheet. The process of copying and pasting the scores into the Excel spreadsheet was then double-checked to ensure accuracy and prevent any potential errors.

Finally, all the scores within the Excel file were transferred to R Studio for further statistical analysis.

3.13 Analytical procedures aligned with research questions

As was mentioned before, a qualitative approach was employed to code the FSs found in the first three chapters of the textbook. This coding method was applied to answer RQ1, which involved observing participants' use of FSs taught in their curriculum via speaking performance. Additionally, it was utilised to address RQ5, focusing on participants' perceptions of their ESP experience. Furthermore, it was used to explore further comments that the respondents made in the open-ended questions in the questionnaire.

In order to answer RQ2, I segmented, coded and scored the transcribed dialogue based on the measures chosen for evaluating the CAF triad of entire speech (see Section 2.3.6). Measuring oral proficiency can be extremely demanding because of the difficulty in ensuring reliable scores (e.g., in gaining consistent scores from separate judges) (Boers et al., 2006). To guarantee that the segmentation and scoring of the transcripts as well as the CAF scores in FSs were reliable, 10 per cent of the data were segmented, coded and scored blind by a second expert.

In order to answer RQ3a and RQ4a, a correlation matrix was created to investigate the correlation between the variables used. Two types of correlation analysis tests are available: the Pearson correlation coefficient (r), which is commonly used for normally distributed data, and Spearman's rho, which is more applicable for data that are not normally distributed. Pearson's r correlation coefficients was computed to assess the correlation between pairs of variables. All scores were normally distributed, except those for the FS test ($p = .011$ based on the Shapiro-Wilk test) and the vocabulary test ($p = .046$ based on the Shapiro-Wilk test). I will return to this issue in the next chapter.

In order to answer RQ3b and RQ4b, the correlation analysis was followed by a regression analysis, which is a standard way of modelling quantitative relationships between observed variables (Cohen et al., 2017). A correlation was used at the first stage to determine whether, and to what extent, there is a relationship (or degree of association) between two variables (Cohen et al., 2017); thus if the variables correlate, a regression analysis can be conducted (the second stage). Linear regression analysis was applied to relate X to Y through an equation of the form: Y = a + bX (Y dependent variable, a=intercept, b=slope, X-independent variable); in other words, the slope tells us how the dependent variable (Y) changes for every one unit increase in the independent (X) variable on average (Cohen et al., 2017). The main outcomes of interest (C, A and F of the entire speech sample for RQ3b and C, A and F of FSs in speaking tasks for RQ4b) are continuous scores, so a linear regression analysis was used. In order to be able to compare the effect of different predictors on the outcomes of interest, all variables were standardised (Z-scores were computed as these predictors were measured on different scales previously) (Field, 2013). As a result, significant independent variables from univariate analysis were included in the multivariate analysis. In other words, the analysis was extended from a univariate case to the multivariate. Thus, multivariate linear regression analysis was conducted to evaluate the adjusted association between the independent variables and the dependent variables to explore which predictor had a significant effect on the dependent variable. However, because there are many variables that affect CAF (as will be shown in the regression models) and given limitations of space, this review does not cover the univariate analysis and multivariate analysis in detail. All univariate regression analyses and multivariate regression analysis associating with CAF results can be found in Appendices 18 and 19, 20, 21 and 22. Following the univariate regression and multivariate regression, the analysis will be extended to include a hierarchical linear regression model. Thus, results for RQ3b and RQ4b will directly start with the hierarchical linear regression analysis.

This type of model uses a nested hierarchical form (Winter, 2019), a hierarchical design being one which involves variables of interest being combined and/or used at different stages of the model to show how one level of the model can influence the other (Cohen et al., 2017). In this study hierarchical analysis is only applied to RQ3b and RQ4b, specifically to ascertain whether newly added variables indicate a significant improvement in R^2 (the amount of variance in the dependent variable explained by the model). As a follow-up analysis, the hierarchical regression analysis was conducted in the opposite direction (studying the effect of Z while controlling for X) to examine whether both X and Z make an independent contribution to the model (Appendices 23 and 24).

The alpha level for all multiple statistical tests of significance was set by using Bonferroni's adjustment in order to decrease the possibility of a Type 1 error. This adjusted level was obtained by dividing the alpha level (.05) by the number of times the test was computed. Because this study includes a number of multiple comparison tests, different Bonferroni alpha levels were used for different multiple statistical tests, depending on the number of models. For RQ3b in the hierarchical linear regression analyses, an alpha level of $p = .01$ (.01 = .05/5 number of models) was set for C-SA; $p = .012$ (.012 = .05/4 number of models) for both A-SA and F-SA. Finally, for RQ4b in the hierarchical linear regression analysis an alpha level of $p = .012$ (.012 = .05/4 number of models) was set for C-FS, and an alpha level of $p = .016$ (.016 = .05/3 number of models) for both A-FS and F-FS.

3.14 Inter-coder/rater reliability

To ensure the reliability of the coding and the data analysis, a 10% subset of the entire dataset was re-analysed by a second rater. This re-analysis occurred at various stages of the data analysis process. To start with, the researcher re-evaluated all measures of the two CAFs: entire speech and FSs for all 51 participants. This procedure was repeated three times to ensure the consistency of results.

Subsequently, scores related to the complexity measures of FSs, such as native speaker judgments, were cross-checked with another monolingual English speaker, achieving an inter-rater reliability correlation of 94%. The transparency of complex FSs was verified through cross-checking between the researcher's scores and those of the monolingual English speaker, revealing a high Pearson correlation coefficient of 92%.

After that, 10% of the complexity, accuracy and fluency measures for both entire speech and FSs were cross-checked by a monolingual English expert, an applied linguist. The Pearson correlation coefficient demonstrated high agreement between the researcher and the rater for entire speech measures: complexity (88%), accuracy (92%) and fluency (94%), as well as for FSs measures: complexity (95%), accuracy (90%) and fluency (95%).

The syllable counts in English, conducted through an online tool, were manually double-checked by the researcher for 10% of the data, resulting in 100% agreement between the researcher's and the tool's calculations (Cohen's Kappa of 1.00). The achieved high inter-rater reliability confirmed the consistency and robustness of the data transcription, coding and scoring procedures, providing the researcher with confidence to proceed with data analysis.

3.15 Summary of the chapter

This chapter provided a detailed overview of the study, including its aims, research questions and hypothesis, study design, participants selection and the research instruments used for data collection, with explanations for their choices. An account of the pilot study was provided, followed by the ethical considerations that were applied to the study. The processes of the data collection, transcription, coding and analysis have been discussed, evaluated and justified. The novelty and original contribution of this study, which aimed to operationalise and measure the complexity of FSs, have been discussed. A rationale behind the operationalisation was provided, with the hope of developing a framework for assessing the complexity FSs. Finally, the chapter concluded with a detailed description of how the data were analysed and prepared for statistical analysis. The forthcoming chapter will present the results of the study, using both descriptive and inferential analyses to address the research questions.

Chapter 4 Results

4.1. Introduction

This chapter provides an in-depth discussion of the statistical analyses carried out in order to answer the research questions (RQs). The findings are presented with respect to each RQ.

1. Do ESL beginner college learners use formulaic language included within the curriculum and textbook when performing an oral proficiency task? If so, how do they use these FSs?

2. To what extent is the CAF of students' entire speech sample related to their use of FSs in a speaking task?

3. a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their personal background variables on the one hand, and CAF of the entire speech sample on the other?

   b) Which variables (general language proficiency, vocabulary test scores, working memory test scores, personal background variables) best predict speaking task performance as measured through CAF scores of the entire speech samples?

4. a) Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and personal background variables on the one hand, and CAF(FSs) on the other?

   b) Which of the predictors (general language proficiency, vocabulary test scores, working memory test scores, personal background variables) best predict speaking task performance as measured through CAF scores for FSs?

5. What are the students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency?

Section 4.2 of this chapter details various analyses carried out in order to check whether a number of statistical assumptions for these analyses were met, particularly the evaluation of the linear regression analysis. Section 4.3 includes a narrative describing the descriptive results of this study, including the language background tests, the cognitive tasks (WM tasks), and the duration of participants' speech samples, before reporting the results of the five RQs. This chapter ends with a summary of the key findings in Section 4.5.

4.2 Evaluation of assumptions

Before running both the univariate regression analysis and the multivariate regression analysis, I ran a number of analyses to check the fulfilment of assumptions for these analyses. First of all, the distribution of the scores on the independent variables (IVs) and the dependent variables (DVs) was checked by exploring normality with plots (histogram plot) and conducting normality tests (both the Shapiro–Wilk test and Kolmogorov–Smirnov test). It was found that some of the scores (i.e., complexity-SA, accuracy-SA and fluency-SA – both Rate A and Rate B) were not significantly different from the normal distribution (see Appendix 25). However, log-transformation was needed for the remaining dependent variables to transform the distribution to one that was as close to normal as possible: (a) complexity, (b) accuracy, (c) fluency of FSs; and (d) total number of FSs used in the speech sample, and for the independent variables (i.e., a general proficiency test and a WM test). A log10 transformation method was used to reduce skewing. This method was chosen over other logarithms with reference to prior research in the field (e.g., Carrol & Conklin, 2020; Winter, 2019). However, according to the Shapiro–Wilk test the vocabulary test (NVLT) and the FS test were still not normally distributed in spite of the log transformations (sig. value of less than .05). Although these did not meet the assumptions for carrying out a linear regression for these two variables specifically, I continued with the analysis for two reasons. First, the two variables (NVLT and FS test) were also checked for the assumption of normality using Kolmogorov-Smirnov test. The result showed that the distribution was not significantly different from the normal distribution. Second, a slight deviation from normality may not have a significant influence on the results of an analysis especially with small sample sizes (Field, 2017), which was the case in the current study (N = 51). On the other hand, outliers can impact statistical analysis, including the mean and standard deviation, which can affect the interpretation of the overall conclusion (Field, 2017). However, this was not the case in this study.

Second, all measures were checked for outliers. Despite the presence of three participants with proficiency levels ranging from B1 to C1, the boxplot and scatterplots detected no outliers in any variables, implying there were no scores above +3.0 SDs or below -3.0SDs from the mean, which is a cut-off point widely used in Second Language Acquisition studies (Larson-Hall, 2010). Influential cases were also tested through Cook's Distance values for each participant. According to Cohen et al. (2017), values should not go beyond 1, otherwise, they are likely to be significant outliers and therefore would have exerted excessive influence on the results. There were no influential instances predetermining the model. For this reason, results from all 51 participants were retained for the analysis.

Third, all variables were checked for multicollinearity to guarantee that the final model comprised variables which shared hardly any variance, so that multicollinearity did not play a role in the multiple regression analysis (Tabachnick & Fidell, 2014). Multicollinearity refers to the occurrence of high intercorrelations among two or more independent variables (Cohen et al., 2017). Therefore, collinearity statistics Variance Inflation Factor (VIF) and tolerance values were computed to assess multicollinearity in the data (Field, 2009). In line with Field (2013), tolerance values should be greater than 0.1 and VIF scores less than 10. This was the case in the study (see Appendix 26). Therefore, there was no cause for concern regarding issues of multicollinearity. For every pair of variables with a correlation value higher than .80 (Cohen et al., 2017), only the variable closest to the aim of the study was retained. For this study, two of the predictors – the NVLT and the FS test – strongly correlated with each other ($r > .8$) which was to be expected as FSs are part of vocabulary knowledge. This means that the two predictors represent closely related constructs. It was decided to remove the NVLT as a predictor at the multivariate level, because the FSs test was more closely related to the focus of the study, namely the CAF of FSs.

Fourth, the assumption of the independence of the values of the residuals in the regression model was checked. It was found that the Durbin-Watson statistic was close to 2 for all models in this study (see the Durbin-Watson statistics in Appendix 27). This proximity to 2 means that there was little or no autocorrelation (serial correlation) in the errors of the regression model, as indicated by Cohen et al. (2017) and Field (2009).

Fifth, visual inspection of scatterplots of standardised residuals versus standardised predicted values revealed that the dots were similar at each point in the models (see Appendix 28), supporting the assumption of homoscedasticity and linearity (Winter, 2019; Field, 2009). P-P plots for the models were also inspected to ascertain whether the residuals were normally distributed. The more adjacent the dots to the diagonal line, the closer to normal the residuals are distributed, satisfying the assumption of normality of residuals.

After checking all the assumptions of the linear regression analysis were met, the data were analysed using R Studio, as will be presented in the next section. I turn next to the findings of this study.

4.3 Results

This section first presents the descriptive results, which include the results of the language background tests (Table 4.1), the WM tasks (Table 4.2), and the duration of participants' speech samples (Table 4.3). Beginning with the language tests, the proficiency test revealed a mean score of 20.35, corresponding to the A2 level on the CEFR scale, as indicated by the QPT. However, as detailed in Section 4.2, three participants achieved proficiency levels ranging from B1 to C1, explaining why the maximum score in Table 4.1 is 49. These participants were retained in the study as their scores did not qualify them as outliers (see Section 4.2). This distribution supports the decision to uniformly categorise the participants as mainly beginners, given the majority fell within the A1 to A2 range, with only a small number at higher proficiency levels. The results of the language tests highlight the variability in participants' performance and comprehension, as evidenced by scores on the NVLT, which ranged from a minimum of 7 to a maximum of 116, with a mean score of approximately 50.73. For example, in the NVLT, participants achieved scores ranging from a minimum of 7 to a maximum of 116, with a mean score of approximately 50.73. A standard deviation of 27.59 suggests considerable variability, with some participants scoring significantly higher or lower than the tests' average. Similarly, the FSs test reveals a diverse range of scores, with participants scoring as low as 1 and as high as 44 out of a total possible score of 50. The presence of a score as low as 1 implies minimal recognition of the tested FSs by at least one participant, contributing to the overall dispersion in scores. The average scores on the language tests suggest that, on average, the participants in this study are indeed lower-proficiency L2 learners of English, as the mean is always below half of the total score of each test. However, an exception was observed for the TF-FSs, as this test consists of FSs taken from the students' textbook (the same chapter from which the speaking task was adapted).

Table 4.1. Descriptive results from language background tests

Language tests	N	Minimum	Maximum	Mean	Std deviation
Proficiency test (out of 60)	51	6	49	20.35 (33.92%)	9.46
Vocabulary test NVLT (out of 120)	51	7	116	50.73 (42.28%)	27.59
FSs test (out of 50)	51	1	44	19.29 (38.58%)	11.58
TF-FS (out of 14)	51	5	14	11.73 (83.79%)	2.136

As mentioned in the previous chapter, the WM span tasks in this study consisted of three components, digit span forward (DSF), digit span backward (DSB), digit span sequencing (DSS). Table 4.2 below demonstrates the results of these digit span tasks, including the generation of a total score by summing up the scores from all three components. Examining DSF, the 51 participants in his study yielded an average score of 7.88 (SD 2.30), with scores ranging from a minimum of 4 to a maximum of 15. For DSB and DSS, the average scores were 8.14 (SD 3.16) and 8.12 (SD 2.55) respectively. The total generated score ranged from a minimum of 15 to a maximum of 44, with half of the sample scoring around the average.

The scores observed in this study align with those from the study by Abdelhamid et al. (2017) on WM tasks for Arabic speakers. The study reported a WM mean score of 27.55, which is comparable to the mean score in this study (M = 24.41). This may suggest that, on average, participants are performing similarly to other participants in digit span tasks, as expected as they are conducting tasks in their native language (see Chapter 3, The Current Study, Section 3.6.2,).

Table 4.2. Descriptive results from WM tasks

Working memory tests	N	Minimum	Maximum	Mean	Std deviation
Digit span forward (out of 16)	51	4	15	7.88 (49.25%)	2.30
Digit span backward (out of 16)	51	4	16	8.41 (52.56%)	3.16
Digit span sequencing (out of 16)	51	4	14	8.12 (50.75%)	2.55
WM total score (out of 48)	51	15	44	24.41 (50.85%)	6.64

Before moving into the results regarding the duration of participants' speech samples, it is important to clarify that the amount of speech produced by the participants during the task was sufficiently long for the analysis of fillers (*M* = 173.4 words, 95% CI [150.7, 196.2], *SD* = 80.8 words, range: 91–423 words). The duration of speech samples produced by participants in this study is comparable to earlier L2 speech studies that assessed lexical performance, considering around 100 words as the minimal length of spontaneous speech for adequate lexical analyses. For example, Saito (2020) considered a range of 95–424 words as samples of sufficient duration for robust vocabulary analyses, while Uchihara and Clenton (2018) found that the average number of words in production ranged from 81 to 319 words.

Table 4.3. presents the results for the duration of the participants' speech samples. As can be seen, the descriptive analysis suggests that the duration time tends to decrease as the speech becomes more focused. This suggests that participants, particularly those at lower proficiency levels, may not yet have enough vocabulary to produce a complete speech in the L2, leading to pauses and the use of L1. The score for excluding in-between L1 discussion had an average of 176.75, with a minimum score of 77 and a maximum score of 305. Ultimately excluding both in-between L1 discussion and pauses resulted in an average duration of 108.10 seconds, with a minimum score of 37 and a maximum score of 262. These results indicate the differences in the amount of time spent actually talking with the goal of completing the task, compared to the time spent negotiating meaning, asking for help, hesitations and confirmations while formulating an utterance. These results form the basis of the measurement of Rate B (number of meaningful syllables per minute of speech).

Table 4.3. Descriptive results from the duration of the speech samples

Speech sample rate (in seconds)	N	Minimum	Maximum	Mean	Std deviation
Full phonation time	51	93	407	222.16	71.58
Phonation time excluding in	51	77	305	176.75	56.01

between L1 discussion					
Phonation time excluding L1 discussion and pauses	51	37	262	108.10	47.39

4.3.1 Research Question 1

The first research question asked: *Do ESL beginner college learners use formulaic language (FSs) included within the curriculum and textbook when performing an oral proficiency task? If so, how do they use these FSs*? Findings showed that all participants used at least one FS (either from the textbook or outside the textbook) in their task performance. The majority of participants (92%) used FSs from the textbook, as shown in Table 4.4. The mean number of FSs taken from the textbook used by participants is 2.59 (SD = 1.458) (see Table 4.5). The descriptive results from Table 4.6 also indicate that all students were exposed to the fixed expressions found in the first chapter of the textbook, as shown by the test of familiarity of FSs (TF-FS) (*Min* = 5, *Max* = 14). Participants achieved a mean score of 86 per cent on that test. This might indicate that some students were only able to make use of some of the FSs from the textbook but did not use other FSs when speaking. Appendix 29 shows some examples of FSs elicited from the participants; these examples are chosen because they are representatives of the data collected. As discussed in the previous chapter, these multi-word units (from both the textbook and speech samples) have been identified as FSs on the basis of indices from TAALES (2.2) (Kyle et al., 2018) (and see Chapter 3, Section 3.10.2).

Table 4.4. Descriptive results of participants' source of FSs used (percentages)

	N (%)
Participants that used FSs from the textbook	47 (92%)
Participants that did not use FSs from the textbook	5 (8%)

Table 4.5. Descriptive results of number of FSs taken from the textbook

Number of FSs taken from the textbook				
N	Minimum	Maximum	Mean	Std deviation
51	0	6	2.59	1.458

Table 4.6. Descriptive results of participants' scores derived from the TF-FSs

Number of FSs taken from the textbook				
N	Minimum	Maximum	Mean	Std deviation
51	5	14	11.73	2.136

Findings from the speech samples also showed that most of the participants (N = 47) used FSs belonging to a less formal register or simpler than the ones provided in the textbook, while opting for FSs which fulfilled the same communicative function. For example, participants preferred to share their contact details before ending the conversation by uttering FSs including *this is my contact number*, *this is my phone number*, *my email address is,* instead of using more complex constructions such as *if you would like to keep in contact, this is my number*. Another example is *hope to X* instead of *looking forward to X* – again highlighting preferences for simplicity over complexity.

4.3.2 Research Question 2

Research Question 2 examined the following: *To what extent is the CAF of students' entire speech related to their use of FSs in a speaking task?* First, four different univariate regression analyses were conducted, each with Complexity, Accuracy and Fluency (Rate A and Rate B) of the entire speech sample (CAF-SA) as the dependent variables, and Number of FSs used by each participant as the sole predictor. Conducting the univariate analysis multiple times allowed for a detailed examination of the correlation coefficient (beta number) for each dependent variable. The univariate regression analysis showed that CAF-SA was statistically associated with the number of FSs elicited from the speaking task (Tables 4.7, 4.8, 4.9 and 4.10). Specifically, accuracy was the most predictive variable over complexity and fluency, as it has the highest beta coefficient (β = .817, p <.001). The results suggests that all CAF measures predict use of FSs to a great extent, with accuracy of speech emerging as the best predictor. In other words, the higher the number of FSs used, the higher the accuracy of speech is.

Table 4.7. Univariate linear regression analysis associating complexity of speech with number of FSs

Predictor	Complexity of speech					
	B	*p*	S.E	95% CI	*β*	*F*
Number of FSs	.709	<.001	.101	.507–.912	.709	F (1, 49) = 49.583

Table 4.8. Univariate linear regression analysis associating accuracy of speech with number of FSs

Predictor	Accuracy of speech					
	B	*p*	S.E	95% CI	*β*	*F*
Number of FSs	.817	<.001	.082	.651–0.982	.817	F (1, 49) = 98.121

Table 4.9. Univariate linear regression analysis associating Rate A (fluency of speech) with number of FSs

Predictor	Fluency (Rate A)					
	B	*p*	S.E	95% CI	*β*	*F*
Number of FSs	.726	<.001	.098	.528–.923	.726	F(1, 49) = 54.555

Predictor	Fluency (Rate B)					
	B	*p*	S.E	95% CI	*β*	*F*
Number of FSs	.720	<.001	.099	.521–.919	.720	F(1, 49) = 52.806

Table 4.10. Univariate linear regression analysis associating Rate B (fluency of speech) with number of FSs

In summary, the univariate regression analysis suggested that there is a statistically significant association between the number of FSs and each of the three CAF variables (complexity, accuracy and fluency) as measured on the entire speech samples.

4.3.3 Research Question 3

The third research question has two subsections. Research Question 3a asked: *Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their language history questionnaire (LHQ) on the one hand, and CAF-SA on the other, in speaking?* The correlation matrix, presented in Table 4.11 below, shows that the test of familiarity of FSs (TF-FSs) was most strongly correlated with CAF-SA, with correlation coefficients between .633 and .665 (p <.001). This evidence suggests a mid-strength correlation exists between these two variables (TF-FSs and CAF-SA). Significant correlations were also observed between the Proficiency test (.526 < r>.568, p <.001), NVLT (.497< r> .567, p <.001) and FS test (.364 < r> .434, p <.001). It seems reasonable to argue that Proficiency test was the second-best predictor for CAF-SA, as this implies that the higher the proficiency of the learner, the more language exposure they have had.

Comparatively, some of the background variables correlated less strongly with the CAF scores of the entire speech sample. Interestingly, however, the age at which participants started using language software to practise their L2 correlated negatively with two of the dependent variables, namely complexity of speech and Rate A (fluency of speech), with *r* -.280 for complexity of speech -and -.304 for fluency of speech (Rate A), with a level of significance <.05 for both complexity and fluency of speech. In other words, the younger the participants were when they started using L2 via language software, the higher their CAF scores were at the time of data collection of the speech sample.

Before moving on to the findings of the next research question, it is worth reporting the significant correlation between WM and complexity of speech (r = .369, p =.008), while WM did not correlate with accuracy or fluency. This significant correlation suggests those learners with a higher WMC tend to produce more complex language in their speech.

Table 4.11. Correlation matrix correlating learners' language tests, WM test and the results of LHQ on the one hand, and CAF-SA on the other, in speaking

variables	**Complexity of speech**	**Accuracy of speech**	**Fluency of speech (Rate B)**	**GP test**	**NVLT test**	**FS test**	**TF-FS**	**WM test**

Complexity of speech	1	0.792**	0.760**	0.568**	0.567**	0.434**	0.647**	0.369**
Accuracy of speech		1	0.832**	0.556**	0.507**	0.403**	0.665**	0.233
Fluency of speech (Rate B)			1	0.526**	0.497**	0.364**	0.633**	0.203
General proficiency test				1	0.733**	0.694**	0.641**	0.128
NVLT test					1	0.843**	0.532**	0.302*
FS test						1	0.484**	0.162
Test of familiarity of FS							1	0.028
Working memory test								1

Levels of significance ***<.001, **<.01, *<.05

For Research Question 3b – *Which of the predictors (general proficiency, vocabulary, working memory, variables from the LHQ) best predict speaking task performance as measured through CAF scores of entire speech?* – further univariate regression analyses and multivariate linear regression analyses were carried out. This was because there were significant correlations between the key variables of interest (C, A, and F of the entire speech sample) and the language tests – background variables from the LHQ and WM, as reported under the results for RQ3a. Variables associated with outcomes from the univariate analyses – one independent variable versus one dependent variable (see Appendix 19 for descriptive statistics of the univariate analyses of CAF-SA) were assumed to be suitable for inclusion in the multivariate linear regression analyses (see Appendix 20 for descriptive statistics of the multivariate analyses of CAF-SA). The participants' WM score was only included for complexity of entire speech (F (1, 49) = 7.71, p = .008) as it was not associated with scores for the remaining dependent variables; for example, the model for accuracy of entire speech (F (1, 49) = 2.804, p = .10) and (F (1, 49) = 2.114, p = .152) for Rate B failed to reach a statistically significant level with WM scores. These non-significant results indicate that individual difference variables other than WM may potentially have a more significant impact on mediating the accuracy and fluency of entire speech, at least for beginners. The focus of the results will be on the hierarchical linear regression analysis, as this type of analysis reveals the extent to which each predictor uniquely explains variance. Thus, the hierarchical linear regression analysis will start with complexity of speech (C) as the dependent variable, moving on to accuracy of speech (A) and finally to fluency of speech (F), both Rate A and Rate B. It is important to note, as discussed in the previous chapter (Section 3.12), that the alpha level for all multiple statistical tests of significance was set using Bonferroni's adjustment. Therefore, the alpha level of p = .01 will be applied for C-SA, and p = .012 for both A-SA and F-SA .

4.3.3.1 Explaining variability in Complexity of Speech (C-SA)

A hierarchical regression model was built with C-SA as the dependent variable. According to Field (2009), in a hierarchical regression model the researcher should decide the order in which predictors would be entered into the model based on research. Thus, predictors were added one by one in each model, starting from the least important (i.e., variables that may have the least contribution to the dependent variable) to those that are assumed to have the greatest contribution. This enables the researcher to determine whether the key variables of interest contribute significantly to explaining variance in the dependent variable over and above the contributions of other variables.

Although individuals with higher working memory capacity are reported to be better able to plan and organise their speech as their large WMC enables them to produce more complex utterances with fewer errors (e.g., Baddeley, 2003, 2015; Guará-Tavares, 2013; Mitchell et al., 2019), the main focus of this study is not on cognitive variables such as WM but on FSs and their effects. For this reason, FSs are treated as more important variables and the study focuses on the value of learning FSs and their impact on speech. It is important to note that the variable 'FSs used by participants in speech samples' is always placed last in the model, indicating that it is the most important predictor as this measurement is more important than formulaic sequences test scores (e.g., FS test and TF-FSs) because it captures the participants' actual performance in a speaking task, reflecting their ability to use formulaic language in real-time interaction. Therefore, the models were ordered as follows: Model (1) WM test; Model (2) WM test + Proficiency test; Model (3) WM test + Proficiency test + FS test; Model (4) WM test + Proficiency test + FS test + TF-FSs; and Model (5) WM test + Proficiency test + FS test + TF-FSs + FSs used by participants in speech samples. Regressions were computed with complexity of the entire speech sample (C-SA) as the dependent variable.

Model 1 (Table 4.12) shows a statistically significant association between C-SA and WM score (F (1, 49) = 7.710, p = .008) and WM explained 13 per cent of the variance in C-

SA. In particular, the model shows that the C-SA score would increase by .369 units for every 1 unit in the WM score.

Model 2 includes both the WM scores and Proficiency test scores; this emerged as a stronger model, explaining an additional 27 per cent of unique variance in C-SA (F (2, 48) = 16.828, p < .001). In total, the variance explained by the model is 38 per cent. While the former predictor did not exhibit any significant results, the latter displayed significance, with a beta value of .530 and p value of <.001 (Table 4.12).

Model 3 brings in the FS test score, which did not affect the magnitude or the direction of the Model 2 association. Notably, the WM variable remained statistically insignificant, and the Proficiency test also lost its statistical significance (p = .01) after Bonferroni adjustment (p = .01).

Model 4 introduces the TF-FSs scores, and interestingly proves to be a powerful model (F (4, 46) = 17.861, p < .001), accounting for an additional 18 per cent of the variance in C-SA. After adjusting the variables, the p value of WM decreased from .019 to .003, regaining its statistical significance (after the Bonferroni adjustment) and indicating a significant association between WM and the complexity of speech. Further, the beta coefficient increased, with a value of .016 in Model 3, so that for every 1-unit increase in the WM score, the C-SA increased by .293 units after controlling for the other variables. In particular, the new variable added in this model – TF-FSs – had the most powerful role in increasing scores for C-SA effect (β = .510; p = <.001). Neither the Proficiency test nor the FS test were associated with the scores for C-SA (Table 4.12).

In the final model (Model 5) (Table 4.12), the FSs used by participants in speech samples variable was added. Model 5 accounts for 66 per cent of the variance and is also significant (F (5, 45) = 21.215, p < .001). Results indicated that after adjusting for all predictors, a statistically significant association was found between C-SA and WM score (*β* coefficient = .235, *p* =.008). The number of FSs used by participants in speech samples was the most powerful potential predictor (*β* = .407, *p* = <.001). In contrast, the TF-FSs no longer exhibit statistical significance with the outcome variable after applying the Bonferroni adjustment (*p* = .017). An important finding of this study is that WM was found to affect the complexity of the speech dimension even after adjusting the association between the predictor and the dependent variable and controlling for the effects of all other variables. Similar to correlation analysis, linear regression does not imply a cause-and-effect relationship between variables, but it can provide information about the strength and direction of the relationship between variables. Thus, the higher the WM scores of the participant, the more complex the participant's speech was. The analysis also indicated that the number of FSs used by participants in speech samples contributed significantly to the model, while the TF-FSs could to some extent predict C-SA.

Table 4.12. Hierarchical regression analysis of predictors of C-SA

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* Change	ANOVA
Model 1					.118	.136	.008	*F* (1, 49) = 7.710, *p*=.008
WM test	.369	.133	.008	.102–.636				
Model 2					.388	.276	<.001	*F* (2, 48) = 16.828, *p*<.001
WM test	.301	.112	.010	.239–.719				
GP test	.530	.112	<.001	.306–.754				
Model 3					.388	.013	.314	*F* (3, 47) = 11.571, *p*<.001
WM test	.277	.114	.019	.048–.507				

GP test	.419	.156	.010	.104–.733				
FS test	.162	.159	.314	-.158–.483				
Model 4					.574	.184	<.001	F (4, 46) = 17.861, p<.001
WM test	.293	.095	.003	.102–.484				
GP test	.173	.141	.225	-.110–.456				
FS test	.118	.133	.381	-.150–.386				
TF-FS	.510	.110	<.001	.289–.731				
Model 5					.669	.094	<.001	F (5, 45) = 21.215, p<.001
WM test	.235	.085	.008	.063–.407				
GP test	.079	.127	.535	-.176–.334				
FS test	.178	.119	.140	-.061–.417				
TF-FS	.283	.114	.017	.053–.512				
FSs used by participants	.407	.108	<.001	.189–.625				

Significant at p = .01 after Bonferroni adjustment.

As mentioned in Chapter 3 (The Current Study, Section 3.12), a hierarchical regression analysis was also conducted in the opposite direction. This was needed to examine whether the predictors remain significant regardless of the order in which the variables are entered into the model. For complexity of speech, the models for the hierarchical analysis in the opposite direction were added as follows: Model (1) FSs used in speech samples; Model (2) FSs used in speech samples + TF-FSs; Model (3) FSs used in speech samples + TF-FSs + FS test; Model (4) FSs used in speech samples + TF-FSs + FS test + Proficiency test; and Model

(5) FSs used in speech samples + TF-FSs + FS test + Proficiency test + WM test, the results for which can be found in Appendix 23 (Table 1).

In summary, Model 1 shows a statistically significant association between C-SA and FSs used by participants in speech samples variable (β = .709; p = <.001), while subsequent models introduced additional variables and showed how they contributed to the model's explanatory power. The final model (Model 5) emphasises the importance of WM and reinforced the significance of FS usage (both FSs used in speech samples and TF-FSs) in enhancing C-SA (as shown in Appendix 23, Table 1).

4.3.3.2 Explaining variability in Accuracy of Speech (A-SA)

For A-SA, the models were added starting with the following variables: Model (1) Proficiency test; Model (2) Proficiency test + FS test; Model (3) Proficiency test + FS test + TF-FSs; and Model (4) Proficiency test + FS test + TF-FSs + FSs used in speech samples. Although the Proficiency test is an important pedagogic L2 tool that helps to assess the level of learners' overall linguistic knowledge (e.g., speaking), it was chosen to be entered first in the model (as the least important predictor) for two reasons. First, scores from the Proficiency test do not necessarily reflect speaking aptitude and subsequently the CAF of overall speech. Second, similar to the previous argument in Section 4.3.3.1, the focus of this study is on knowledge of FSs and their contribution in L2 learning/teaching. Thus, it was decided that FSs used in speech samples would be included as the final variable in the model. This decision stems from the fact that FSs constitute the key variable of interest in the current study, representing the knowledge of FSs elicited from a live performance of the task.

Model 1 was significant (F (1 , 49) = 21.892, p < .001) and the Proficiency test scores explained almost 31 per cent of the variance in A-SA. When the FS test was added in Model 2, scores from the Proficiency test were still shown to be the only statistically significant variable in the model (Table 4.13). However, no significant associations were found between the Proficiency test and the FS test and A-SA when the TF-FSs (β = .604, p = <.001) were added in Model 3, explaining additional 26 per cent of the variance. Finally, Model 4 reveals that when the total number of FSs elicited from speech samples was added, it was shown to be the best predictor of A-SA (p <.001). The TF-FSs was further attenuated with the beta coefficient (β = .286), still highly statistically significant (p = .006) and explaining 74 per cent of the variance (Table 4.13).

Table 4.13. Hierarchical regression analysis of predictors of A-SA

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig.
Model 1					.295	.309	<.001	*F* (1,49) = 21.892, *p*<.001
GP test	.556	.119	<.001	.317–.794				
Model 2					.283	.003	.664	*F* (2,48) = 10.861, *p*<.001
GP test	.504	.169	.004	.164–.843				
FS test	.074	.169	.664	-.266–.413				
Model 3					.542	.258	<.001	*F* (3,47) = 20.698, *p*<.001
GP test	.212	.146	.153	-.082–.505				
FS test	.027	.135	.844	-.245–.299				
TF-FS	.604	.114	<.001	.375–.833				
Model 4					.742	.193	<.001	*F* (4,46) = 36.912, *p*<.001
GP test	.083	.111	.460	-.141–.307				
FS test	.088	.102	.391	-.117–.294				

TF-FS	.286	.100	.006	.085–.488				
FSs used in speech samples	.575	.094	<.001	.386–.764				

Significant at p = .012 after Bonferroni adjustment.

In terms of accuracy, the models for the hierarchical analysis were added as follow: Model (1) FSs used in speech samples; Model (2) FSs used in speech samples + TF-FSs; Model (3) FSs used in speech samples + TF-FSs + FS test; and Model (4) FSs used in speech samples + TF-FSs + FS test + Proficiency test – the results for which can be found in Appendix 23 (Table 2). In summary, in Model 1, a statistically significant association was found between A-SA and total number of FSs used by participants in speech samples, explaining 66 per cent of the variance (F (1 , 49) = 98.121, p < .001). Mode1 2, with the addition of TF-FSs, shows a minimal 7 per cent effect but remained significantly associated with the outcome. Model 3 introduces FS test scores, which did not affect the association, and these scores were not associated with A-SA. In the final Model 4, even after adjusting for all predictors, both the number of FSs used in speech samples (β = .575, p = <.001) and TF-FSs scores (β = .286, p = .006) remained statistically associated with accuracy of speech. This indicates that participants who uttered more FSs in their speaking tasks achieved an increase of .575 units in accuracy of speech compared to those who used fewer FSs (Appendix 23, Table 2).

4.3.3.3 Explaining variability in Fluency of Speech (F-SA)

For Rate A and Rate B, the models were built up separately, starting with the following variables: Model (1) Proficiency test; Model (2) Proficiency test + FS test; Model (3) Proficiency test + FS test + TF-FS; and Model (4) Proficiency test + FS test + TF-FS + total number of FSs used by participants in speech samples. For Rate A and Rate B, Model 1 shows that Proficiency test scores were statistically significant (p <.001 for both Rate A and Rate B). Model 2 brings in the FS test, which failed to show any contribution for both Rate A and Rate

B. The Proficiency test, however, was still statistically significant for both Rate A (p = .011) and Rate B (p = .009). In Model 3, the TF-FSs was added, making this variable the only predictor which has an association with both Rate A (β = .488, p <.001) and Rate B (β= .509, p = <.001). Model 4, after the total number of FSs elicited from speaking task variable was introduced, showed that this specific variable was the only statistically significant variable to have an association with the dependent variables Rate A (p <.001) (Table 4.14) and Rate B (p <.001) (Table 4.15). This finding suggests that the use of FSs plays a mediating role in increasing fluency scores in speech, measured by both Rate A and Rate B.

Table 4.14. Hierarchical regression analysis of predictors of fluency (Rate A) of entire speech

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig.
Model 1					.280	.295	<.001	F (1, 49) = 20.488, p<.001
GP test	.543	.120	<.001	.302–.784				
Model 2					.275	.009	.436	F (2, 48) = 10.472, p<.001
GP test	.449	.170	.011	-.107–.790				
FS test	.133	.170	.436	-.208–.475				
Model 3					.438	.168	<.001	F (3, 47) = 14.004, p<.001
GP test	.213	.161	.194	-.112–.538				
FS test	.095	.150	.527	-.206–.397				
TF-FS	.488	.126	<.001	.234–.741				
Model 4					.591	.152	<.001	F (4, 46) = 19.085, p<.001
GP test	.099	.140	.484	-.183–.381				

FS test	.150	.128	.248	-.108–.409				
TF-FS	.206	.126	.108	-.047–.460				
FSs used in speech samples	.510	.118	<.001	.272–.748				

Significant at p = .012 after Bonferroni adjustment.

Table 4.15. Hierarchical regression analysis of predictors of fluency (Rate B) of entire speech

	β	*SE*	p	95% CI	R^2	R^2 change	Sig. F change	Model sig.
Model 1					.262	.276	<.001	F (1, 49) = 18.717, p<.001
GP test	.526	.122	<.001	.282–.770				
Model 2					.249	.003	.653	F (2, 48) = 9.309, p<.001
GP test	.471	.173	.009	.123–.818				
FS test	.078	.173	.653	-.269–.426				
Model 3					.428	.183	<.001	F (3, 47) = 13.482, p<.001
GP test	.224	.163	.175	-.103–.552				
FS test	.039	.151	.800	-.265–.343				
TF-FS	.509	.127	<.001	.253–.765				
Model 4					.572	.144	<.001	F (4, 46) = 17.728, p<.001
GP test	.113	.143	.433	-.175–.402				
FS test	.092	.131	.488	-.173–.356				

TF-FS	.235	.129	.075	-.024–.494				
FSs used in speech samples	.496	.121	<.001	.253–.740				

Significant at *p* = .012 after Bonferroni adjustment.

In terms of re-running the hierarchical analysis in the opposite direction for both Rate A and Rate B, the models were built up separately, starting with the following variables: Model (1) total number of FSs used by participants in speech samples; Model (2) total number of FSs used by participants in speech samples + TF-FS; Model (3) total number of FSs used by participants in speech samples + TF-FS + FS test; and Model (4) total number of FSs used by participants in speech samples + TF-FS + FS test + Proficiency test – the results for which can be found in Appendix 23 (Tables 3 and 4). In summary, Model 1 revealed a significant association between the number of FSs used by participants in speech samples and fluency of speech measured by both Rate A and Rate B ($p<.001$). In Model 2, which includes TF-FSs scores, additional explanatory power is introduced, accounting for approximately 57 per cent of the variance in Rate A and 56 per cent in Rate B. This highlights the influential roles of both the number of FSs used in speech samples and scores from the TF-FSs in increasing scores for both Rate A and Rate B. Model 3 introduced FS test scores, but the total number of FSs used in the speaking task remained statistically significant (β = .526, p = <.001). In the final Model 4, after adding the Proficiency test, only the total number of FSs used in the speech task remained significantly associated with both Rate A (β = .496 and p = <.001) and Rate B (β = .496 and p = <.001) (Appendix 23, Tables 3 and 4). Taken together, these models show the significant contribution of the number of FSs used in speech samples to the influence on fluency scores measured by both Rate A and Rate B, even when the variables are re-arranged in reverse order.

4.3.4 Research Question 4

Research Question 4a asks the following: *Is there a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and personal background variables on the one hand, and CAF(FSs) on the other?* Again, the findings for this research question support the previous one with regard to the language tests. Notably, TF-FSs was best correlated with CAF scores for FSs, with *r* between .597 and .632 and a level of significance *p* <.01 as shown in the correlation matrix below (Table 4.16). The following results are for the Proficiency test (.433 <*r*> .473, *p* = <.01), NVLT (.378 <*r*> .412, *p* = <.01) and FS test (.277 <*r*> .317, *p* = <.05). The strong correlations between the TF-FSs and CAF scores for FSs may suggest that familiarity allows learners to overcome challenges in speech production, enabling learners to produce language of higher complexity, accuracy and fluency.

As for variables taken from the LHQ, the age of using English via language software correlated negatively with all three of the dependent variables (C, A and F of FSs) with *r* around -.368 for complexity of FS, -.289 for accuracy of FS, and -.296 for fluency of FS and a level of significance <.008 for complexity and <.05 for both accuracy and fluency of FSs. This means that the younger the participants were when they started using English language software, the higher their scores on C, A and F of the FSs. Another interesting finding from the LHQ is the age of participants using English via online games, which also correlated negatively with two of the dependent variables (C and A) and which approached statistical significance with fluency of FSs – with *r* around -.340 for complexity of FS, -.278 for accuracy of FS and -.271 for fluency of FS, and a level of significance *p* = .015 for complexity, *p* = .048 for accuracy and *p* = .054 for fluency of FSs. Thus, the younger the participants were when starting to use English via online games, the higher their scores on CAF(FSs). These results indicate that learning through English language software at an earlier age may be related to participants' better performance in terms of CAF. Unexpectedly, WM scores did not statistically associate with any of the CAF of FSs variables, implying the speaker's WM was not related to a higher rate of CAF of FSs (Table 4.16).

Table 4.16. Correlation matrix correlating learners' language tests, WM test and personal background variables on the one hand, and CAF(FSs) on the other

variables	**Complexity of FS**	**Accuracy of FS**	**Fluency of FS**	**GP test**	**NVLT test**	**FS test**	**TF-FSs**	**WM test**
Complexity of FS	1	0.978**	0.964**	0.473**	0.412**	0.317*	0.632**	0.170
Accuracy of FS		1	0.982**	0.469**	0.398**	0.315*	0.609**	0.147
Fluency of FS			1	0.433**	0.378**	0.277*	0.597**	0.167
GP test				1	0.757**	0.705**	0.539**	0.128
NVLT test					1	0.885**	0.497**	0.352*

FS test						1	0.419**	0.263
TF-FS							1	0.060
WM test								1

Levels of significance *** <.001, ** <.01, *<.05

Turning to Research Question 4b – *Which of the predictors (general proficiency, vocabulary test scores, working memory test scores, variables from the LHQ) best predict speaking task performance as measured through CAF scores for FSs?* – again, the significant correlations (from RQ4a) between the dependent variables (C, A, and F of FSs) and the independent variables (language tests, variables from the LHQ and the WM) were appropriate for univariate linear regression analysis (See Appendix 21 for descriptive statistics of the univariate analysis of CAF-FS) and multivariate linear regression analysis (See Appendix 22 for descriptive statistics of the multivariate analysis of CAF-FS). Results will begin directly with the hierarchical regression analysis, starting with the complexity of FSs as a dependent variable, moving on to accuracy of FSs and finally to fluency of FSs. It is also important to note that the Bonferroni-adjusted alpha level was set to p = .012 for C-FS, and p = .016 for both A-FS and F-FS, as mentioned in Section 3.12.

4.3.4.1 Explaining variability in Complexity of FSs (C-FS)

The models for C-FS were added as follows: Model (1) Age of using English via online games; Model (2) Age of using English via online games + Proficiency test; Model (3) Age of using English via online games + Proficiency test + FS test; and Model (4) Age of using English via online games + Proficiency test + FS test + TF-FSs. Like RQ3b, predictors were added one by one in each model, starting from the least important (i.e., variables that may make the least contribution to the dependent variable) to those that could make the greatest contribution. The results can be found in Table 4.17. Stated simply, age of using English via online games was added in the first model and was assumed to be the least important variable; however, it was one of the variables found in the background questionnaire and revealed positive correlations with complexity. In contrast, TF-FSs was chosen to be inserted last in the model (i.e., being the most important predictor) as this test presents FSs that participants are exposed to (FSs taken from the same chapter that the speaking task was adapted from; see Chapter 3, The Current Study, Section 3.6.2 for more details).

Model 1 shows that after using the Bonferroni corrections, no statistically significant association was found between age of using English via online games and scores for C-FS in a speaking task (see Table 4.17 below).

Model 2 introduces the Proficiency test, which was found to be vital to the overall model (F (2, 48) = 12.433, p <.001) and the adjusted coefficient of determination (r^2= .314), showed that the predictors explained 31 per cent of the variability accounted for in the model. Results showed that after adding scores from the Proficiency test, the p value of age of using L2 via online games had dropped in the hierarchical regression analysis from p = .049 (in Mode 1) to p = .005 (in Model 2), regaining its statistical significance with the outcome variable (after the Bonferroni corrections). The Proficiency test was also statistically significant (p<.001) with a beta value of .519.

Model 3 shows that when FS test score was added, the overall model was still statistically significant (F (3, 47) = 8.324, p <.001); however, the FS test did not affect the magnitude or the direction of the Model 3 association, with the remaining variables still being statistically significant.

In the final model (Model 4), after the TF-FSs test had been added, the overall model was significant (F (4, 46) = 12.666, $p < .001$) and the adjusted coefficient of determination (r^2= .483), showed that the predictors explained 48 per cent of the variability in the model. It was shown that scores from the TF-FSs made the most important contribution to the complexity of FSs ($\beta = .516$, $p <.001$) (Table 4.17). Thus, it seems reasonable to argue that the more exposure to FSs learners had, the easier it was for them to identify the sequences and their correct meanings in a test – a point discussed earlier in this chapter. After adjusting for multiple comparisons using the Bonferroni correction, the significance of age of using English via online games in the final model of the hierarchical linear regression analysis was no longer evident ($\beta = -.250$; $p = .024$). Clearly, age of using English via online games does not influence the complexity of FSs. Moreover, neither scores from the Proficiency test nor scores from the FS test influenced C-FS.

Table 4.17. Hierarchical regression analysis of predictors of C-FS

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig.
Model 1					.058	.077	.049	*F* (1, 49) = 4.086, *p* = .049
Age of using English via online games	-.277	.000	.049	-.001–.000				
Model 2					.314	.264	<.001	*F* (2, 48) = 12.433, *p*<.001
Age of using English via online games	-.346	.000	.005	-.001–.000				
GP test	.519	.118	<.001	.281 –.756				
Model 3					.305	.006	.524	*F* (3, 47) = 8.324, *p*<.001
Age of using English via online games	-.357	.000	.005	-.001 –.000				
GP test	.596	.170	<.001	.255 –.937				
FS test	-.108	.168	.524	-.446–.230				

Model 4					.483	.177	<.001	*F* (4, 46) = 12.666, *p*<.001
Age of using English via online games	-.250	.000	.024	-.001 –.000				
GP test	.316	.161	.056	-.008 –.641				
FS test	-.125	.145	.392	-.418 –.167				
TF-FSs	.516	.125	<.001	.265–.767				

Significant at p = .012 after Bonferroni adjustment

When re-running the hierarchical analysis in the opposite direction for C-FS, the models were built up starting with the following variables: Model (1) TF-FSs; Model (2) TF-FSs + FS test; and Model (3) TF-FSs + FS test + Proficiency test – the results of which can be found in Appendix 24 (Table 1). In summary, TF-FSs is consistently shown to be a significant variable across all models, explaining a substantial proportion of variability in the complexity of FS. In Model 1, it accounts for 43 per cent of the variability, in Model 2, 42 per cent, in Model 3, 43 per cent, and in the final model, it explains 48 per cent of the variability. These findings emphasise the significant role that TF-FSs scores play in mediating the C-FS, even with a reversed order (Appendix 24, Table 1).

4.3.4.2 Explaining variability in Accuracy of FSs (A-FS)

In terms of A-FS, the models for the hierarchical linear regression analysis were added as follows: Model (1) Proficiency test; Model (2) Proficiency test + FS test; and Model (3) Proficiency test + FS test + TF-FSs. Model 1 was significant (F (1 , 49) = 13.797, p < .001) and accounted for 20 per cent of the variance, showing a statistically significant association between A-FS and scores for the Proficiency test (Table 4.18). Model 2 introduced the scores from the FS test (F (2, 48) = 6.777, p = .003). The new variable (FS test) did not show a contribution to the model and was not associated with A-FS (Table 4.18).

Model 3 exhibited a strong fit to the data (F (3, 47) = 12.436, p < .001). It is worth mentioning that the adjusted coefficient of determination showed that the predictors explained 41 per cent of the variability accounted for in the total number of accurate FSs. The final model confirmed that scores from the TF-FSs were the best predictor for A-FSs (p<.001) elicited from participants' speaking task. The TF-FSs had a beta value of .561, which means that for every 1 standard unit increase in the TF-FSs score, the A-FSs score increases .561 units (95% CI: .300-.821) (Table 4.18). This finding could be explained by the participants' limited proficiency, given that the TF-FSs consist of FSs extracted from the participants' textbook.

Table 4.18. Hierarchical regression analysis of predictors of A-FS

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig.
Model 1					.204	.220	<.001	F (1, 49) = 13.797, p<.001
GP test	.469	.126	<.001	.215–.722				
Model 2					.188	.000	.863	F (2, 48) = 6.777, p=.003
GP test	.491	.180	.009	.129–.852				
FS test	-.031	.180	.863	-.393 –.330				
Model 3					.407	.222	<.001	F (3, 47) = 12.436, p<.001
GP test	.219	.166	.192	-.114 –.553				
FS test	-.075	.154	.629	-.385 –.235				

TF-FSs	.561	.130	<.001	.300–.821				

Significant at *p* = .016 after Bonferroni adjustment.

For the hierarchical analysis in the opposite direction for A-FS, the models were built up starting with the following variables: Model (1) TF-FSs; Model (2) TF-FSs + FS test; and Model (3) TF-FSs + FS test + Proficiency test – the results of which can be found in Appendix 24 (Table 2). In summary, all three models – Model 1, Model 2 and Model 3 – are statistically significant in their association with A-FS. Across these models, the only consistently significant variable is TF-FSs, which explains a substantial amount of variability. In Model 1, it accounts for 40 per cent of the variability, followed by 39 per cent in Model 2, and 40 per cent in Model 3. This finding supports that the TF-FSs variable is the best predictor for A-FS, even when the variables are re-arranged in reverse order (Appendix 24, Table 2).

4.3.4.3 Explaining variability in Fluency of FSs (F-FS)

In terms of running the hierarchical linear regression analysis for F-FS, the models were built up with the following variables: Model (1) Proficiency test; Model (2) Proficiency test + FS test; and Model (3) Proficiency test + FS test + TF-FSs. An overview of the models can be found in Table 4.19.

Model 1 showed that scores from the Proficiency test showed a statistically significant association with the dependent variables (p = .002). In particular, the unadjusted model showed the Proficiency test would increase by .433 units for every 1-unit increase in F-FS (Table 4.19). Model 2 introduced the scores from the FS test (F (2, 48) = 5.597, p = .007). Neither the scores from the Proficiency test nor the FS test had any impact on the magnitude or the direction of the Model 1 association (Table 4.19).

In the final model, after TF-FSs was added (F (3, 47) = 11.262, p <.001), results showed that the TF-FSs was the only variable to have a statistically significant association with the dependent variable F-FS (p <.001). The β value for total number of F-FSs was .569, which means that for every 1 standard unit increase in the score of TF-FSs, the total number of fluent FSs score increases by .569 units (95% CI: .303–.836) after controlling for the other variables (Table 4.19), implying the more FSs one knows, the more fluent their speech seems to be.

Table 4.19. Hierarchical regression analysis of predictors of F-FS

	β	SE	p	95% CI	R^2	R^2 change	Sig. F change	Model sig.
Model 1					.171	.187	.002	F (1, 49) = 11.307, p=.002
GP test	.433	.129	.002	.174–.692				
Model 2					.155	.002	.758	F (2, 48) = 5.597, p=.007
GP test	.473	.183	.013	.104–.842				
FS test	-.057	.183	.758	-.425 –.312				
Model 3					.381	.229	<.001	F (3, 47) = 11.262, p<.001
GP test	.198	.169	.250	-.143 –.539				
FS test	-.101	.157	.523	-.417 –.215				
TF-FSs	.569	.132	<.001	.303–.836				

Significant at p = .016 after Bonferroni adjustment.

For the hierarchical analysis in the opposite direction for F-FS, the models were built up starting with the following variables: Model (1) TF-FSs; Model (2) TF-FSs + FS test; and Model (3) TF-FSs + FS test + Proficiency test – the results for which can be found in Appendix 24 (Table 3). In summary, the TF-FSs was the only variable displaying a statistically significant relationship with the dependent variable F-FS ($p < .001$) across all three models. This suggests that there is an association between one's familiarity with FSs and the fluency of FSs. In other words, the greater one's familiarity with FSs, the more fluent speech tends to be.

4.3.5 Research Question 5

Research Question 5 examined the following: *What are the students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency?* Thematic analysis of the qualitative data collected from the post-task interviews, where participants expressed their opinions on the ESP course, discussed their English learning methods, expressed and justified their needs, revealed five central themes. The primary themes derived from the participants' responses in the post-task interviews are presented in Table 4.20 below. The table further shows the frequency and percentages of each theme.

Table 4.20. Thematic analysis of learners' perceptions of their ESP experience

Themes	Frequency	%
Speaking practice		
• Need for increased speaking practice • Practicing enhances speaking	33	64.7%
• Speaking boosts confidence	23	45.1%
• Practising Enables/encourages noticing mistakes in L2	6	11.8%

	2	3.9%
Importance of FSs • Learners recognising FSs' importance in learning • Facilitate speaking • Handle daily interactions • Facilitate memorisation • Sentence construction is difficult	28 26 11 2 28	54.9% 51% 21.6% 3.9% 54.9%
Vocabulary learning • Classroom instruction emphasised individual words • Learning words from a list • Learned vocabulary is reviewed • No vocabulary reviewing during the course	41 33 4 36	80.4% 64.7% 7.8% 70.6%
FSs learning • We learn phrases	29	56.9%
Prioritising speech quality • Prioritising speech accuracy	12	23.5%

The data in Table 4.20 reveals that a large number of participants (64.7%) emphasised the need for increased speaking practice in their ESP course. For instance, Participant 11 stated, "*I'm consistently surprised by the neglect of speaking skills in class, considering their importance in our future careers, such as in the bank*". Nearly half of the responses (45.1%) highlighted that including speaking practice in the classroom was necessary because it significantly enhanced speaking abilities. Many participants reported struggling with the speaking task due to a lack of speaking practice, including Participant 12, who said, "*I struggled to perform this task due to a lack of speaking practice".*

Furthermore, other participants (11.8%) emphasised that incorporating speaking into the curriculum was important because it boosts their confidence. For instance, Participant 15 stated, "*If only speaking activities would be incorporated into our classroom; I believe this would boost my confidence in conversing in English."* A similar view was given by Participant 29, who noted, "*I would like more speaking time in the class; it would boost my confidence in English speaking".* This implies that students are aware of the necessity for increased speaking practice to increase their confidence in communication.

Interestingly, two participants (36 and 47) highlighted the importance of speaking practice in identifying and correcting language mistakes. Participant 36 clearly stated, "*If we have speaking activities, we will be able to notice and correct our mistakes".* Participant 47 elaborated on this idea, emphasising that speaking time is important for students to notice mistakes in their language use, allowing the teacher to provide corrections that improve language proficiency and speaking ability. This suggests that participants are aware of the importance of speaking practice within the classroom, emphasising its role in enhancing speaking abilities and boosting confidence and providing the opportunity to notice and correct language mistakes when using the L2.

The next question focused on FSs learning in the classroom. More than half of the sample (54.9%) showed their awareness of the importance of learning FSs in the L2 classroom. This awareness might be attributed to their exposure to FSs in the textbook. This was evidenced during the post-task interviews when participants were shown the TF-FSs to assess their knowledge (Chapter 3, The Current Study, Section 3.9). The majority (94%) confirmed having seen those FSs before, while the remaining participants did not provide a response on this matter (see Table 4.21).

Table 4.21. Descriptive results of participants' familiarity with FSs from TF-FSs (percentages)

	N (%)
Participants exposed to FSs from TF-FSs	48 (94.1%)
Participants who did not provide a response	3 (5.9%)

Learners had varied justifications for the importance of learning FSs, where a large number elaborated that this is because sentence construction is difficult (54.9%), it facilitates speaking (51%), helps one handle daily interactions (21.6%) and facilitates memorisation (3.9%). The majority of participants who commented on the importance of FSs said they eased sentence construction. Participant 5 explained, "*I was searching for sentences, but I struggle to create them*". Participant 29 explicitly stated, "*Learning phrases is important, especially for students like me who have limited knowledge of English. Phrases are helpful because they provide complete sentences, unlike single words which can be challenging to use in context*". It was interesting to observe that some participants (18, 21, 22, 24, 29, 33, 37, 39, 41 45, 49) considered learning FSs useful for handling daily interactions, like ordering from a restaurant, interacting in a business context, travelling, or making simple requests such as asking to use the restroom. One of the interesting responses was that from Participant 37, who explained that the challenge was not the speaking task's difficulty or the limited planning time. Instead, it was the minimal vocabulary she had to convey her thoughts. She stated:

> *Planning time was enough, but my vocabulary is restricted when it comes to expressing my ideas. I believe that knowing more phrases would facilitate handling the task and various real-life situations, such as using English in restaurants, business, and more.*

Highlighting another benefit of FSs, Participant 33 and Participant 40 emphasised the greater importance of learning phrases over single words. Participant 33 explained, "*Learning phrases is advantageous, forgetting one word within a phrase allows context from preceding words to help in recalling missing elements, unlike learning individual words*". Similarly, Participant 40 emphasised, *"Learning phrases facilitates memorisation; if one word is forgotten within a phrase, context from other words can help me recall, making phrases easier to remember."* Therefore, the use of FSs was perceived as beneficial for its ability to address learners' challenge in terms of: (1) constructing sentences from scratch, (2) facilitating speaking, (3) managing daily interactions and (4) ultimately enhancing memorisation.

Moving to the next question, Table 4.19 shows that 80.4% of the participants reported that classroom instruction primarily emphasised learning individual words. Many of these participants (64.7%) elaborated on their vocabulary learning methods, emphasising the use of word lists. For example, Participant 48 mentioned, "*We learn words from a list for the exam",* while other participants (N = 10) provided additional details, stating, *"Teachers provide word lists with their Arabic and English meanings."*

Around 7% of participants explained that the vocabulary they learn is revisited in subsequent courses or classroom materials, but the majority (about 70%) indicated that once they learn the new word, they seldom encounter it again except during exams when they are being assessed.

Furthermore, when participants were asked whether they also learned phrases, the data showed that 56.9% of participants acknowledged learning phrases within the classroom. Some (N = 7) further elaborated that they learned phrases through word rearrangement exercises, while others (N = 6) highlighted the practice of identifying phrases during group text reading. Notably, some participants (N = 14) expressed that although they focused on phrases, teachers primarily assessed them based on individual words. In other words, although FSs are taught, they are not subject to assessment, with more focus placed on individual words for vocabulary learning. FSs are being practiced receptively in the classroom, such as through word rearrangement exercises and highlighting full phrases with their meanings.

Finally, the data demonstrated that approximately 23% of the participants placed a high priority on accuracy of speech (Table 4.20). This observation emerged when they were asked about the reasons for taking longer to formulate their utterances during task performance. Some of the elicited quotes included: "*I felt frustrated because I did not want to make any mistake*s"; "*I was deeply focused on the task, aiming to speak without making any mistakes*" and "*I tried my best to avoid mistakes in my speech*".

Although the quantitative data answered the research questions, it did not provide insights into students' perceptions of learning and employing FSs to improve their spoken language when practiced in the classroom. The addition of the qualitative data, focusing on students' perceptions, supplemented the understanding of learners' needs in the L2 classroom. This study identified several benefits of FSs, such as easing the stress of constructing a sentence from scratch and reducing the pressure associated with striving for accuracy of speech. The analysis highlights the importance of FSs in language leaning, especially for learners confronted with producing speech in a live performance under time constraints. This challenge is a key factor preventing effective task performance in real-life situations and significantly shapes learners' perceptions of their language learning needs. Detailed discussions of the emerging themes will follow in the subsequent discussion chapter, aligning with the findings of this study and previous research on spoken language.

4.4 Summary of key findings

Thus far, this analysis has yielded key results that provided answers to the research questions raised in this study. In answer to the first research question, the descriptive results showed the FSs in speech samples were simpler than the examples presented in the textbook chapter from which the speaking task had been taken; this happened despite the fact that many of the participants knew the meanings of the complex FSs (as shown from the TF-FSs). The results from the analyses carried out to answer the second research question showed that the number of FSs in speech samples was highly significantly related to the complexity, accuracy and fluency of the samples, with accuracy emerging as the best predictor. In other words, the positive correlation suggests the higher the number of FSs used in a speech sample, the higher the CAF of SA. One of the key findings in RQ3a and RQ4a was that the TF-FSs was a strong predictor for both CAF scores for the entire speech samples and CAF scores for FSs elicited from the participants' speech; further, RQ3b was answered through a hierarchical linear regression analysis showing that learners with higher WMC can produce more complex utterances than learners with lower WMC. The same statistical procedures were followed when investigating RQ4b – one important finding from RQ4b is that the TF-FSs always appeared to be the most important predictor for increasing scores for the complexity, accuracy and fluency of FSs. Finally, RQ5 revealed learners' perceptions regarding the necessity for increased speaking practice in the L2 classroom and their recognition of the role of FSs.

These results will be examined more fully and discussed in detail in the following discussion chapter.

Chapter 5 Discussion

5.1 Introduction

This chapter interprets and discusses the results presented in the previous chapter in relation to the research questions and hypotheses of the study. It will evaluate the current results in light of the relevant literature and the findings reported in previous studies. Additionally, the qualitative data obtained through post-task short interviews with the students will be used to gain insights into students' perspectives with regard to the importance of FSs for language learning in general and, more specifically, for facilitating fluency.

The chapter is organised into three main sections. I begin with an overview of the key findings from this study, followed by detailed discussions on: 1) students' application and usage of FSs, 2) the relationship between CAF of students' overall speech and FSs, 3) the correlation between ESL learners' individual differences and CAF-SA in speaking, 4) the learners' individual differences that best predict scores of CAF-SA, 5) the correlation between ESL learners' individual differences and CAF-FS in speaking, 6) the learners' individual differences that best predict scores of CAF-FSs, and 7) students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency. Since the learning of FSs is also of interest, the next section will be dedicated to exploring the pedagogical implications closely related to the ongoing discussion within this chapter. The chapter concludes with a summary of the key points discussed throughout.

5.2 Overview of findings

The key findings of the study shed light on several important aspects of the use of FSs in teaching and learning. Firstly, the study found a predominant use of simpler FSs by most participants during speaking tasks, despite having the knowledge of the more complex FSs that serve the same communicative purpose (as shown by the test of familiarity of FSs and their responses in the post-task interviews). This prevalent use of simpler FSs may have been influenced not only by factors like time pressure and WM constraints but also by the distinction between explicit and implicit knowledge of such use, although the design of the current study did not allow for these constructs to be investigated. Additionally, the results showed that complexity, accuracy and fluency of entire speech had almost similar effects on the number of FSs used in a speaking task. This relationship suggests the important role of FSs in facilitating more complex, accurate and fluent language use. Accuracy emerged as the most predictive variable for the number of FSs used in a speaking task, suggesting that more accurate language usually included more FSs. Moreover, individual differences, especially WM, was associated only with complexity of speech. This finding suggests that learners with higher WMC can store and retain more complex speech compared to those with lower WMC.

The test of familiarity of FSs was the only individual difference measure that correlated with CAF of FSs. This finding indicates that familiarity allows learners to overcome challenges in speech production, enabling learners to produce language of higher complexity, accuracy and fluency. Additionally, a moderately positive correlation was observed between vocabulary test scores and CAF-FSs in a speech production. This was explained by two factors: 1) Vocabulary tests may not fully align with the specific FSs used in speech, especially within the context of self-introduction within a working life context; 2)

Low-proficiency learners may compensate for limited vocabulary by relying more on FSs, given their prior exposure to these sequences. As expected, general proficiency test scores showed a moderately positive correlation with CAF-FSs. This highlights the importance of general proficiency in language learning, although with a potentially less pronounced impact on the utilisation and integration of FSs into spoken language. Furthermore, participants, as revealed through the test of familiarity of FSs and post-task interviews, prioritised accuracy of speech in their ESP course. This tendency toward prioritising form (accuracy) over meaning (fluency) reflects a conscious effort to produce error-free speech.

Overall, these findings highlight the value of integrating FSs into pedagogy to enhance various aspects of speech. In the following sections, I discuss these findings in terms of how they relate to those from earlier studies, as well as the qualitative and quantitative data gathered in this study, and I attempt to offer some explanations.

5.3 Research Question 1: the application and usage of FSs

Research Question 1 explored the usage of formulaic language (FSs) incorporated in the textbook by college learners of English (mainly beginners). Additionally, it investigated how these FSs were employed in the learners' speech samples. The results showed that all participants (N = 51) utilised at least one FS, sourced either from the textbook or external resources. It is reasonable to argue that this finding is naturally expected, as a large proportion of human language relies on FSs. However, it is important to highlight that this finding suggests that even beginner learners of English who have limited knowledge of the language was found to have formulaic language such as *work hard* and *for example* in their speech samples. This reiterates Hall's (2010) findings among novice learners, which concluded that formulaic expressions were detectable in the speech production of all participants. Therefore, the findings from Hall's (2010) study and the current study add to results obtained by Tavakoli and Uchihara (2020), illustrating that even A1- to A2-level learners use FSs in their speech, aligning with the pattern observed among more advanced learners in their study. The frequency and complexity of these expressions, however, may differ based on the proficiency levels across the three studies.

The significance of FSs in language use is underscored in the literature (e.g., Goncharov, 2019; Liang, 2017; Pellicer-Sánchez & Boers, 2019) with various reasons for their importance (see Chapter 2, Literature review, Section 2.2 for a detailed explanation). Indeed, Martinez & Schmitt (2012) indicated that FSs often carry specific meanings or serve particular communication purposes. In the current study, participants used FSs to help them convey meanings in an ESP context, specifically in scenarios such as self-introduction in a business context, including FSs like *nice to meet X* and *thank you for X*. This is of course explained not only by their widespread usage of FSs – approximately 60% of spoken language (Thornbury, 2019; Erman and Warren, 2000; see Chapter 2, Literature review Section 2.2) – but also by the fact that FSs span across various genres (Biber et al., 2004; Hyland, 2008; Wulff et al., 2009), including their utilisation in business-related genres. This study identified specific business-related FSs, such as *communicate with, keep in touch, deal with* and *buy and sell.*

Furthermore, it was hypothesised that in the speaking task, learners would predominantly use simpler FSs instead of more complex ones (both available in their textbook) that serve the same communicative purpose. The results confirmed the predicted use of less complex FSs that belonged to a less formal register compared to the more complex ones presented in the textbook – for example, using a FS such as *want to X* instead of *would like to X* when expressing preferences or requests. One way to explain this is that when speakers are under time pressure, their primary focus is expressing their message/meaning, and using complex language is not a priority for them. This finding implies that complex speech can overload their working memory (Ahmadian & Tavakoli, 2011), and therefore they prioritise meaning over form (Long, 1991; Skehan, 1998). The observed preference for simplicity over complexity in participants' use of FSs aligns with previous literature findings regarding the CAF of overall speech (Yuan & Ellis, 2003; Ellis & Yuan, 2005).

Despite the limited use of the complex forms, many of the participants knew the meaning of the complex FSs that served the same communicative purpose in their speech samples, as was shown from the test of familiarity of FSs (FSs taken from the students' textbook – the same chapter from which the speaking task was adapted), where the participants achieved a mean score of 83.79 per cent (see Chapter 4, Results, Section 4.3). However, their preference for simplicity over complexity of FSs could also be explained by fear of failure. This idea was also supported in the post-task interviews, where six students mentioned that one of the major reasons for advocating more attention on speaking within the L2 classroom is to boost their confidence. This was particularly relevant for participants who identified themselves as shy and afraid of making mistakes in the post-task interview. Indeed, cultural issues affecting language learners are discussed in the literature, outlining reasons for students' reluctance in speaking, often due to fear of making mistakes or inherent shyness (Alharbi, 2015; Gudu, 2015; Stupar-Rutenfrans, Ketelaars, and van Gisbergen, 2017; Pakula, 2019). Therefore, the findings of the current study suggest that language educators need to equip learners with FSs relevant to the task. This provision can reduce unfamiliarity, subsequently enhancing learners' confidence in their ability to communicate. As discussed in Section 2.4.5, the limited exposure to and interaction with a range of FSs by many L2 learners, including the participants in this study, makes it important to provide these learners with a repertoire of FSs to fulfil their linguistic needs (Van Vu & Peters, 2022).

A further possibility is that learners in this study may have merely noticed the meanings of FSs in the test without the ability to apply their knowledge of FSs in practice, emphasising the distinction between explicit and implicit knowledge. This highlights the difference between FSs recognition and FSs production/use. Noticing, considered central to the initial phase of L2 learning (Damanhouri, 2018; Doughty, 2001; Wen et al., 2021; Schmidt; 1990, 1994, 2001) does not ensure successful application of FSs in participants' speech, as observed in this research.

Repeated exposure, through practice and rehearsal, can mediate explicit knowledge (learning in declarative memory) into implicit knowledge (procedural memory) and ultimately into automatic performance (Ullman & Lovelett, 2018). The activation of procedural knowledge inevitably requires practice and repetition, which in turn facilitates the process of acquisition (Segalowitz, 2010). If use of FSs reduces demands on WM and repetition facilitate the learning of these FSs, then the combination of FSs and repetitive practice is likely to benefit quality of speech and presumably its CAF. This process can hypothetically foster learning through consistent practice, ultimately leading to automatization. As a result, this progression significantly enhances the CAF of FSs, thereby improving overall speaking ability.

5.4 Research Question 2: the relationship between CAF of students' overall speech and FSs

Research Question 2 investigated the relationship between CAF of students' entire speech samples and their use of FSs in a speaking task. The hypothesis was formed in line with previous studies indicating that fluency improves when the content is familiar and tasks require less intricate planning (Bui and Huang, 2018; Foster & Skehan, 2013). Aligning with this hypothesis, the results demonstrated that the number of FSs elicited from the speaking task has a substantial impact on the fluency of the entire speech. This finding suggests that an increase in the usage of FSs in a task is associated with a corresponding improvement in fluency of speech, specifically when measured by Rate A. Notably, Rate A demonstrated a slightly higher beta value of .726 compared to Rate B's beta value of .720. This nuanced difference implies that among lower proficiency learners, a higher utilization of FSs is associated with a higher rate of syllables per minute in their speech. These findings explain the intricate relationship between FS usage and fluency, providing insights for understanding language dynamics in the context of learners with lower proficiency.

However, the findings also showed that complexity and accuracy were of almost equal influence when associated with the number of FSs, with accuracy standing out as the most predictive variable for the number of FSs used in speech samples, exceeding complexity and fluency (β = .817, p <.001, as detailed in the previous chapter). Two significant implications arise from these findings. Firstly, complexity (β = .709), accuracy (β = .817), and fluency, including Rate A (β = .726), and Rate B (β = .720) of entire speech, show almost similar effects on the number of FSs used in a speaking task, and this effect is strongly statistically significant (p < .001), deepening our knowledge on the pivotal role of FSs. This highlights the importance of the use of FSs, which can be used to promote opportunities for producing more complex, accurate and fluent language use (e.g., Boers et al., 2006; Wood, 2016, Wray, 2002). Secondly, the prominence of accuracy, from the remaining overall qualities of speech, as the best predictor for the number of FSs used by participants is interesting. This is similar to Levelt's monolingual speech production model, which has been adapted for L2 speakers (e.g., De Bot, 2020; Kormos, 2006; Segalowitz, 2010). According to this model, both L1 and L2 speakers are constantly monitoring their speech to ensure successful communication. Participants explicitly noted in the post-task interview that their primary focus was on producing error-free speech. Therefore, this emphasis on accuracy implies a continuous monitoring process, which can indirectly explain the relationship between accuracy and monitoring.

It is also possible that accuracy being the best predictor might stem from the method by which FSs are learned – through the memorisation of complete, accurate chunks stored, retrieved and produced as a whole. Given that FSs are reported in the literature to be stored and retrieved as one whole unit (Tavakoli & Uchihara, 2020; Wood, 2020), they are not subject to analysis, and therefore they are largely produced with a high level of accuracy. The rapid retrieval and usage of FSs as prefabricated units support their error-free nature, free from lexical, morphosyntactic and syntactic errors. Consequently, as the number of FSs in a speech sample increases, the overall accuracy of the speech also increases. The findings of this research question align with those of Boers et al. (2006), where the experimental group significantly outperformed the control group with regard to fluency. However, unlike that by Boers et al., this study revealed an additional association with accuracy. Nevertheless, it is important to remember that this study follows a correlational design and correlation should not be equated with causation (Field, 2013, p. 270).

It would be interesting to conduct a similar analysis with advanced L2 speakers to explore the correlation between the number of FSs used in their speech samples and accuracy of speech. This comparative study could reveal whether the number of FSs used in participants' speech samples is related to the degree of accuracy of their utterances – whether their frequency of FSs is linked to accuracy. In other words, determining whether their speech is accurate because it incorporates numerous FSs could shed light on the role of FSs in the accuracy of advanced speakers.

5.5 Research Question 3a: correlation between ESL learners' individual differences on the one hand, and CAF-SA in speaking, on the other

Research Question 3a asked whether there is a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their language history questionnaire, and their performance in CAF-SA. One of the interesting predictions was that a higher WMC should positively associate with higher scores of complexity, accuracy and fluency of entire speech. This hypothesis was corroborated by Baddeley (2001, 2003, 2015), Conway et al. (2007), Skehan (2015) and Mitchell et al. (2019), suggesting that WM plays a crucial role in performing cognitively demanding tasks, including L2 speaking. However, contrary to these expectations, the results from the current study showed that WM only correlated with complexity of speech, while no significant correlations were observed with accuracy or fluency. Thus, the higher the participants' WM scores, the more complex the participant's speech. This result confirmed the findings of Gilabert (2010), Kormos and Trebits (2011) and Awaad and Tavakoli (2022), who reported a correlation between WM and complexity. The findings of these studies taken together suggest that the relationship between WM and syntactic complexity exists across levels of proficiency, as the current study recruited lower proficiency learners compared to intermediate and advanced learners within those other studies. However, those studies also found that WM correlated not only with complexity but also with other qualities of speech like accuracy or fluency. One possible explanation for the absence of a correlation with accuracy in the present study could be attributed to the lower number of FSs elicited from the participants. Given that many of them were novice learners, the reduced number of FSs might contribute to the absence of a clear relationship with accuracy in this specific context (i.e., the case for lower proficiency learners in this study).

Furthermore, predictions included strong correlations between scores from the TF-FSs and the CAF of entire speech, and a high correlation specifically between the FS test scores and fluency of speech. Additionally, it was hypothesised that vocabulary knowledge would have a high correlation with the fluency of entire speech. The results confirmed the high correlation predicted between scores from TF-FSs and CAF of entire speech. Surprisingly, the FS test scores seemed to be the least powerful predictor of CAF-SA. This discrepancy may be attributed to the FS test's broad range of expressions, some potentially unfamiliar to the students, and encompassing FS characteristics that participants were not yet familiar with, such as transparency and length of multi-word units (see Chapter 2, Literature review, Section 2.2.1, for a review of the identification of FSs). Thus, given that the TF-FSs was familiar to students as it consisted of FSs taken from their textbook, and the literature emphasises the importance of familiarity and exposure in language learning (Ellis, 2003; Dai & Ding, 2010; Van Vu & Peters, 2022; Wulff, 2019), the observed correlation between scores of TF-FSs and CAF-SA could be interpreted to suggest that the variable of familiarity may be considered as a proxy for exposure. However, it is important to note that correlation does not imply causation. Additionally, it is important to note that exposure, in the current context, remains undetermined, an aspect that was not operationalized or controlled for. Therefore, exposure as an independent variable requires further exploration in subsequent studies. In this sense, the TF-FSs comprises expressions derived from the first chapter, which students were expected to have encountered, and this chapter is also the source of the speaking task. Participants recognised these FSs, leading to improved scores in this particular test, a finding supported by their answers in the interview (94%, see Chapter 4, Results, Section 4.3.5). This finding is in line with the literature, which stresses the crucial role of familiarity with the domain knowledge, which, in turn, may enhance the complexity, accuracy (Thai and Boers, 2016) and fluency (Bui & Huang, 2018) of speech production. In other words, the TF-FSs, being more specific and aligned with the students' encountered content, demonstrated a higher correlation. It may also be the case that the Martinez (2011) FS test was too advanced to accurately capture beginner learners' knowledge or to establish an association, considering that the participants were mostly lower-proficiency learners. However, since the number of FSs used in speech samples in RQ2 associated positively with C, A and F of speech, then it may be the case that more advanced learners would show a positive correlation between the results of the FS test and the CAF of overall speech.

It was hypothesised that the general proficiency test would show a moderate correlation with the C, A and F of entire speech. Contrary to what was anticipated, the outcomes obtained in the current study demonstrated a different pattern with the relationship between the scores from the general proficiency test and CAF-SA. However, this assertion that the proficiency test scores were considered the second-best predictor for CAF-SA could stem from the implication that learners with greater proficiency have likely accumulated more language experiences, resulting in better performance with the CAF-SA in speech samples. This is similarly noted in Kyle (2016), which suggested that from a usage-based perceptive, higher-proficiency learners should have more experiences with the language.

Surprisingly, the NVLT showed a less significant correlation with the CAF of speech, suggesting the performance on the NVLT test did not account for variations in CAF-SA. This weaker correlation might indicate that the NVLT did not encompass lexical items relevant to the speaking task, which was focused on giving a self-introduction in a business context. Additionally, the participants, mainly beginners in an early stage of vocabulary development, might not have integrated their vocabulary learning sufficiently into their spoken production. Although it may appear to be contradictory to arguments by Hinkel (2020), Van Vu and Peters (2022) and Webb and Nation (2017) on the importance of vocabulary for coherent and authentic speech, it is important to consider the beginner-level proficiency of these learners. This low proficiency level may have failed to indicate a correlation between the scores of NVLT and CAF of speech, thus obscuring the importance of learning vocabulary at this stage. Furthermore, this finding may imply that at lower proficiency levels, it could be beneficial for learners to focus on learning FSs before individual vocabulary items.

5.6 Research Question 3b: learners' individual differences predictors that best predict scores of CAF-SA

Research Question 3b looked into variables *(general proficiency, vocabulary, working memory, variables from the LHQ)* that best predicted scores of CAF-SA. It was anticipated that the best predictor of CAF-SA elicited by participants would be the FS test scores, followed by WM test scores, as presumed by the literature highlighting the importance of FSs in language use (Goncharov, 2019; Pellicer-Sánchez & Boers, 2019; Tavakoli & Uchihara, 2020; Vu & Peters, 2022; Wood, 2020), in addition to the role of WM in language learning (Ahmadian, 2013; Baddeley, 2001, 2003, 2015; Conway et al., 2007; Mitchell et al., 2019; Skehan, 2015). The following subsections discuss the findings of the current study with regards to the dimensions of speech production, addressing them individually in the sequence of complexity, accuracy and finally fluency.

5.6.1 Learners' individual differences that best predict scores of complexity of overall speech

For complexity of speech, the results were largely in line with the initial expectations, with the number of FSs used by participants in speech samples being the most powerful potential predictor of the complexity of entire speech. Surprisingly, the FS test by Martinez (2011) did not show any significant contribution to the scores of complexity of entire speech. The prominence of the number of FSs used as the leading predictor expands on the results of Wood's (2009, 2010, 2016) studies, which consistently showed that an increased exposure to FSs assisted with improvement in fluency, suggesting that FSs not only mediate fluency but also seem to facilitate speech complexity. In contrast, the scores from FS test by Martinez (2011) failed to show any contribution, supporting my previous argument regarding the importance of familiarity and exposure (as explained in RQ3a). This result also aligns with the explanation provided in RQ3a, suggesting that the test might have been too challenging for most of the participants' proficiency levels.

Interestingly, and despite participants performing relatively well on the TF-FSs, this test did not predict C-SA. This could imply that the ability to recognise the meanings of FSs in a test might not directly mediate the complexity of their L2 speech. This finding is in line with Nergis's (2021) study, which reported that FS instructions did not increase oral complexity scores; instead, academic vocabulary instructions were more influential. This might imply that, given the task's focus on an ESP context, technical business vocabulary could have potentially enhanced the complexity scores of the entire speech. Another possible explanation could be linked to the idea that increased language complexity often signals learners' willingness to experiment with new linguistic structures (Bui & Skehan, 2018). In this study, participants might have refrained from using complex language in their speech performance, possibly explained by the earlier findings highlighted in RQ1 regarding their tendency to avoid using complex FSs.

The pattern observed in the hierarchical linear regression analysis, coupled with those in the correlational analysis (RQ3a), seem to suggest a significant association between WM and the complexity of speech (p = .003). This finding is in line with studies by Gilabert and Muoz (2010), Guará-Tavares (2013), Kormos and Trebits (2011) and, more recently, Awaad and Tavakoli (2022) – who found that WM correlated with complexity. Additionally, the results from RQ1 may be linked to findings from RQ3a and RQ3b, emphasising that learners with higher WMC can store and retain more complex linguistic items and structures compared to those with lower WMC.

5.6.2 Learners' individual differences that best predict scores of accuracy of overall speech

As for accuracy of speech, the findings partially aligned with the initial hypothesis. Again, although I expected to see the FS test scores (Martinez, 2011) to show a relationship with accuracy, it did not show any contribution to scores of accuracy of speech. However, the number of FSs utilised by participants in their speech samples emerged as the most influential potential predictor ($p < .001$) of accuracy of speech, with the scores from the TF-FSs (p = .006) serving as the second-best predictor of accuracy of speech. This suggests that these two predictors, specifically the number of FSs in speech samples and TF-FSs scores, were particularly effective because of their incorporation of FS usage. The processing advantages associated with FSs (e.g., easy storage and retrieval of error-free chunks) may account for their prominent role in predicting accuracy (Martinez & Schmitt, 2012). However, it was interesting to observe that the TF-FSs is a significant predictor of accuracy of speech but not the complexity of speech (as shown previously). This could support the hypothesis underlying RQ2 in that these FSs were stored and retained as a whole and were not subject to grammatical and lexical analysis that could lead to errors. These participants may have studied the phrases in the same chapter that the speaking task was taken from, and for this reason those who scored better in accuracy of entire speech may have done so because they retained these FSs from the test of familiarity – especially because these FSs shared the same context as the participants' speaking monologue task.

Research evidence, as demonstrated by Bui and Huang (2018), indicates that knowing FSs is advantageous, particularly in specific domains, where meanings and communicative functions of FSs are often associated with specific formulas. Therefore, the findings of the current study suggest that participants' high performance in a test is attributed to their familiarity with domain-specific FSs. Providing learners with sufficient FSs related to the task can enhance their performance, allowing them to perform the task with increased complexity, accuracy and fluency. This aligns with the study's advocacy for the vital role of learning FSs, emphasising their importance in reducing planning time (e.g., Christiansen & Chater, 2016; Khodadady & Shamsaee, 2012; Wray, 2002) and functioning as short-cuts for both the acquisition of L2 facets and processing grammar and vocabulary (See Chapter 1, Introduction, Section 1 and Chapter 2, Literature review, Section 2.1 for an overview).

It is possible to argue that the scores from the FS test by Martinez (2011) did not contribute to the model because the FSs in this test are more general and are unfamiliar to many of the participants. Given that the participants fall within the low proficiency range (between A1 and A2 according to the CEFR), the FS test by Martinez (2011) may have been beyond their current proficiency level, as has been argued previously. This is especially true with the general proficiency test, which also did not affect the association of interest after adjusting the model, and thus was not associated with participant's accuracy of speech elicited from the speaking task (p = .460). The lower proficiency levels of participants likely contributed to their reduced familiarity with the language in general. This would explain the results of the general proficiency test and FS test and may explain why the correlation with the familiarity of FSs test scores was stronger in the correlation matrix (as shown in Chapter 4, Results) and why it is a better predictor than results from both the general proficiency test and the FS test in the hierarchical linear regression model.

The results of the hierarchical linear regression analysis did not show any significant association between the WM and A-SA. This finding contradicts the initial hypothesis that WM would have an association with accuracy of speech because greater WM plays a crucial role in retrieving information (e.g., accurate utterances), suggesting that learners with higher WMC are more adept at retrieving information during the execution of demanding cognitive tasks, leading to enhanced fluency in their performance (Guará-Tavares, 2013). However, the current study's finding, indicating that WM had no relationship with accuracy of overall speech, is consistent with the results of some other studies in the literature. These studies revealed a significant correlation between WMC, fluency and complexity, but not accuracy (e.g., Gilabert & Munoz, 2010; Guará-Tavares, 2013). In contrast, the results of the current study contradict the results recorded by Awaad and Tavakoli (2022) as well as de Paula et al. (2021). Awaad and Tavakoli (2022) found that WM was associated with accuracy across two task types (i.e., complex and less complex) on L2 speech performance. However, it is worth mentioning that the researchers employed tasks that were based on a metric of reiterating a narrative extracted from two videos involving varying degrees of intentional reasoning, whereas in this study the oral task was based on a monologic task adapted from the students' textbook, which students should have encountered in class. Thus, the task undertaken in this study may be considered as having relatively low cognitive load, compared to the cognitively demanding and complex task used in the study by Awwad and Tavakoli, which required the students to watch the video and retell it simultaneously. From a methodological perspective, Awaad and Tavakoli's (2022) study also differs from the current study in that it did not provide any pre-task planning time. Thus, the different results can be interpreted in the light of the different task demands and performance condition (specifically, the provision of planning time).

On the other hand, in the study by de Paula et al. (2021), the variance in results with regard to accuracy could potentially be attributed to the utilisation of a different memory span test. Specifically, de Paula et al. (2021) employed a speaking span test to measure participants' WM, while the current study used a digit span test. The latter is considered language-independent, and known to mitigate potential influences of language proficiency on WM scores (Harrington & Sawyer 1992; Wright, 2010). This discrepancy in testing methods might account for the differences observed in the findings.

Further, a conclusion can be reached based on the mixed findings reported above. The differences in results can be interpreted in light of the varying levels of proficiency among the participants in each study. In Gilabert and Muoz's (2010) study, participants were high-intermediate/advanced, while in the studies by Guará-Tavares (2013) and de Paula et al. (2021), they were intermediate level. Awaad and Tavakoli's (2022) study included participants with a range of proficiency, whereas the current study mainly involved beginners. These differences in language proficiency levels may have contributed to the varied outcomes observed across the studies, and it is notable that Gilabert and Muoz's (2010) study indicated that differences in WM were significantly correlated with the performance of only the participants in the high-proficiency group. This finding supports Mizera's (2006) argument that WM becomes especially beneficial for L2 learners during later stages of acquisition. While everyone possesses WM, its role and importance may vary at different stages of development. In the initial phases of development, learners might allocate more cognitive resources to store immediate knowledge than advanced learners would. During this early phase, learners may rely heavily on attentional and memory resources to access and retrieve lexical items, and structure messages with appropriate syntax and morphology, making differences in WM less significant (Gilabert & Munoz, 2010). However, as learners progress and become more proficient, the role of WM becomes increasingly crucial in facilitating their language learning process.

5.6.3 Learners' individual differences that best predict scores of fluency of overall speech

As for fluency of speech, the total number of FSs elicited from the speaking task was the only variable that showed a significant association with both Rate A (p <.001) and Rate B (p <.001). This finding suggests the more FSs one knows the more fluent speech seems to be. This finding supports previous studies, such as those by Boers et al. (2006), Goncharov (2019); Mirzaei et al. (2016); McGuire & Larson-Hall (2018); Nergis (2021); Tavakoli and Uchihara (2020) and Wood (2009), who found that use of FSs enhances fluency of speaking production. This adds support for teaching FSs rather than single vocabulary units – a point that Martinez and Schmitt (2012) argued for. The argument for not simply focusing on single vocabulary units was explored in more detail during the participants' post-task interviews, which will be discussed in RQ5. Approximately 80 per cent of participants mentioned that their classroom instruction mainly focused on individual words. For example, Participant 37 argued that "*learning words is crucial, but gaining proficiency in phrases is even more important*", demonstrating a high awareness of the relevance of FSs in language usage.

Participant 49 also remarked, "*We do learn phrases, but more focus is allocated to single vocabulary items*".

Unexpectedly, scores from the test of familiarity of FS did not statistically associate with either Rate A or Rate B, supporting the predictions of Skehan's trade-off hypothesis (i.e., in the Limited Attentional Capacity). This hypothesis suggests that when the speaking task is cognitively demanding, L2 learners have limited attentional resources available to them and therefore will need to prioritise meaning (fluency) or form (complexity and accuracy), with an improvement in some but not all aspects of CAF. In other words, when one facet of C, A and/or F in speech increases there may be a trade-off with the other (see Section 2.4.3). This echoes the findings of Yuan and Ellis's (2003) study, in which participants showed significant effects in complexity and accuracy but not fluency. However, the trade-off observed in fluency within the current study is not in line with Nergis's (2021) findings, which reported effects in fluency and complexity but not accuracy. The trade-off effects may also vary from one individual to another, as each speaker possesses a specific priority in determining which quality of speech to focus their attention on. For example, in post-task interviews, 12 participants highlighted their focus on avoiding making errors in their speaking performance, thereby allocating a focus to accuracy.

The hierarchical linear regression analysis did not reveal a statistically significant association between scores on the general proficiency test, the FS test and fluency. Again, this may imply issues of limited proficiency and familiarity of the language, as has been discussed earlier.

It is also noteworthy that the results in this study contradicted the hypothesis that WM would be one of the best predictors of fluency. Findings suggest that there is no association between WM and fluency of speech in the current study; these results are in line with those of

Kormos and Trebits (2011) and Awaad and Tavakoli (2022), who did not observe effects of WM on various aspects of L2 speech performance, including fluency. The question of why WM was not associated with fluency of speech could be explained by the association between vocabulary knowledge and fluency, suggesting that L2 learners with wider lexical knowledge (e.g., FSs) and faster access to it could perform vocabulary searches more effortlessly and quickly (Koizumi & In'nami, 2013). Therefore, vocabulary knowledge might explain more variance in lexical access and retrieval compared to WMC. Still, it appears that there are previous studies that found WM correlated with fluency (e.g., de Paula et al., 2021; Gilabert & Munoz, 2010; Guará-Tavares, 2013; Mizera, 2006). However, in line with the previous argument regarding accuracy of speech, WM becomes especially beneficial for L2 learners as they progress to more advanced stages of language learning (Mizera, 2006), as all of the studies that identified a relationship between WM and fluency had a common factor: they involved participants with intermediate to advanced levels of language proficiency.

5.7 Research Question 4a: correlation between ESL learners' individual differences on the one hand, and CAF-FS in speaking, on the other

Research Question 4a asked whether there is a correlation between ESL learners' general language proficiency, vocabulary knowledge, working memory and the results of their language history questionnaire, and their performance in CAF-FSs. This question was built on the hypothesis that because vocabulary knowledge is one of the many variables that influence speaking production (de Jong et al., 2012) and given that FSs are a specific subcategory of vocabulary, a strong correlation between vocabulary knowledge and FSs was predicted. Conversely, a relatively weaker correlation was anticipated between general proficiency and CAF-FSs. Additionally, it was hypothesised that as learners' WMC increases, there would be a corresponding increase in the production of FSs in participants' speech. Finally, the hypothesis proposed that the more hours learners spend using everyday spoken

English (as obtained from the LHQ), the better they will be at utilising complex FSs (as was stated in Section 3.3).

The findings revealed that the scores for TF-FSs were best correlated with CAF scores for FSs, suggesting familiarity with the FSs could predict the participants' ability to incorporate and utilise these FSs in their spoken language, thus promoting their complexity, accuracy and fluency scores. This finding aligns with those of previous studies (Boers et al., 2006; Bui & Huang, 2018; Wood's, 2009; 2010; 2016) in which participants benefitted from prior knowledge in a particular domain in order to speed up idea retrieval. Familiarity with FSs seems to indeed allow learners to overcome challenges in speech production, enabling learners to produce language of higher complexity, accuracy and fluency. The study conducted by Boers et al., (2006) identified that familiarity (i.e., FSs) led to improved performance in terms of fluency and range of expressions but not accuracy. Additionally, both Bui and Huang's (2018) study and Wood's series of studies (2009, 2010, 2016) demonstrated gains in fluency. In contrast, the current study has shown a positive correlation between familiarity (i.e., TF-FSs) and all three facets of speech production: complexity, accuracy and fluency of FSs in speech. However, I should note that Boers et al. (2006) did not investigate complexity, and Bui and Huang's (2018) study, along with Wood's series of studies (2009, 2010, 2016), did not specifically explore complexity or accuracy. Moreover, while Boers et al. (2006), Bui and Huang (2018) and Wood (2009, 2010, 2016) explored the familiarity of FSs in relation to the accuracy or fluency of overall speech, this research question places particular emphasis on investigating the familiarity of FSs concerning the CAF of FSs within speech samples. Notably, direct comparisons with other studies exploring the relationship of CAF within the context of FSs are limited in the existing literature. Future replication studies are needed to determine whether similar results are obtained when measuring the complexity and accuracy of FSs; similar significant results may be observed for complexity and accuracy.

Further, the strong correlation between CAF scores for FSs and the TF-FSs scores can be accounted for by the shared domain between the speaking task and the assessed FSs. This is due to the fact that the CAF scores are derived from a task within the same domain as the familiarity test. However, if the FSs found in the TF-FSs were obtained from a different domain, the correlation could have differed. Of course, the same rationale can also be extended for the CAF of overall speech.

While a strong relationship was anticipated between the vocabulary test scores and CAF-FSs, the result showed that the correlation was only a moderately positive one (r between .378 and .412, $p = <.01$) suggesting vocabulary knowledge can only predict a small variation in the complexity, accuracy and fluency of FSs in a speech production. Two explanations can be given based on this finding. First, it is worth considering that vocabulary tests, in general, might not comprehensively cover or closely align with the specific FSs produced in participants' speech, especially in the current study, where the FSs employed by participants primarily revolved around the context of self-introduction within a working life context. Another way to interpret this result is to highlight that since most participants in the current study are low-proficiency learners, they might have compensated for their limited vocabulary by relying more on FSs to enhance the complexity, accuracy and fluency of their speech, particularly considering that these participants were likely to have encountered and been exposed to these FSs previously. It is in this context that recent research is advocating for the importance of FSs in language learning (e.g., Goncharov, 2019; Liang, 2017; McGuire & Larson-Hall, 2018; Tavakoli & Uchihara, 2020; Van Vu & Peters, 2022). Martinez and Schmitt (2012) highlighted a notable advantage of FSs – meanings and communicative functions (contexts; the purpose of utterance) are often associated with specific formulas, which therefore facilitates effective interaction (See Chapter 2, Literature review, Section 2.2 for a full discussion of formulaic sequences and their importance).

The general proficiency test scores showed a moderately positive correlation with CAF-FSs (*r* between .433 and .473, *p* = <.01) which is in line with the hypothesis. This finding suggests that participants' general proficiency can only predict a small variation in their production of complex, accurate and fluent FSs. This could be attributed to the fact that FSs primarily comprise of phrases that are often retrieved from the mental lexicon as a whole, as prefabricated units (Boers et al., 2006; McGuire & Larson-Hall, 2018; Tavakoli & Uchihara, 2020; Wood, 2010, 2020; Wray, 2002; see also Chapter 2, Literature review, Section 2.2). Consequently, speakers might not engage in the process of formulating new messages in terms of both vocabulary and syntax. This finding implies that although general proficiency is an important factor in language learning, its influence might be less potent when it comes to the utilisation and integration of FSs into spoken language. This is because FSs are typically learned as chunks, allowing for their rapid and effective employment (Segalowitz & Hulstijn, 2005) without necessarily relying on general language knowledge to assemble these multi-word combinations.

It was also interesting to see that variables taken form the LHQ, the age of using English via language software (*r* around -.368 for complexity of FS, -.289 for accuracy of FS and -.296 for fluency of FS), negatively correlated with all three of the dependent variables (C, A and F of FSs). This finding indicates that the earlier the participants started using English via language software, the more complex, accurate and fluent the FSs in their speech samples. Additionally, the age of participants using English via online games also showed a negative correlation with two of the dependent variables (C and A), approaching statistical significance in its correlation with fluency of FSs (*r* around -.340 for complexity of FS, -.278 for accuracy of FS and -.271 for fluency of FS). These findings are interesting because both the age of learning English through language software and the age of using English via online games are important in strategies of learning facilitated by software. The correlation between these variables and the dependent measures (C, A and F of FSs) suggests that the utilization of these learning programs could potentially explain the improved performance observed in participants. Although it is beyond the scope of this paper to discuss the role of learning L2 via software further, it is important to keep in mind the important role learning through software may provide with regard to the age of learning onset and diversifying the sources for interacting with the target language.

Surprisingly, the scores for WM did not correlate with any of the CAF of FSs variables. Direct comparisons to previous studies are not possible because to the best of my knowledge this is the first study that has investigated the relationship between CAF of FSs and the cognitive variable of WM. However, the results show that more investigations is required to inspect the relationship between CAF of FSs and WM.

5.8 Research Question 4b: learners' individual differences that best predict scores of CAF-FSs

Research Question 4b investigated variables *(general proficiency, vocabulary, working memory, variables from the LHQ)* that best predicted scores of CAF-FSs. It was hypothesised that both vocabulary test scores and the hours of English used in daily activities would serve as the most effective predictors for elevating the scores of CAF of FSs. Note that unlike the previous discussion, where the individual dependent variables (C, A and F) were addressed separately in response to the research questions, the following discussion of this research question will combine the C, A and F of FSs, as these dependent variables shared one common predictor – the test of familiarity of FSs.

5.8.1 Learners' individual differences that best predict scores of complexity, accuracy and fluency of FSs

In terms of the complexity, accuracy and fluency of FSs, findings revealed that scores from the TF-FSs made the only significant contribution in enhancing the predictability of C, A and F of FSs scores derived from speech samples (β = .516, p = <.001 for complexity of FSs; β =

.561*, p* = <.001 for accuracy of FSs; and β = .569*, p* = <.001 for fluency of FSs). The question of why, out of all the language tests, only the TF-FSs emerged as a significant predictor for the C, A and F of FSs could be answered by referring to the descriptive results presented earlier in the previous chapter (Chapter 4, Results, Section 4.3). There was an exception shown specifically with regard to the TF-FSs, where participants seemed to perform notably well (mean score 83.79%). This outcome suggests that it is likely that familiarity and exposure are responsible for this exceptional result: the more exposure to FSs, the higher the possibility that learners can identify them and their correct meanings in a test. This perspective echoes Wray's (2002) and N. Ellis's (2003) assertions that frequent encounter and utilisation of learned items (e.g., FSs) enable leaners to readily recall them. Along the same lines, Schmidt (1990, 2001) emphasised the significance of "noticing", suggesting that learners' awareness of language components contributes to their language development. However, it is important to consider that noticing alone does not ensure successful application of FSs in participants' speech, particularly with complexity, as observed in RQ1 and RQ3b.

As FSs are useful in different contexts, such as giving commands, making requests and incorporating politeness markers and hedges (Wray, 2002), these phrases can be used in daily routines and can be transferred to suit any given situation (such as business interactions). However, it needs to be kept in mind that most of the participants in this study are lower-proficiency learners, which may be the reason why the test of familiarity of FSs has been shown to be the best predictor. For higher-proficiency learners, the general FS test (i.e., Phrasal Vocabulary Size test) by Martinez (2011) might be a better predictor.

The necessity of learning pre-selected FSs tailored to individual needs to facilitate speaking interactions was a topic that captured the interest of participants in the current study.

This was confirmed during the post-task interviews, where more than half of the participants (55%) exhibited an understanding of the importance of learning FSs. They are aware that FSs are constructed as a string of words that are retrieved as a whole to express their ideas, rather than having to deal with adding individual words together to formulate an entirely new utterance. In other words, participants seem to acknowledge that FSs effectively minimise the time needed for speech planning, echoing points raised by several researchers, including Christiansen & Chater (2016), Khodadady & Shamsaee (2012) and Wray (2002). FSs function as short-cuts for acquiring both the target language aspects and processing the complexities of grammar and vocabulary (Wray, 2002; Wray & Perkins, 2000), fostering confidence in learners to speak, as producing novel utterances is extremely challenging (Hall, 2010; Wray, 2002). This is especially relevant as 55 per cent of the participants emphasised in their post-task interviews that constructing a new sentence is difficult, elaborating on the reason for their poor performance in their speaking task. The remaining participants, accounting for 45 per cent of the total, did not mention this topic. Furthermore, other participants raised some noteworthy perspectives regarding the importance of learning FSs. For example, Participants 41 and 42 stressed that using phrases would facilitate speaking as they are constructed as a whole, enabling them to be memorised as a complete unit instead of having to assemble individual words to generate a novel sentence/utterance. Participant 48 raised a very interesting point, reflecting her awareness of a crucial advantage of FSs, stating, "*Studying phrases is valuable because I can memorise the entire phrase more effectively. It is better than learning single words only because if I forget a word within the phrase, the other words will serve as a reminder.*" Eleven participants further elaborated on the role of FSs in enabling them to handle situations and daily interactions, such as requesting, travelling, conveying appreciation to someone, or even meeting up with someone. This shows that even lower-proficiency learners are aware of the value of learning FSs for the development of their speaking productions in terms of the CAF of overall speech. Furthermore, they are aware of the benefits of learning FSs over single vocabulary items in terms of reducing the overall effort of learning new items. This finding is consistent with Ding's (2007) view, as highlighted in his study exploring the success of three accomplished Chinese L2 learners (of English) in winning nationwide English-speaking competitions. Ding (2007) advocated planning and rehearsal through text memorisation (i.e., FSs) and imitation, as these practices enable learners to integrate what they have recited and noticed so that it evolves to be their individual language expression. However, as explained previously, memorising every single utterance is impractical. Therefore, to optimise the efficacy of memorisation, a balance should be achieved, allowing for the retrieval of extensive linguistic units when required and modified to suit the requirements of the context (Wray & Fitzpatrick, 2008; see also Chapter 2, Section 2.4.5 for full discussion and examples).

5.9 Research Question 5: students' perceptions of the importance of FSs for L2 learning and enhancing fluency

Research question 5 looked into participants' perceptions of the importance of FSs for L2 learning, with a specific focus on their role in enhancing fluency. It aimed to ascertain whether participants were introduced to FSs and to what extent these new lexical items/FSs were integrated into their speaking performances, if such use occurred. Additionally, the purpose of this research question was to complement and supplement the findings, where possible, with those from previous research questions in this study. Each theme deriving from this RQ is observed to relate to earlier RQs. RQ1 is linked to the need for enhanced speaking practice, RQ3b encompasses themes related to vocabulary learning, learning of FSs and prioritisation of speech accuracy, while RQ4b is associated with the recognition of the importance of learning FSs.

The thematic analysis of the post-task interviews revealed five main themes: 1) the need for enhanced speaking practice, 2) recognition of the importance of learning FSs, 3) the need for/the practice of vocabulary learning, 4) issues related to learning FSs, and 5) the prioritisation of accuracy when speaking.

The findings showed that around 70 per cent of the participants expressed a need for increased speaking practice within the classroom. This aligns with the hypothesis and echoes the challenges faced by English learners in different contexts, including those in Kuwait, specifically within the CBS, due to limited emphasis on speaking in the L2 classroom.

A subset of participants (23.5%) cited the reason for taking longer to formulate their utterances during task performance was due to placing a high priority on accuracy of speech. The emphasis on accuracy adds to the findings from previous research questions, specifically RQ2 and RQ3b. The relationship between FSs and accuracy suggests that FSs function as a significant source of accuracy. In other words, the number of FSs predicts accuracy, indicating that a substantial portion of accuracy within an utterance is attributed to the presences of FSs, especially at the beginner proficiency level. Moreover, it highlights the significance of accuracy as a predictor, especially in comparison to fluency and complexity. In addition, it corroborates with the observations from the post-task interviews, where two participants stressed that increasing classroom speaking helps in identifying mistakes in L2 production. Despite its relatively lower contribution (3.9%) to the overall perception of the need for speaking in class, increased classroom speaking to help identify mistakes in L2 is crucial for beginner learners. The observed perception also provides additional support to the argument that learners in this study prioritise error-free speech. This emphasis is further elaborated in their post-task interview responses, which attribute this prioritisation to cultural issues such as shyness and fear of making mistakes. As previously discussed in RQ1, the impact of cultural issues on language learners has been addressed in the literature review

(Alharbi, 2015; Gudu, 2015; Stupar-Rutenfrans, Ketelaars, and van Gisbergen, 2017; Pakula, 2019).

However, speaking fluently and automatically requires a solid foundation of linguistic knowledge, such as FSs, and likewise for automatic processing of such knowledge. The qualitative thematic analysis identified an important recognition among learners regarding the importance of FSs, or as learners' commonly referred to them, "expressions". One key rationale identified was the difficulty in sentence construction, accounting for 54.9% of the responses. This aligns with research by Lightbown and Spada (2021), who argued that fluent speakers utilise strings of words that frequently co-occur instead of creating new sentences word by word. It seems reasonable to argue that the general proficiency tests did not significantly associate with complexity, accuracy and fluency of speech, thus not predicting complex, accurate or fluent utterances. Clearly, participants who face challenges with vocabulary and syntax might encounter difficulties in constructing sentences because of their lower proficiency level. Furthermore, classroom instruction primarily focused on individual words (as reported by 80.4% of learners) in this study. This finding further explains the challenge for participants in transitioning from individual lexical units to generating sentences, highlighting the value of relying on prefabricated FSs. Time pressure impacts speech processes, often leading to pauses and corrections in speech, as noted by Bygate (2001, 2018), while insufficient L2 vocabulary and syntax for beginner learners make using FSs a more practical choice over creating sentences from scratch, as suggested by Wray (2002).

Moreover, participants highlighted that FSs facilitate speaking (51% of the responses), assist in handling daily interactions (21.6% of the responses) and facilitate memorisation (2% of the responses). These findings underscore the multi-faceted role of FSs in language learning, supporting the emphasis of FSs in L2 teaching advocated by Pellicer-

Sánchez and Boers (2019). This perspective aligns with Martinez and Schmitt (2012), recognising FSs as an integral part of vocabulary essential in both language learning and practical usage.

In summary, the findings from this research question revealed participants' substantial need for increased speaking practice, particularly emphasising accuracy. These findings contribute to triangulating the interpretations of the empirical results obtained in this study. The identified themes, interconnected with previous research questions, strengthen our understanding of the challenges faced by English learners, especially beginners, within the CBS context. The significant relationship between FSs and accuracy highlights the pivotal role of prefabricated expressions in L2 learning. In addition, the recognition of FSs as vital for overcoming challenges in sentence construction and their benefits in addressing time pressure and limited vocabulary contribute to a nuanced understanding of learners' linguistic needs. The multifaceted roles of FSs, from facilitating speaking to managing daily interactions and memorisation, underscore their importance in L2 learning. This aligns with current pedagogical perspectives advocating for integrating FSs into language learning.

While providing participants FSs and encouraging their use in speech has demonstrated gain in fluency (Wood; 2009, 2010, 2016), it is noteworthy that beginners can also enhance accuracy and complexity, as shown in RQ2, RQ3a and RQ3b. The memorisation and rehearsal of FSs as error-free whole chunks can elevate accuracy and complexity, further boosting participants' confidence in speaking, as supported by research such as that by Wray and Fitzpatrick (2008). However, this progress relies heavily on practice. Without ample practice, declarative knowledge may not transfer/transform to procedural knowledge. If FSs are not encountered and used regularly, their retention might be compromised, impacting not only fluency but also other qualities of speech.

5.10 Pedagogical implications

Having discussed the results and observations derived from this study, I will now discuss the pedagogical implications. In this discussion, I mostly draw upon the literature review encompassing pedagogical strategies aimed at enhancing L2 learners' performance in speaking tasks, particularly the use of FSs. This includes a reflection on pedagogical methods like task repetition, memorisation and performance, as discussed in Sections 2.4.4 and 2.4.5. Additionally, this section discusses automatisation and Levelt's stages of speech production, elucidating pedagogical implications for the learning of FSs to facilitate speech production – an important aspect of this study.

Research suggests that learners require multiple exposures to and practice of lexical items for effective learning to occur (N. Ellis, 2003; Van Vu & Peters, 2022; Wulff, 2019). Explicit knowledge alone cannot causally contribute to the learning of implicit knowledge without practice (e.g., DeKeyser, 2018; N. Ellis, 2005; Hulstijn, 2007; Jeong & De Keyser, 2023; Paradis, 2009). In other words, repeated encounters and rehearsals are important for new knowledge to be productively used and automatised in speech production. The study's findings, particularly from the TF-FSs, revealed that participants were familiar with the meanings of complex FSs presented in their textbook. Post-task interviews confirmed their prior exposure to these FSs, as indicated by 91% responses (see Chapter 4, Results, Section 4.3.5). However, despite this explicit knowledge, the results from RQ1 indicated that participants predominantly used simpler FSs during the speaking task. This discrepancy suggests that their explicit knowledge did not seamlessly transfer to actual use.

Therefore, the pedagogical implication for the classroom is that complex FSs might be more difficult to learn, and that they require more focused and dedicated teaching strategies when compared to simpler forms. These could potentially be operationalised through employing strategies to facilitate the learning and use of FSs, thereby fostering speech with more complexity, accuracy and fluency. In my study, I explored a pedagogical framework designed to facilitate the learning of FSs via a speech production, known as the ACCESS approach (Gatbonton and Segalowitz, 2005). While the ACCESS approach is not the only pedagogical framework applicable in the L2 classroom, it has garnered attention in the literature (Dao et al., 2017). As discussed thoroughly in the Literature Review (Section 2.4.2) the framework advocates for the automatization of FSs through the employment of communicative activities. ACCESS involves learning strategies such as repetition, memorisation and rehearsal, aiming to mediate explicit knowledge into procedural knowledge (practice), thereby facilitating the development of automatisation (Jeong and DeKeyser, 2023; O'Malley & Chamot, 1990).

These strategies play a crucial role as they allow learners to have opportunities to recycle FSs effectively in speaking performance within context, as advocated by ACCESS-recommended activities like interviews and retelling a situation to different peers. While ACCESS encourages activities such as roleplays to enhance communicative abilities among learners, I would recommend to exclude this task for lower-proficiency learners based on findings from the pilot study, indicating its unsuitability for beginner learners (see The Current Study, Section 3.7). Nevertheless, with higher proficiency learners, role-paly activities could be really beneficial. Therefore, the effectiveness of these activities are likely to be more or less useful depending on the context and the proficiency level of the learners, highlighting the need for flexibility in pedagogical approaches.

Moreover, considering the rapid process of speech production and that effective application necessitates automatization (Bygate, 2015; De Bot, 2020, Levelt, 1989), speakers often face a trade-off in deciding which aspect of performance to prioritise during L2 speaking – retaining meaning (fluency) or form (accuracy and complexity). Indeed, the findings from RQ3b in this study highlighted observed trade-off effects, particularly between accuracy and fluency associated with scores from the test of familiarity of FSs with CAF of speech. As shown previously, this trade-off was further confirmed through post-task interviews, where 12 participants stressed their focus on avoiding errors in their speaking performance, emphasising accuracy over fluency. This preference may stem from various reasons, including cultural factors such as the desire to save face from making mistakes. As explored in the literature review (Chapter 2, Sections 2.4.3 and 2.4.4), SLA researchers have investigated this trade-off concern and explored various methodologies to enhancing all qualities of speech. These methodologies include practice and repeating tasks that focus on different facets of facilitating learners' oral production (Bygate, 2018; de Paula et al., 2021; Gatbonton & Segalowitz, 2005; Ghazi-Saidi & Ansaldo, 2017; see also Chapter 2, Literature review, Sections 2.4.4 and 2.4.5).

Despite the satisfactory performance of learners in the TF-FSs in the current study (mean score 83.79%) and their presumed familiarity with the speaking task (given that this task was adapted from the first chapter of their textbook), the students might not have had sufficient opportunities to engage in practice and repetition of these FSs in speaking performance to recycle them effectively. This teaching/learning strategy, advocated by many researchers (e.g., Galian-Lopez, 2018; Garbati and Mady, 2015; Jeong & De Keyser, 2023; Nation, 2007; Suzuki, 2024; Thai & Boers, 2016), involves going over learned vocabulary items for improved retention. This point was further confirmed through the post-speaking task interview, in which around only 8 per cent of participants indicated that they would go over the learned vocabulary item during the course, while around 71 per cent of participants reported that their teachers did not employ this methodology. Instead, they further indicated that vocabulary items are introduced once, with the focus solely on learning the meanings and providing the L1 translation of the lexical item. The remaining 22 per cent did not provide any mention of this topic. The notable percentage of 'No' responses suggests that repetition and practice is indeed not being used in the L2 classroom, which therefore answers the question of why the test of familiarity of FSs has not been a significant predictor of fluency of speech (both Rate A and Rate B).

Although participants may have been exposed to this task before, it is important to note that this does not necessarily indicate that they have actually learned these FSs. Consequently, the automatisation of this linguistic knowledge (i.e., FSs used to facilitate the completion of the speaking task) should not be assumed in a live task performance. This is because the three components of speech production – Conceptualiser, Formulator and Articulator – demand the integration of repetition and practice, as automatisation can then be achieved. As already discussed in section 2.4.4, repetition enables L2 learners to rely on previously conceptualised task content and recycle recently acquired linguistic knowledge to produce their utterances fluently (Lambert, Kormos & Minn, 2017).

Speaking proficiency requires extensive practice, as highlighted in studies by researchers such as Gatbonton & Segalowitz (2005), Jeong & De Keyser (2023) and Mitchell et al (2019), who advocated for automatization – a vital aspect of fluent speech developed through repeated practice. Such practice could be designed through teaching methods based around noticing errors. Research evidence suggested the usefulness of noticing (Damanhouri, 2018; McGuire & Larson-Hall, 2018) in retrieving knowledge and avoiding errors (Schmidt, 2012). For example, the study by McGuire and Larson-Hall (2018) focused on noticing the use of FSs and identified improvements in participants' speech fluency. One of the sub-themes elicited when the participants expressed a need for increased speaking practice within the classroom (see RQ5) was helping in identifying mistakes in L2 production. This aligns with the L2 learning literature advocating for error noticing, a point reinforced by researchers such Juffs and Harrington (2011), Lightbown & Spada (2021), McGuire & Larson-Hall (2018), Mitchell et al. (2019), Schmidt (2001) and Wen and Skehan (2021). Teachers could dedicate more time to activities that enable learners to notice errors in their speech, either individually or within peer interactions. This can also be demonstrated through activities like recording, where the learner speaks about a given topic (e.g., arranging a meeting) to the recorder, listens to it to check whether corrections could be made, and then re-records the speech (Nation, 2007). Error detection is important for producing messages effectively and signifies the potential benefits of feedback for improving their linguistic abilities. The learners' focus on identifying errors indicates a desire to bridge the gap between intended production and actual proficiency, indicating an eagerness to enhance linguistic precision or accuracy. This recognition could guide educators in tailoring teaching methods that encourage error detection and feedback, thereby fostering a more effective learning environment for beginners. Educators can draw on these nuanced insights to develop effective pedagogical strategies for teaching FSs to beginner language learners and when assessing their proficiency in FSs through speech production. Employing a nuanced assessment allows for a more detailed understanding of errors, enabling teachers and learners to identify specific areas of improvement and tailor teaching approaches accordingly.

Although not explored in the research, another tentative implication could be that the process of strengthening the association of FSs in learners' memory requires them to notice how these sequences are structured when broken down into individual units. Given learners' interest in accuracy of speech, understanding how FSs are composed of smaller, comprehensible units may play a pivotal role in their integration and retention within working memory. This methodology aligns with the Schmidt's Noticing Hypothesis (1990, 2001), directing learners' attention towards the forms and meanings of FSs. For example, learners could learn how to apply the FS *my name is X* using different subjects and verbs, expanding their knowledge construction, access and variation (Schmidt, 2001; Schmidt, 2012). Exploring this potential benefit in future research could offer valuable insights into the cognitive processes associated with integrating FSs into learners' domain knowledge and retention. This investigation would contribute to a deeper understanding of language learning dynamics, offering enhanced perspectives on how learners incorporate and retain FSs within their domain-specific language proficiency.

It is also worth noting that RQ5 revealed deep insights for second language pedagogy, as elicited through the short-task interviews. Firstly, participants expressed a distinct need for increased speaking practices within the classroom (64.7% of responses). Secondly, they recognised the importance of learning FSs (54.9% of responses), citing benefits such as facilitating sentence construction, facilitating spoken communication, assisting in handling daily interactions and facilitating memorisation of sequences (Chapter 4, Results, Section 4.3.5). Some even mentioned the importance of learning individual words, but highlighted that gaining proficiency in phrases holds even greater importance, demonstrating a high awareness of the relevance of FSs in language usage. This alignment with participants' intuitions suggests a positive correlation with the research aims. Moreover, the findings suggest that the perceived disconnect does not stem from a misalignment between learners' internal needs and desires. Instead, it appears to be a gap between these needs and the reality of classroom teaching. Notably, classroom instruction tends to emphasise individual words (80.4% or responses), indicating the participants may not encounter new lexical items, something that contradicts the recommendations of Nation (2007) and more recently Pellicer-Sanchez and Boers (2019). These scholars assert that repetitive encounters with lexical items play a crucial role in strengthening retention and fostering familiarity, particularly benefiting learners with minimal prior English instruction, like the participants in this study.

The qualitative data that I used for this project serves as an accessible research paradigm that educators can use to make gradual improvements to their teaching practices. A simple way to begin with involves surveying learners at the beginning of the L2 course, seeking their insights on what they appreciate about their L2 learning and how they believe L2 learning could be improved. After collecting learner opinions, the teacher can then reflect on this feedback along with their own observations and consider ways to improve classroom learning. Adjustments can then be tailored to meet learners' needs, and subsequent surveys can gauge the effectiveness of these adaptations. Creating a simple feedback loop like this will help teachers to comprehend learners' needs within their context, culture and preferences.

5.11 Summary of the chapter

In this chapter, I presented the study findings addressing the five research questions. I explored the utilisation of FSs among college learners of English, mainly beginners, examining how these FSs were employed into their spoken samples. I linked the findings emerging from the descriptive results to the relevant literature in that area. I suggested that even beginners use formulaic sequences to a certain extent in their speech production. I argued that participants' preferences for simplicity over complexity in their use of FSs in speech samples may be influenced by factors such as time pressure, WM and anxiety about making mistakes. Furthermore, this finding highlights the distinction between explicit and implicit knowledge, suggesting the need to encourage learners to practice FSs for them to sound natural in their production. Research Question 1 concluded with practical implications for the classroom, including discussion about the integration of FSs in pedagogy to enhance L2 learners' speaking tasks.

The second research question investigated the relationship between the complexity, accuracy and fluency of students' entire speech samples and their utilisation of FSs during a speaking task. Results revealed accuracy as the most predictive variable. The relationship between accuracy and FSs was emphasised, suggesting that as FS usage increases, accuracy of the speech samples improves.

The third research question consisted of two parts. The first part aimed to examine correlations between ESL learners' individual differences and their performance in CAF-SA. It was emphasised that an important discovery was made regarding the relationship between WM and speech complexity. I explained that higher WM capacity could potentially assist students in producing more complex sentences. Furthermore, the vital role of familiarity and exposure in language learning was discussed. Findings showed scores from TF-FSs that participants had already encountered correlated best with their CAF scores in speech, as opposed to the remaining tests, including the general FSs test by Martinez (2011), which participants had not been exposed to. This observation is noteworthy when considering the complexity aspect of certain FSs, including transparency aspects and length of multi-word units.

The second part of RQ3 focused on identifying the best predictors among learners' individual differences for predicting scores in CAF-SA. I discussed the findings emerging from hierarchical linear regression analysis, and the association between TF-FSs and accuracy was highlighted, suggesting the retention of FSs from familiar tests. Additionally, participants prioritised accuracy over complexity and fluency, aligning with Skehan's trade-off hypothesis. It was also found that WM only associated with scores of complexity of speech. I have related this finding from the hierarchical linear regression to the relevant literature in this area. Given the mixed findings with regards to the association of WM and accuracy and fluency of speech, I argued that there are other factors that should be taken into consideration, including the methodological approach and the participants' proficiency levels. It is clear that future research is required on the association of WMC and CAF scores of speech, and a specific framework of the metric of measuring WM should be further investigated thoroughly and systematically.

Research question 4 involved a twofold investigation. The first part examined correlations between ESL learners' individual differences and their performance in CAF-FS in speaking. The second part of RQ4 aimed to investigate which individual differences serve as the best predictors for CAF-FSs scores. I presented the findings derived from hierarchical linear regression analysis. The scores from TF-FSs were further confirmed to be the best and only significant predictor of scores of C, A and F of FSs in speech. This emphasises the importance of incorporating certain pedagogical methodologies in the L2 classroom, such as memorisation and consistent practice. Additionally, I referred to Gatbonton's and Segalowitz's (2005) ACCESS approach, advocating for the automated mastery of selected utterances (like FSs) to enhance their application across various communicative interactions as a beneficial framework that language educators can refer to in the classroom. On the other hand, I clearly noted that the TF-FSs might have emerged as the best predictor for other reasons, such as lower proficiency levels, and it is important to acknowledge that results could have differed at higher levels of proficiencies.

The final research question RQ5, delved into students' perceptions regarding the importance of FSs for language learning, with a specific focus on enhancing fluency. The aim of this RQ was to complement earlier findings in this study. The high prioritisation of accuracy of speech was underscored, in line with findings from previous research questions and emphasising accuracy's significance as a predictor, especially in comparison to fluency and complexity. I argued that this suggested that L2 learners often prioritise planning their utterances with accuracy, with the goal of minimising speech errors. While providing participants with FSs and encouraging their use in speech has demonstrated gain in fluency

(Wood; 2009, 2010, 2016), it is noteworthy that this study showed further impact of L2 speech production. Beginners not only improved fluency but also demonstrated enhancements in accuracy and complexity, as shown in RQ2, RQ3a and RQ3b. This extends the scope of Wood's findings with regards to fluency, emphasising the variety of benefits of FSs in speech production. It was also interesting to see that beginner learners were aware of the importance of FSs, especially in overcoming difficulties in sentence construction, reinforcing the value of relying on prefabricated FSs. Overall, this chapter contributes to a comprehensive understanding of L2 learning dynamics, highlighting the central role of FSs in addressing diverse challenges faced by English learners.

Finally, this chapter concluded with pedagogical implications to be implemented in the classroom. I argued for the adoption of methodologies such as repetition, memorisation and rehearsal, emphasising their potential to effectively facilitate the learning of FSs, especially among beginner learners of English. I recommended that language educators adopt the ACCESS approach, and I have provided a justified rationale for this recommendation. These pedagogical insights can serve as valuable guidance for educators seeking to optimise language learning experiences and outcomes for their students.

Chapter 6. Conclusion

6.1 Introduction

In this chapter, I present the main findings of this study, emphasising their significance for SLA and research into Complexity, Accuracy and Fluency (CAF). I discuss the contributions that this thesis has made at both theoretical and methodological levels. Furthermore, I outline the implications of the present research within theoretical and methodological domains. Additionally, I provide further insights into pedagogical implications beyond the scope of the earlier discussion chapter. To conclude, I address some of the limitations of the current research and suggest potential directions for future research in the field.

6.2 Conclusions from the findings

This study was designed to investigate the use of formulaic sequences (FSs) in speech production by examining, 1) students' application and usage of FSs, 2) the relationship between CAF of students' overall speech and their use of FSs, 3) the correlation between learners' individual differences and CAF in speaking, 4) the learners' individual differences that best predict scores of CAF, 5) the correlation between learners' individual differences and CAF of the FSs in speaking, 6) the learners' individual differences that best predict scores of CAF of the FSs, and 7) students' perceptions of the importance of FSs for language learning in general and, specifically, for enhancing fluency. The participants in this study performed a monologic task focused on presenting a self-introduction in a work–life context, a topic derived from their textbook specifically from a chapter that they had already encountered. Subsequently, they participated in a post-task interviews to share their perceptions towards FSs learning and speaking in the classroom. Participants also underwent various background languages tests, a Working Memory test and completed a language history questionnaire regarding their L2 learning context and language use practices. Their speech performances were recorded, transcribed and analysed for measures of complexity, accuracy and fluency of entire speech. The novelty of the current study was that the FSs extracted from participants' speech samples were analysed for complexity, accuracy and fluency. The qualitative data from the post-task interviews complemented the quantitative analysis, providing a more comprehensive understanding of the findings.

The findings of this research revealed that the use of FSs had a significant impact on the complexity, accuracy and fluency of speech produced by Kuwaiti beginner learners of English, who have yet to acquire the first 2,000 words of the language. Regarding the first research question concerning participants' usage of FSs in a speaking task, the findings indicated that beginner learners did incorporate FSs in their speech samples. However, it was noteworthy that the majority of participants used simpler FSs, characterised by a less formal register compared to those found in the textbook. This is despite their familiarity with the meanings of more complex FSs, as confirmed by the test of familiarity of FSs. This result is supported by Ahmadian and Tavakoli's (2011) finding that under time pressure, speakers prioritise expressing their message over using complex language, as complexity may overload their working memory. Additionally, this finding corresponds with Ullman & Lovelett's (2018) discussion on explicit versus implicit knowledge. The fact that participants are familiar with complex FSs yet choose simpler alternatives that serve the same communicative purpose highlights the distinction between declarative knowledge (knowing FSs) and procedural knowledge (using these FSs in practice). Clearly, participants in this study demonstrate explicit knowledge of the meaning of FSs, yet exhibit a gap in their ability to effectively apply this knowledge in practical speech.

The second research-question explored the relationship between CAF of students' entire speech samples and their use of FSs in a speaking task. Univariate regression analyses revealed a statistically significant association between CAF of speech and the number of FSs elicited from the speaking task, with accuracy emerging as the most predictive variable among complexity and fluency. This result suggests, given FSs are stored and retrieved as one whole unit (Tavakoli & Uchihara, 2020; Wood, 2020), they are produced with a high degree of accuracy since they are not subject to analysis. Indeed, the results of this study showed a high degree of accuracy in the use of FSs. When errors did occur in the FSs, they rarely caused disfluency, with the exception of one example found in this study. Thus, this study suggests that FSs used by L2 learners could be inaccurate but still likely maintain fluency.

The finding aligns with Boers et al. (2006), who also identified significant results in the relationship between the use of FSs and fluency. However, the current study found a highly significant association between the utilisation of FS and accuracy—a relationship that was not emphasised in Boers et al.'s interventional study. This result suggests that, at this level of proficiency (i.e., beginners), the greater the number of FSs used in speech samples, the more accurate learners' speech becomes.

The first part of research question 3 (RQ3a) explored the correlation between learners' general language proficiency, vocabulary knowledge, working memory and the results of their language history questionnaire, and their performance in CAF of speech. The findings showed a significant correlation between working memory (WM) and the complexity of speech, supporting findings of Gilabert (2010), Kormos and Trebits (2011) and Awaad and Tavakoli (2022) on the effect of WM on speech complexity. However, the absence of significant correlations with accuracy or fluency implies that the relationship between accuracy and fluency has yet to be investigated, especially considering the mixed findings seen in prior studies.

Furthermore, the correlation matrix revealed a strong relationship between familiarity and the CAF of the entire speech sample. This underscores the importance of familiarity and context specificity for speech production. In contrast, the FS test (Martinez, 2011), encompassing unfamiliar FSs, demonstrated a weaker association with CAF of entire speech, reinforcing the importance of prior exposure and familiarity with the domain knowledge.

The second part (RQ3b) extended the investigation to identify variables predicting CAF scores. The hierarchical linear regression analysis confirmed the pivotal role of WM as significant predictor of speech complexity, which confirms findings of the available literature (Gilabert & Muoz, 2010; Guará-Tavares, 2013; Kormos & Trebits, 2011; Awaad and Tavakoli, 2022). Additionally, the findings revealed associations between complexity, accuracy and fluency of speech with the number of FSs elicited from speech samples, expanding on the positive impact of increased exposure to FSs on fluency observed in previous studies (Bui & Huang, 2018; Wood, 2009, 2010, 2016). Therefore, the findings suggested that FSs not only impact fluency but also contribute to enhanced speech complexity and accuracy. Interestingly, the observed trade-off effects between accuracy and fluency, as indicated by scores from the familiarity test and participants' emphasis on producing error-free speech, as highlighted in post-task interviews where some participants explicitly mentioned their focus on avoiding errors in their speaking performance, align with Skehan's trade-off hypothesis. This finding also resonate with Yuan and Ellis's (2003) study on the effects of planning on CAF in speech, were a trade-off of accuracy over fluency was found among all groups in their study. However, the study suggests that these trade-off effects may vary among individuals, emphasising the role of personal preferences in prioritising specific speech qualities (i.e., form or meaning) (Section 5.6.3).

The first part of research question 4 (4a) explored correlations between learners' individual difference variables and CAF-FS in speaking. Notably, scores from the familiarity test demonstrated the strongest correlation with CAF for FSs. This finding supports previous studies (Boers et al., 2006; Bui & Huang, 2018; Wood, 2009; 2010; 2016) where participants benefitted from prior knowledge in specific domains, facilitating idea retrieval. While Boers et al. (2006) focused on familiarity with FSs leading to improved performance in terms of fluency and range of expressions, additional support comes from both Bui and Huang's (2018) study and Wood's series of studies (2009, 2010, 2016), demonstrating that familiarity results in notable gains in fluency. The present study further establishes a positive correlation between familiarity with FSs (specifically TF-FSs) and all three facets of speech production: complexity, accuracy and fluency of FSs. Therefore, familiarity with FSs seems to enable learners to overcome challenges in speech production, enabling them to produce language of higher complexity, accuracy and fluency. Furthermore, the moderate positive correlation between vocabulary test score, general proficiency scores and CAF of FSs suggested that individual difference variables can only predict a small proportion of the variability in the complexity, accuracy and fluency of FSs in speech production. This finding can be attributed to the nature of FSs, primarily consisting of prefabricated units retrieved from the mental lexicon as a whole (Boers et al., 2006; McGuire & Larson-Hall, 2018; Tavakoli and Uchihara, 2020; Wood, 2010, 2020; Wray, 2002). Therefore, speakers might not engage in formulating new messages from scratch in terms of both vocabulary and syntax. This brings in the advantage of learning FSs as chunks, which allows for their rapid and effective employment (Segalowitz & Hulstijn, 2005).

In the second part (4b), the investigation into variables *(general proficiency, FS, familiarity with FSs—TF-FSs, working memory, age of using English via online games*) predicting scores of CAF-FSs revealed that TF-FSs scores made the only significant contribution. This outcome suggests that familiarity and exposure play a crucial role in enhancing the predictability of complexity, accuracy and fluency of FSs scores derived from speech samples. Increased exposure to FSs allows learners to identify them and their correct meanings in a test, which emphasises the importance of familiarity in predicting FSs performance.

Finally, the last research question (RQ5) investigated participants' perceptions of the importance of FSs for L2 learning, specifically in enhancing fluency. The needs and insights expressed by participants in the short-task interviews correspond with the research aims, as participants were already aware the FSs facilitate speech production in various ways. Additionally, the study underscores the gap between learners' internal needs and desires and the current reality of classroom teaching practices. Notably, tailoring a curriculum to align with learners' perceptions and intuitions highlights the importance of needs analysis in shaping a more effective and learner-centric language learning environment, as shown from the findings of this study. This extends to ensuring the curriculum's suitability for students' proficiency levels. Learners with a vocabulary of fewer than 2000 words in the language cannot effectively be taught ESP, as detailed in Sections 3.5 and 3.6.3. This recognition emphasises the need for alignment between curriculum design and the actual proficiency levels of the learners.

6.3 Contributions of this research

To begin with, unlike previous research, such as studies by Khodadady & Shamsaee (2012) and McGuire and Larson-Hall (2018), which often focused on exploring the relationship between a single variable, such as the frequency of FSs, and learners' level of speaking ability, this study takes a more expansive approach. It explores multiple linguistic and cognitive variables, aiming to explain which variables affect the use of FSs among L2 learners. This examination contributes to a more holistic understanding of the intricate relationships governing the incorporation of FSs in language production, providing insights for future research and language instruction strategies.

Secondly, while previous studies of FSs mostly targeted at least intermediate-level proficiency level students, this study stands out as one of the few that focuses on beginner learners (CEFR A1-A2) and that explores linguistic, cognitive and background variables affecting speech production. Given the crucial role of accommodating the needs of learners at various language proficiency levels, especially in the formative beginner proficiency, and recognising the importance of FSs in enhancing learner fluency; therefore gaining insights into variables that mediate the learning of FSs in speech production at lower proficiency levels become even more essential.

Third, in addition to examining learners at lower levels, this research is the first, to the best of my knowledge, to specifically address those who need to learn the minimum amount of L2 for their future careers. This involves learning specific FSs familiar to students' language domains, especially tasks relevant to their future needs, in a scaffolded and supported context to enable them to perform at or near their maximum potential. In simpler terms, it focuses on equipping learners with FSs that facilitate communication in their business future careers, allowing them to express themselves clearly. Unlike some studies that investigated content familiarity within students' language domains, such as Bui and Huang's (2018) study on nursing and computer majors, which did not explore the use of FSs and involved participants at an upper-intermediate level, and Boers et al.'s (2006) examination of familiarity with FS use in speech production, where participants were at upper-intermediate to advanced levels and focused on fluency, range of expressions and accuracy, this study extends its scope to include complexity of FSs, which had received little attention so far. This investigation also revealed that familiarity with FSs had a significant association with all aspects of CAF: complexity of FSs, accuracy of FSs and fluency of FSs in speech. These findings are of particular interest especially to the domain of pedagogy, and can be used by publishers, researchers and educators to improve the design SLA curricula, in terms of what FSs should be taught and how they can be integrated in SLA textbooks. These findings will be particularly relevant for the Kuwaiti context, as they can be used to enhance the learning opportunities for Kuwaiti learners and may address possible shortcomings found in their textbook.

Fourth, unlike previous studies that focused solely on the relationship between the cognitive individual difference variable—WM and CAF of entire speech (e.g., de Paula et al., 2021; Kormos and Trebits, 2011; Awaad and Tavakoli, 2022), this study takes a more expansive approach. The key distinction lies in the broader perspective adopted, wherein this study not only investigates the relationship between WM and CAF of entire speech but also extends its exploration to include the relationship between WM and CAF of FSs. Given the crucial role of WM in language learning, specifically in speaking ability, understanding how learners' WM predicts their use of FSs in live performance was even more necessary. This investigation aimed to determine whether learners with higher WM retrieved more complex, accurate and fluent FSs in their speech performance. Therefore, an important contribution of the current study lies in the exploration of WM across the three qualities of speech—complexity, accuracy and fluency of FSs. This study took a step toward understanding the relationship between WM and the CAF of FSs but did not reveal any significant correlations (refer to Results Chapter, specifically RQ4a on Page 168). However, this result is not necessarily discouraging at this stage. There is a possibility that potential changes or enhancements to the methodology, particularly in investigating the role WM in the CAF of FSs elicited from live performances, especially for learners at lower proficiency levels, might be required to capture a relationship. Given that this is an initial exploration and there is limited literature available in the field regarding the role of WM and CAF of FSs, it is not yet time to dismiss the influence of WM on the CAF of FSs. Further investigating is required.

Moreover, unlike many studies (e.g., Awaad & Tavakoli, 2022; de Paula et al., 2021; Guará-Tavares; 2013) that have explored the role of WM and CAF of entire speech, often focusing among intermediate or advanced learners, this study contributed by directing attention to how WM impacts speech production among beginners. The findings showed that beginner learners followed the same pattern as in other studies involving intermediate and/or advanced learners with regard to complexity. The research contributed in showing that WM had an impact on the complexity of speech for lower proficiency learners, but it did not affect the accuracy or fluency of their speech.

An important methodological contribution of the current research lies in the operationalisation of complexity of FSs. My reading on the relevant literature has helped me to design certain criteria to measure the complexity of FS. This included counting the number of syllables that comprise a FS, addressing transparency issues and incorporating monolingual speaker judgements. This is a novel contribution in the literature, as no prior research paper has undertaken the assessment of complexity of FS, unlike the common and standardised practice of measuring the complexity of entire of speech samples. This contribution would help researchers, educators and those interested in pedagogy to assess the complexity of FSs, providing insights into how the complexity of these sequences can enhance overall complexity of speech. Furthermore, it enables the evaluation of learners' proficiency in using FSs by measuring the complexity of their sequences. Additionally, it can help teachers to take informed decisions about which expressions are suitable for different proficiency levels–more complex FSs for more advanced learners, etc. This would enable language educators to equip learners with language tailored to their proficiency levels.

Consequently, learners would be motivated to learn and use the language more effectively in their communication.

6.4 Implications of this research

Researching the relationship between FSs and speech production holds significant importance for both the research community and practitioners. Gaining insights into how to integrate FSs into the English classroom, with a specific focus on enhancing speaking production, is crucial. Drawing on theories related to speech production processes, previous research in CAF research has predominantly focused on understanding the pivotal role these processes play in comprehending the mechanisms involved in learners' L2 performance (de Bot, 1992, 2020; Kormos, 2006; Levelt, 1989; Segalowitz, 2010). The investigation of speech processes in the L2 has been recognised as cognitively challenging, emphasising the need to explore the psycholinguistic aspects underlying L2 fluency development, including gaining automaticity in using FSs. In the majority of studies, the relationship between FSs and speech production has been explored primarily for intermediate proficiencies, with a predominant focus on fluency of speech (Goncharov, 2019; McGuire & Larson-Hall, 2018; Mirzaei et al., 2016; Tavakoli and Uchihara; 2020). Consequently, L2 researchers have drawn conclusions about speakers' performance with FSs, highlighting how the use of FSs facilitate fluency in speech. For instance, learning FSs can prepare students for speaking exams like IELTS, enhancing their chances of achieving better scores, as these tests often have fixed questions that heavily rely on the use of FSs (Goncharov, 2019; Khodadady & Shamsaee, 2012; Mirzaei et al., 2016). Additionally, there is evidence suggesting that practising the use of FSs contributes to increased fluency of speech (McGuire & Larson-Hall, 2018; Tavakoli and Uchihara; 2020; Wood, 2009; 2010; 2016). Very few studies have attempted to analyse speech samples to determine to what extent there is a correlation and association between various individual difference variables (including linguistic and cognitive variables) and FSs produced via a speaking task. Moreover, the question of which individual difference variable most predicts beginner learners' speech samples in terms of both CAFS—entire speech samples and FSs—remains largely unexplored. The current research has been an attempt to shed some light on these questions and explores the role of linguistic variables, cognitive variables and background variables in explaining the complexity, accuracy and fluency of both entire speech and FSs. A number of theoretical, methodological and extended pedagogical implications have emerged from the findings, which are discussed in the following sections.

6.4.1 Theoretical implications

When examining the studies conducted by Khodadady and Shamsaee (2012) and McGuire and Larson-Hall (2018) in Chapter 1, Introduction, Section 1.1, it was noted that these research often focused on exploring the relationship between a single variable, such as the frequency of FSs, and learners' proficiency in speaking. In contrast, this study adopts a more expansive approach, delving into multiple linguistic and cognitive variables. The goal is not only to establish correlations but also to predict the use of FSs. This examination offers a more nuanced understanding of how various individual difference variables impact the incorporation of FSs in language production. The ability to predict a greater variety of variables and determine correlations among these variables and the complexity, accuracy and fluency of both entire speech and CAF of FSs in a speaking task, as demonstrated in the current study, provides deep insights for other researchers in the field. For example, a key finding was that the cognitive variable—WM-significantly contributed to explaining variances in the complexity of speech among beginners. This opens avenues for researchers to build upon this work and examine deeper the relationship between WM and complexity of speech, exploring why WM did not mediate accuracy or fluency of speech among beginner learners. For practitioners, such as language educators, a more comprehensive understanding of FSs use in a speech production can inform their teaching strategies. Educators can improve the effectiveness of pedagogical strategies by using learners' domain-specific FSs knowledge. Furthermore, an increase in familiarity with FSs related to a specific domain can lead to enhanced fluency within that domain, emphasising the importance of building familiarity with both the subject matter (i.e., domain) and associated FSs.

Further, when discussing Levelt's speech production model in Section 2.3.3, it was argued that the processes underlying speech production are not always smooth and efficient. This is because the string of words speakers produce is often interrupted by a variety of disfluencies, including short pauses, repetitions, repairs and filler words (e.g., uh and um) (de Jong et al., 2018). Such disfluencies sometimes occur when speakers have difficulty deciding how to formulate or articulate their speech (de Jong et al., 2018). I explained that this is especially the case in L2 speech since learners' linguistic knowledge is not extensive or yet automatic. In this study, it was found that familiarity with a specific domain facilitates the use of FSs, especially in L2 speech where learners' linguistic knowledge may not yet be extensive or automatic. The findings imply that using FSs helps speakers avoid the effort of putting together the components of an expression every time they want to use it, as easier access to FSs allows for automaticity in response, thereby facilitating speech production. Specifically, the study showed that low-proficiency learners may compensate for limited vocabulary by relying more on FSs, given their prior exposure to these sequences. In other words, the use of FSs can serve as a compensatory mechanism for low-proficiency learners, allowing them to overcome limitations in their vocabulary and fluency. One theoretical perspective could involve transitioning from teaching individual vocabulary items to focusing on teaching FSs. Once learners, especially beginners, acquire a set of these sequences, deriving meanings of individual words from FSs can become achievable by breaking down the components of the FSs. Additionally, the findings suggest that the use of FSs can contribute to the automaticity of speech production, as speakers can access these sequences more quickly and effortlessly compared to constructing novel utterances. These findings support suggestions made in previous research (e.g., Bui and Huang, 2018) and provide a better understating of L2 fluency performance. More importantly, these findings could make valuable contributions to the development of a better-informed L2 speech production model. For example, the formulator, responsible for converting conceptual representations into specific linguistic forms, including words, phrases, and grammatical structures, as well as encoding them phonologically, can be accommodated by the use of FSs. This suggests that incorporating FSs into L2 speech production models can enhance their explanatory power and provide a more comprehensive account of the processes involved in speech production.

In addition to low-proficiency learners compensating for limited vocabulary by relying more on FSs due to their prior exposure to these sequences, this study observed that scores from the general proficiency test only predicted a small variation in learners' production of complex, accurate and fluent FSs. This reinforces previous research suggesting that FSs primarily consist of prefabricated phrases that are often retrieved from the mental lexicon as whole units (Boers et al., 2006; McGuire & Larson-Hall, 2018; Tavakoli & Uchihara, 2020; Wood, 2010, 2020; Wray, 2002). Consequently, speakers might not engage as much in the process of formulating new messages in terms of both vocabulary and syntax, as they can rely on FSs as prefabricated united retrieved from the mental lexicon as whole phrases. One implication is that while general proficiency is important in language learning, its influence might be less significant regarding the utilisation and integration of FSs into spoken language. The use of FSs can compensate for second language learners' challenges in constructing sentences from scratch, both in terms of vocabulary and syntax. This means that speakers can compensate for processing difficulties in the formulation stage by using FSs, which are typically learned as chunks, allowing for their rapid and effective employment

(Segalowitz & Hulstijn, 2005), without necessarily relying on general language knowledge to assemble these multi-word combinations.

Another interesting implication concerning L2 speech production models involves the increased demand on cognitive resources, particularly WMC, when learners are under time pressure (Bygate, 2018; Kormos, 2011). It was found that WMC did not predict the use of FSs in speaking performance. In other words, this study did not find a relationship between WM and the CAF of FSs. This suggests that the use of FSs functions as a compensatory strategy, offering an alternative means to overcome the challenges associated with limited WMC. In essence, the use of FSs may accommodate learners with lower WM capacities. This highlights the importance of integrating FSs into language learning curricula, as they could be used as a strategy to enhance learners' language use in terms of complexity, accuracy and fluency, even for individuals with lower working memory capacity.

6.4.2 Methodological implications

The current findings have a number of methodological implications that can offer valuable insights to researchers in the fields of SLA and fluency. The first of these relates to the study's focus on beginner learners (CEFR A1-A2). Previous research on FSs in speech production mainly concentrated on intermediate to advanced proficiency levels (Goncharov, 2019; McGuire & Larson-Hall, 2018; Nergis, 2021; Tavakoli and Uchihara; 2020). Therefore, this study attempts to enhance the current methodology by investigating linguistic, cognitive and background variables specifically in the context of beginner learners, thus providing a comprehensive methodological framework tailored to this proficiency level. For example, the measure of accuracy of FSs in the literature is often measured in a binary way—either considering them accurate or inaccurate. This study, however, deviates from this approach, as detailed in Chapter 2, Literature review, Section 2.3.7, where some exceptions are outlined. These exceptions include cases where verb agreement (third person singular -s), prepositions and articles are incorrectly uttered or are omitted. Recognising the challenges faced by beginner learners, especially considering their limited language exposure, this non-binary measure for assessing accuracy of FSs accommodates for variations influenced by the learners' L1.

Another methodological implication is investigating the role of WM in the use of CAF of FSs in speech production. Here I aimed to introduce a cognitive dimension to the methodology of CAF of FSs as it has not been investigated before. This expands the scope of research methodologies by considering cognitive individual difference variables in the analysis of FSs in speech production. This is an important methodological implication as it explores the role of cognitive variables in the use and retention of complex, accurate and/or fluent FSs in a live performance.

The present research identified FSs elicited via participants' speech samples and textbook using computer technology–TAALES software (2.2) (Kyle & Crossley, 2015; Kyle et al., 2018). As was described in Section 3.10.2, FSs were extracted automatically based on their MI score via TAALES. Surely, automatic MI score analyses make it feasible to identify FSs in a data set. Yet, it is relatively uncommon to find studies which provide detailed descriptions of the identification of FSs and how TAALES was used in their research, the challenges faced or the potential issues with using this software. I would like to note that while the software could be very useful in identifying FSs (or trigrams as it is referred to in the software), it requires a very detailed manual inspection of FSs uttered with errors by participants as this software can only find error-free examples. This issue of inaccurate FSs used by L2 learners highlights the importance of doing manual checks when using the software for identifying and extracting FSs.

Finally, the most important methodological implication of this research is the novel method for operationalising the complexity of FSs. As discussed in the previous chapters, while the literature often addresses the complexity of entire speech, the examination of complexity of entire speech remains relatively rare. This methodological innovation holds considerable relevance for SLA researchers and practitioners, enabling a deeper understanding of how the complexity of FSs contributes to the overall complexity of speech.

Measuring the complexity of FSs not only provides insights into how complex the FSs used by a learner is but also allows for a more comprehensive evaluation of a learner's proficiency in using these sequences. The operationalisation of complexity of FSs becomes a valuable tool for aligning learners' proficiency levels with their language outcomes. This alignment, in turn, facilitates informed decisions regarding the selection of expressions suitable for different proficiency levels, advocating for the incorporation of more complex FSs for advanced learners.

6.4.3 Further pedagogical implications

This research was motivated from the very beginning to enhance classroom teaching and learning. It is therefore important to share the various pedagogical implications that emerge from this study and their potential impact on the language teaching field. Perhaps the most pressing question for practitioners is whether to implement oral skills in the classroom, especially when dealing with beginner learners in a non-immersive setting. This decision is complicated by challenges such as limited vocabulary, grammatical knowledge and L2 communicative skills among beginners, coupled with constraints on available classroom time. Moreover, the fact that speaking is often not assessed in the classroom adds another challenge for not directing attention to speaking. My recommendation based on this research experience is clear: if building familiarity with common expressions significantly mediates speech production, then teachers should allocate focused class time for teaching FSs.

The study highlighted the importance of selecting relevant expressions to fulfil learners' specific needs, specifically in using the L2 for their future careers, as illustrated by the participants in this study. The findings showed the more the exposure to FSs, the more learners can identify them and their correct meanings in a test. This suggests that familiarity with the FSs could predict the participants' ability to incorporate and employ these FSs in their spoken language. Thus, in turn, serves as a facilitator for their scores in complexity, accuracy and fluency. Based on these findings, I recommend language educators to implement supportive practices of FS use through speech production followed by a constructive feedback on their performance. This approach, combining production with feedback, not only strengthens the understanding of FSs but also fosters a more robust integration of these sequences into learners' spoken language. Particularly beneficial for learners aiming to acquire the necessary amount of L2 to express themselves and address their future career needs—a common goal among all participants in this study.

In the next section I will consider the limitations of this research and the implications for future methodology.

6.5 Limitations and directions for future research

While every effort has been made to ensure the robustness of the current study's findings, I should also acknowledge some of the limitations. Firstly, this study has been limited to female participants, and the inclusion of male students could have enabled comparison between genders, allowing for broader generalisations to all Kuwaiti students in the college of business. Future research should aim to replicate these results with males students to validate the findings across genders. Additionally, the current study had a limited number of students, and it is recommended to increase the participant pool to enhance both the reliability and generalisability of the findings.

Moreover, a limitation of the current study lies in the choice of task, which involved a relatively simple, monologic format. This reliance on a monologic task may have limited the range of language production observed. Different tasks, such as less structured free-speech tasks like oral interviews, or more structured tasks like picture description, could have produced different results. For example, in a less structured task, participants might rely more on FSs, potentially affecting the fluency or complexity of their speech. Future studies could explore task variations, such as incorporating a narrative task in addition to or instead of the monologic task, to investigate their potential effects on the relationship between FSs and CAF. This exploration could enhance the discussion and provide deeper insights onto the findings.

In the current study, FSs emerged as robust predictors of complexity, accuracy and fluency of speech. However, it is important to acknowledge that the strength of the relationship between FSs and L2 speech proficiency may vary depending on the task structure. For example, Saito and Liu (2021) found that FS measures, such as MI, had a notably stronger predictive power for participants' L2 speech proficiency in picture narrative tasks compared to interview tasks. This difference in predictive power could be attributed to task characteristics: in picture narrative tasks, speakers tend to prioritise accuracy over complexity, as they are provided with content and can allocate cognitive resources to employ FSs they have learned. Conversely, in interview tasks, speakers may focus more on content than linguistic form, leading to a weaker association between FSs and L2 speech proficiency. Future research could delve into how task type/complexity influence the relationship between FSs and CAF in L2 speech production, providing deeper insights into how task demands shape linguistic performance.

Furthermore, a notable limitation in this study stems from the fact that the analysed FSs were extracted from the same text sample used to compute the CAF of speech. This introduces a potential constraint as the independent variable (FS use) is not entirely independent from the dependent variable (CAF of speech), possibly leading to increased correlations between the use of FSs and CAF scores. Despite this limitation, it is important to emphasise that the study also explored the impact of FS familiarity on both overall speech (CAF of entire speech) and the specific use of FSs (CAF of FSs). In this aspect, the test of familiarity of FSs remained independent from the original sample, enabling a more focused examination of participants' familiarity with the FSs. To enhance the robustness of future studies, there is a need to explore strategies for increasing the independence between the independent variable (FSs use) and the dependent variables (CAF of speech). One possible way could involve excluding FSs from speech samples when computing CAF scores. However, it is essential to consider the proficiency level of participants, particularly with novice learners, as this method may pose challenges to lower proficiency level learners due to potential limitations in the availability of speech samples for analysis.

This research has yielded crucial insights into the relationship between CAF of speech and working memory. However, it is noteworthy that this study did not establish a similar association between CAF of FSs and working memory. It is necessary for further future studies to delve deeper into the relationship between CAF of speech and working memory, particularly focusing on beginner learners. Additionally, it is equally important to advance research by exploring the role of WM in shaping CAF of FSs across all language proficiency levels, as this investigation has not been explored in the existing literature. This broader investigation would contribute to a more comprehensive understanding of how WM can account for variability in CAF of FSs. In addition, it is clear that research studies should work towards the standardisation of WM measures to facilitate a better understanding of the role of this variable in SLA. To echo the recommendations of Wright (2015), it is only by utilising instruments that have demonstrated validity and reliability that we can effectively unravel the intricacies of WM's role in language learning.

Another limitation in this study lies in my ability to directly compare this study with others in the literature. This constraint arises due to the limited number of correlational studies examining the relationship between various variables and FSs produced via a speaking task. Even among those utilising a correlation design, the focus tends to be on correlating a single variable, such as the frequency of FSs use, with the learners' level of speaking ability. Therefore, there is a need for future studies to explore specific individual difference variables using a design similar to the one employed in this study, spanning from beginners to advanced proficiency levels. It would be interesting to investigate whether the general FS test, like Martinez's (2011), can explain more variance in CAF of speech and CAF of FSs compared to a more specific test, such as the test of familiarity of FSs, particularly among higher intermediate to advanced learners. Also, it is recommended to recreate situations where participants are familiar with both simple and complex forms of the same FS, as indicated in the findings of this study. Delving into the reasons behind their preference for using the simple form and correlating it with levels of proficiency, planning time constraints and time pressure would be a fruitful area for future work.

Finally, the Covid19 circumstances necessitated the administration of all instruments online, making it challenging to ensure the completion of data collection instruments such as the questionnaire, which is a factor that would have been facilitated through face-to-face access to participants. Also, the range of challenges included the collection of data online, which was a new experience for me and likely for the participants too.

6.6 Final remarks

This thesis began with the main question to what extent there is a correlation and association between Kuwaiti students' oral fluency and their knowledge and use of FSs. The suggestions of the literature as well as my own personal observation of people's L1 speech led me to investigate FSs use and individual difference variables that impact the use of FSs in ones speech samples. Employing a multiple regression design to predict participants' speech performance, this study stands out as the first to examine various individual differences variables predicting two kinds of CAF: of entire speech and of FSs. Additionally, it explored the mediating role of WM on CAF of FSs and investigated FSs use among beginner proficiency learners, contributing novel insights on these relationships and proficiency levels. A significant contribution lies in the development of a new measure—complexity of FSs, as this measure has not been explored before.

While the main focus in this study was on identifying the best predictor for CAF of speech and CAF of FSs among Kuwaiti college learners, specifically those yet to acquire the first 2,000 words of the language, a small part of the study also delved into learners' perceptions and intuitions regarding the importance of FSs for language learning and for enhancing fluency. The findings suggested that familiarity with specific FSs related to a specific language domain was the only individual difference variable that associated with CAF of FSs. This suggests the importance of building familiarity with common expressions in speech production, advocating for dedicated class time to teach FSs.

Furthermore, the study revealed that accuracy of entire speech emerged as the most predictive variable for the number of FSs used in a speaking task. This implies that the higher the number of FSs used, the more accurate speech becomes. Therefore, at the beginners' proficiency level, FSs serve as a valuable resource for developing learner accuracy. Learners, as indicated by the test of familiarity of FSs and post-task interviews, prioritised accuracy of speech in their ESP course, highlighting a conscious effort to produce error-free speech. This observation can be leveraged in L2 classrooms to meet learners' needs effectively.

The research findings demonstrated that WM, was associated only with complexity of speech. Learners with higher working memory capacity (WMC) were found to produce more complex speech, emphasizing the role of WM in shaping speech complexity. Despite the findings indicating that individuals with higher WM tend to use more complex language, it is crucial to note that higher WM should not pose a challenge for those with lower WM. For learners with lower WM, the use of FSs serves as a compensatory strategy, providing an alternative to overcome the challenges associated with limited WMC. FSs are stored and retrieved as a whole prefabricated units, therefore, can be used as strategy for learners facing the constraints of lower WM. This can be explained in that since WM is an individual difference that learners bring to the task of language learning, FSs may serve as an external variable that compensate for individual differences such as WM. This distinction is important as it highlights the role of FSs as a strategic tool to enhance speech production, addressing complexities in language use, accuracy and fluency, as shown from the findings of this study.

Additionally, the research revealed that learners are aware of the benefits of FSs in facilitating speech production. This alignment with participants' intuitions suggests a positive correlation with the research aims. Therefore, the perceived gap does not stem from a misalignment between learners' internal needs and desires. Instead, it appears to be a gap between these needs and the reality of classroom teaching which underscore the importance of tailoring curricula to align with learners' perceptions, intuitions and abilities. Therefore, I emphasise the need for needs analysis in creating a more effective and learner-centric language learning environment, as shown from the findings of this study. Teachers are encouraged to guide learners in using FSs through well-structured methodologies, as outlined in the literature review. These include practices like repetition, memorisation and rehearsal.

By aligning pedagogical methods with relevant theoretical frameworks, educators can foster enhanced learning experiences (Webster, 2015) with a specific focus on addressing the use of FSs in speech production.

References

Abdelhamid, G. S. M., Gómez-Benito, J., Abdeltawwab, A. T. M., Abu Bakr, M. H. S., & Kazem, A. M. (2019). Hierarchical structure of the wechsler adult intelligence scale–fourth edition with an Egyptian sample. *Journal of Psychoeducational Assessment*, *37*(3), 395-404.

Ahmadian, M. J. (2011). The effect of 'massed'task repetitions on complexity, accuracy and fluency: does it transfer to a new task?. *The Language Learning Journal*, *39*(3), 269-280.

Ahmadian, M. J. (2012). The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*, *16*(1), 129-149.

Ahmadian, M. J. (2013). Working memory and task repetition in second language oral production. *Asian Journal of English Language Teaching*, *23*(1), 37-55.

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.

Alharbi, H. A. (2015). Improving Students' English Speaking Proficiency in Saudi Public Schools. *International journal of instruction*, *8*(1), 105-116.

Alonso, R. A. (Ed.). (2018). *Speaking in a second language* (Vol. 17). John Benjamins Publishing Company.

Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior research methods*, *50*(1), 250-263.

Ariyanti, A. (2016). Psychological Factor Affecting EFL Students' Speaking Performance. *Journal of Language Teaching and Applied Linguistics*, *1*(1), 91-102.

Awaad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*, *60*(2), 169-196.

Bachman, L. F., Palmer, A. S., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory Trends in Cognitive Sciences, 4 (11), 417–423.

Baddeley, A. D. (2001). Is working memory still working?. *American psychologist*, *56*, 851-864.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829-839.

Baddeley, A. D. (2015). Working memory in second language learning. *Working memory in second language acquisition and processing*, 17-28.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL quarterly*, *26*(2), 390-395.

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

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*(3), 245-261.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83-110.

Bøhn, H. (2015). Assessing spoken EFL without a common rating scale: Norwegian EFL teachers' conceptions of construct. *Sage Open*, *5*(4), 2158244015621956.

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing*. Sage publications.

Bryman, A. (2004). Social research methods. Oxford: Oxford University Press.

Bryman, A. (2012). Social research methods: OUP Oxford.

Bui, G. (2014). Task readiness: Theoretical framework and empirical evidence from topic familiarity, strategic planning, and proficiency levels. In *Processing perspectives on task performance* (pp. 63-94). John Benjamins.

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.

Burns, A. (2019). Concepts for teaching speaking in the English language classroom. *LEARN Journal: Language Education and Acquisition Research Network*, *12*(1), 1-11.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. *Challenge and change in language teaching*, *136*, 146.

Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language teaching research*, *3*(3), 185-214.

Bygate, M. (2001). Speaking. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 14–20). Cambridge: Cambridge University Press.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. *Current trends in the development and teaching of the four language skills*, *29*.

Bygate, M. (2009). 15. Teaching the spoken foreign language. *Handbook of foreign language communication and learning*, *6*, 401.

Bygate, M. (2015). Domains and directions in the development of TBLT. *Domains and Directions in the Development of TBLT*, 1-349.

Bygate, M., & Alonso, R. A. (2018). Creating and using the space for speaking within the foreign language classroom: What, why and how. *Speaking in a second language*, *17*, 153-166.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), Planning and task performance in a second language (pp. 37–75). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Carroll, S. E. (2006a). Salience, awareness and SLA. In *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006)* (pp. 17-24). Somerville, MA: Cascadilla Proceedings Project.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95-122.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Ernst Klett Sprachen.

Chan, W. M., Chin, K. N., Bhatt, S., & Walker, I. (Eds.). (2012). *Perspectives on Individual Characteristics and Foreign Language Education* (Vol. 6). Walter de Gruyter.

Chen, Z., & Goh, C. (2014). Teacher knowledge about oral English instruction and teacher profiles: An EFL perspective. *Teacher Development*, *18*(1), 81-99.

Chomsky, N. (1959). Chomsky, N. 1959. A review of BF Skinner's Verbal behavior. Language, 35 (1), 26–58.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, *39*, e62.

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. 8th ED. Routledge.

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.

Conway, A. R. A., Jarrold, C., & Kane, M. J. & Miyake, A., & Towse, JN (Eds.) (2007). *Variation in working memory*. Oxford University Press.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Cooper, B. J. (2004). The enigma of the Chinese learner. *Accounting Education*, *13*(3), 289-310.

Correia, R. C. (2016). Assessing speaking proficiency: A challenge for the Portuguese EFL teacher. *e-TEALS*, *7*(1), 87-107.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, *34*(2), 213-238.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative* and qualitative research.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Cutler, S. F. (2020). *Developing formulaicity: memorisation and production of formulaic expressions in L2 speakers of English* (Doctoral dissertation, Cardiff University).

Dai, Z., & Ding, Y. (2010). Effectiveness of text memorization in EFL learning of Chinese students. *Perspectives on formulaic language: Acquisition and communication*, 71-87.

Dahlin, B., & Watkins, D. (2000). The role of repetition in the processes of memorising and understanding: A comparison of the views of German and Chinese secondary school students in Hong Kong. *British Journal of Educational Psychology*, *70*(1), 65-84.

Damanhouri, M. Y. (2018). The effectiveness of the noticing hypothesis in retrieving data from non-Arabic Qur'an memorizers. *Journal of Arts and Humanities*, *7*(12), 18-26.

Dao, P., Iwashita, N., & Gatbonton, E. (2017). Learner attention to form in ACCESS task-based interaction. *Language Teaching Research*, *21*(4), 454-479.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics*, *14*(2), 159-190.

De Bot, K. (1992). Applied linguistics. *Applied linguistics*, *13*(1), 1-24.

De Bot, K. (2020). A bilingual production model: Levelt's speaking model adapted. In *The bilingualism reader* (pp. 384-404). Routledge.

de Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, *9*, 541-560.

de Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237-254.

de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2013). 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., & 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. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5-34.

DeKeyser, R. (2015). Skill acquisition theory. In *Theories in second language acquisition* (pp. 94-112). Routledge.

DeKeyser, R. (2018). Task repetition for language learning. *Learning language through task repetition*, *11*, 27.

de Paula, S. G. C., Tavares, M. D. G. G., & Weissheimer, J. Working Memory Capacity and the Implementation of New Information During Task Repetition/Capacidade de memória de trabalho e implementação de informações novas durante a repetição de tarefa. *Caligrama: Revista de Estudos Românicos*, *26*(2), 107-123.

DiGiovanni, E., & Nagaswami, G. (2001). Online peer review: An alternative to face-to-face?. *ELT journal*, *55*(3), 263-272

Dillon, G. (2015). *A Formula for Fluency?: An Investigation Into the Effect of Instruction in Formulaic Sequences on Oral Fluency in Irish* (Doctoral dissertation, Trinity College Dublin).

Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System*, *35*(2), 271-280.

Ding, Y., & Qi, Y. (2005). Use of formulaic language as a predictor of L2 oral and written performance. *Journal of PLA University of Foreign Languages*, *28*(3), 49-53.

Dörnyei, Z. (2003). Questionnaires in second language research construction, administration, and processing. Lawrence Erlbaum Associates: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007). Research methods in applied linguistics. Oxford University Press.

Doughty, C. (2001). Cognitive underpinnings of focus on form in P. Robinson (ed.) Cognition and second language instruction (pp. 206-257).

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, *30*(1), 58-72.

Ellis, N. C. (1996). Analyzing language sequence in the sequence of language acquisition: Some comments on Major and Ioup. *Studies in Second Language Acquisition*, *18*(3), 361-368.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143-188.

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. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL quarterly*, *42*(3), 375-396.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in second language acquisition*, *9*(1), 1-19.

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

Ellis, R. (2005). Planning and task-based performance: Theory and research. In *Planning and task performance in a second language* (pp. 3-36). John Benjamins.

Ellis, R. (2008). Learner beliefs and language learning. *Asian EFL Journal*, *10*(4), 7-25.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied linguistics*, *30*(4), 474-509.

Ellis, R. (2014). Principles of instructed second language learning. In *Teaching English as a Second of Foreign Language* (pp. 31-45). Heinle Cengage Learning.

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

Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. Routledge.

Ellis, R., & Yuan, F. (2005). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in second Language acquisition*, *26*(1), 59-84.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29-62.

Eskildsen, S. W., & Cadierno, T. (2015). Advancing usage-based approaches to L2 studies. *Usage-based perspectives on second language learning*, 1-16.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Felker, E. R., Klockmann, H. E., & de Jong, N. H. (2018). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, *40*(1), 111-136.

Field, A. (2009). *Discovering statistics using SPSS: Book plus code for E version of text* (Vol. 896). London, UK: SAGE Publications Limited.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles: Sage Publications.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics+ Webassign*. Sage Publications.

Folse, K. S., & Briggs, S. J. (2004). *Vocabulary myths: Applying second language research to classroom teaching* (No. Sirsi) i9780472030293). Ann Arbor: University of Michigan Press.

Foster, P., & Skehan, P. (2013). Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. *Canadian modern language review*, *69*(3), 249-273.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), 354-375.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98-116.

Garbati, J. F., & Mady, C. J. (2015). Oral skill development in second languages: A review in search of best practices. *Theory and Practice in Language Studies*, *5*(9), 1763.

Gass, Susan, Alison Mackey, María José Alvarez-Torres, and Marisol Fernández-García. "The effects of task repetition on linguistic output." *Language learning* 49, no. 4 (1999): 549-581.

Gatbonton, E. (1994). Preparing and implementing a task-based ESL curriculum in an EFL setting: Implications for theory and practice. *TESL Canada Journal*, 09-29.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325-353.

Gavin, B. U. I., & WONG, C. H. (2021). From Linguistic Skills to Pragmatic Competence: The Role of Functional Adequacy in Task-Based Teaching and Learning. *Asian Journal of English Language Teaching*, *30*, 61-76.

Gibbons, P. (2007). Mediating academic language learning through classroom discourse. In *International handbook of English language teaching* (pp. 701-718). Boston, MA: Springer US.

Galian-Lopez, G. (2018). Automaticity in Second Language Vocabulary Learning. *Language Studies Working Papers*, *9*, 26-34.

Ghazi-Saidi, L., & Ansaldo, A. I. (2017). Second language word learning through repetition and imitation: Functional networks as a function of learning phase and language distance. *Frontiers in Human Neuroscience*, 463.

Gilabert, R., & Muñoz, C. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies*, *10*(1), 19-42.

Goh, C. C. M. (2007). *Teaching speaking in the language classroom*. Singapore: SEAMEO Regional Language Centre.

Goh, C. C. M. & Burns, A. (2012). *Teaching speaking: A holistic approach*. New York: Cambridge University Press.

Goncharov, G. (2019). THE EFFECT OF DIRECT INSTRUCTION IN FORMULAIC SEQUENCES ON IELTS STUDENTS'SPEAKING PERFORMANCE. *Advanced Education*, *6*(11), 30-39.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: Theory, analysis and applications*, 145-160.

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing.

Grant, L. E., & Nation, P. (2006). How many idioms are there in English?. *ITL-International Journal of Applied Linguistics*, *151*(1), 1-14.

Guará-Tavares, M. D. G. (2013). Working memory capacity and L2 speech performance in planned and spontaneous conditions: a correlational analysis. *Trabalhos em linguística aplicada*, *52*, 09-29.

Guest, G., & MacQueen, K. (2012). Applied thematic analysis. Los Angeles: Sage Publications.

Gudu, B. O. (2015). Teaching Speaking Skills in English Language Using Classroom Activities in Secondary School Level in Eldoret Municipality, Kenya. *Journal of Education and Practice*, *6*(35), 55-63.

Hale, J. B., Hoeppner, J. A. B., & Fiorello, C. A. (2002). Analysing digit span components for assessment of attention processes. *Journal of psychoeducational assessment*, *20*(2), 128-143.

Hall, T. (2010). L2 Learner-made formulaic expressions and constructions. *Studies in Applied Linguistics and TESOL*, *10*(2).

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, *32*(3), 465-491.

Hamada, M., & Koda, K. (2011). The role of the phonological loop in English word learning: A comparison of Chinese ESL learners and native speakers. *Journal of psycholinguistic research*, *40*(2), 75-92.

Harmer, J. (2007). The practice of English language teaching. U.K.: Longman.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in second language acquisition*, *14*(1), 25-38.

Hennecke, I., Perevozchikova, T., & Wiesinger, E. (2023). Multiword units in multilingual speakers. *International Journal of Bilingualism*, 13670069231200027.

Hinkel, E. (2020). *Teaching academic L2 writing: Practical techniques in vocabulary and grammar*. Routledge.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in psychology*, *4*, 149.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 461-473.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, *32*, 1-20.

Hsu, H. C. (2017). The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics*, *38*(3), 359-385.

Hughes, J., & Naunton, J. (2017). *Business Result: Intermediate: Student's Book with Online Practice*. Oxford University Press.

Hughes, R. (2010). Materials to develop the speaking skill. *English language teaching materials: Theory and practice*, 207-224.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*(1), 4-21.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in second language acquisition*, *24*(4), 541-577.

Jeong, H., & DeKeyser, R. (2023). Development of automaticity in processing L2 collocations: The roles of L1 collocational knowledge and practice condition. *Studies in Second Language Acquisition*, 1-25.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, *44*(10), e12885.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137-166.

Kawauchi, C. (2005). 5. The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. *Language Learning & Language Teaching*, 143-164.

Khodadady, E., & Shamsaee, S. (2012). Formulaic Sequences and Their Relationship with Speaking and Listening Abilities. *English Language Teaching*, *5*(2), 39-49.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition*, *37*(3), 549-581.

Koizumi, R., & In'nami, Y. (2013). Vocabulary Knowledge and Speaking Proficiency among Second Language Learners from Novice to Intermediate Levels. *Journal of Language Teaching & Research*, *4*(5).

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*. Routledge.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*, 39-59.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*(2), 261-271.

Kormos, J., Trebits, A., & Robinson, P. (2011). Working memory capacity and narrative task performance. *Second language task complexity*, 128.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words?. *Language Assessment Quarterly*, *13*(4), 377-392.

Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly*, *49*(4), 757-786.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods*, *50*, 1030-1046.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. *Handbook of bilingualism: Psycholinguistic approaches*, 289-307.

Lahmann, C., Steinkrauss, R., & Schmid, M. S. (2019). Measuring linguistic complexity in long-term L2 speakers of English and L1 attriters of German. *International Journal of Applied Linguistics*, *29*(2), 173-191.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167-196.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. oxford University.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA review*, *6*(1), 10-20.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning*, *54*(3), 399-436.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics*, *22*(1), 1-26.

Laufer, B., & Hulstijn, J. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, *51*(3), 539-558.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, *40*(3), 387-417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In *Perspectives on fluency* (pp. 25-42). University of Michigan.

Leonard, M. K., Torres, C., Travis, K. E., Brown, T. T., Hagler Jr, D. J., Dale, A. M., ... & Halgren, E. (2011). Language proficiency modulates the recruitment of non-classical language areas in bilinguals. *PloS one*, *6*(3), e18240.

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

Levelt, W. (1999). Producing spoken language. *The neurocognition of language*, 83-122.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2013). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 1-7.

Liang, H. (2017). An Empirical Study on the Effects of Computer-Corpus-based Formulaic Sequences on College Students' Oral English Learning. *International Journal of Emerging Technologies in Learning (iJET)*, *12*(08), 67-76.

Liao, G. (2009). Improvement of Speaking Ability through Interrelated Skills. *English language teaching*, *2*(3), 11-14.

Lightbown, P. (2008). Easy as pie? Children learning languages. *Concordia Working Papers in Applied Linguistics*, *1*, 1-25.

Lightbown, P. M., & Spada, N. (2021). *How Languages Are Learned 5th Edition*. Oxford university press.

Lindsay, C., & Knight, P. (2006). *Learning and teaching English: A course for teachers* (Vol. 11, No. 3). New York: Oxford University Press.

Littlewood, W. (2004). The task-based approach: some questions and suggestions. *ELT Journal,58*(4), 319-326.

Liu, Y. (2020). Relating lexical access and second language speaking performance. *Languages*, *5*(2), 13.

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. *Perspectives on formulaic language: Acquisition and communication*, 22-46.

Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*.

Lowie, W., Verspoor, M., & van Dijk, M. (2018). The acquisition of L2 speaking. *Speaking in a second language*, *17*, 105.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221-250.

Lynch, T., & Maclean, J. (2001). ‘A case of exercising’: Effects of immediate task repetition on learners' performance. In *Researching pedagogic tasks* (pp. 141-162). Routledge.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language teaching*, *46*(1), 1-40.

Mackey, A. J., Philp, J. J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (ed.), Individual differences and instructed language learning. Philadelphia, PA: John Benjamins, 181-209

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, *60*(3), 501-533.

Mackey, A., & Gass, S. (2005). Common data collection measures. *Second language research: methodology and design. Mahwah: Lawrence Erlbaum*, 43-99.

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. *Assessing multilingual children: Disentangling bilingualism from language impairment*, 95-124.

Martinez, R. (2011). Phrasal vocabulary size test, BNC version (1-5k).

Martinez, R. (2011, September). Putting a test of multiword expressions to a test. In *IATEFL Testing, Evaluation and Assessment Conference (TEA SIG), University of Innsbruck*. University of Innsbruck.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied linguistics*, *33*(3), 299-320.
Mayne, R. (2023). Straw methods: clearing up misconceptions about ALM. *Language & History*, *66*(3), 265-282.

McCarthy, M. (2006). Fluency and confluence: What fluent speakers do. *Explorations in corpus linguistics*, *1*.

McGuire, M., & Larson-Hall, J. (2018). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal*, *34*(3), 1-25.

McLean, S., Kramer, B. (2015). The creation of a New Vocabulary Levels Test. *Shiken*, *19*(2), 1-11.

Meara, P. (2010). EFL vocabulary tests (2nd ed.). Swansea: Centre for Applied Language Studies, University College.

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

Michel, M. (2017). Complexity, accuracy and fluency in L2 production. *The Routledge handbook of instructed second language acquisition*, *50*, 68.

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

Mirzaei, A., Hashemian, M., & Azizi Farsani, M. (2016). Lexis-based instruction and IELTS candidates' development of L2 speaking ability: Use of formulaicity in monologic versus dialogic task. *Journal of Teaching Language Skills*, *35*(2), 69-98.

Mitchell, R. M., & Myles, F. F., & Marsden, E.(2013). *Second language learning theories*. Routledge.

Mitchell, R., Myles, F., & Marsden, E. (2019). *Second language learning theories*. Routledge.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Mizera, G. J. (2006). *Working memory and L2 oral fluency* (Doctoral dissertation, University of Pittsburgh).

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language teaching research*, *12*(1), 11-37.

Mystkowska-Wiertelak, A., & Pawlak, M. (2017). *Willingness to communicate in instructed second language acquisition: Combining a macro-and micro-perspective*. Multilingual Matters.

Nation, I. S. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, *1*(1), 2-13.

Nation, I. S. P. (2011). Second language speaking. In *Handbook of research in second language teaching and learning* (pp. 444-454). Routledge.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA- word-family-lists.pdf

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. The language teacher, 31(7), 9-13.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge university press.

Nation, I. S. P. & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, *14*(1), 6-19.

Nation, I. S. P., & Waring, R. (2019). *Teaching extensive reading in another language*. Routledge.

Nemoto, T., & Beglar, D. (2014). Likert-scale questionnaires. In *JALT 2013 conference proceedings* (pp. 1-8).

Nergis, A. (2021). Comparative effectiveness of teaching formulaic sequences and academic vocabulary on complexity, accuracy and fluency of EAP speakers. *Konińskie Studia Językowe*, *9*(3), 287-313.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied linguistics*, *30*(4), 555-578.

Nunan, D. (1992). Research methods in language learning. Cambridge University Press.

O'malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge university press.

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. InR. Ellis (Ed.), Planning and task performance in a second language (pp. 78–209).

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder.

Oxford University Press & Cambridge Local Examinations Syndicate. (2004). Oxford quick placement test. Oxford: Oxford University Press.

Pang, F., & Skehan, P. (2014). Self-reported planning behaviour and second language performance in narrative retelling. *Processing perspectives on task performance*, 95-127.

Pakula, H. M. (2019). Teaching speaking. *Apples-Journal of Applied Language Studies*, *13*(1), 95-111.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing.

Paradis, M. (2009). *Declarative and procedural determinants of second languages* (Vol. 40). John Benjamins Publishing.

Parker Jones, Ō., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., ... & Price, C. J. (2011). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex*, *22*(4), 892-902.

Pellicer-Sánchez, A., & Boers, F. (2018). Pedagogical approaches to the teaching and learning of formulaic language. *Understanding formulaic language*, 153-173.

Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL quarterly*, *30*(2), 201-230.

Pulido, M. F. (2022). Why are multiword units hard to acquire for late L2 learners? Insights from cognitive science on adult learning, processing, and retrieval. *Linguistics Vanguard*, *8*(1), 237-247.

Punch, K. F. (2013). Introduction to social research: Quantitative and qualitative approaches.

London: Sage.

Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, *39*(2), 164-174.

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

Read, J. (2019). Assessment of Vocabulary. *The Encyclopedia of Applied Linguistics*.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*(1), 93-132.

Richards, J. C. (2015). *Key issues in language teaching*. Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge university press.

Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes*, *15*(2), 129-157.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211.

Sabbah, S. (2015). Negative transfer: Arabic language interference to learning English. *Arab World English Journal (AWEJ) Special Issue on Translation*, (4).

Saito, K. (2018). To What Extent Does Long-Term Foreign Language Education Help Improve Spoken Second Language Lexical Proficiency?. *TESOL Quarterly*, *1-26*.

Saito, K. (2020). Multi-or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, *70*(2), 548-588.

Saito, K., & Liu, Y. (2022). Roles of collocation in L2 oral proficiency revisited: Different tasks, L1 vs. L2 raters, and cross-sectional vs. longitudinal analyses. *Second Language Research*, *38*(3), 531-554.

Saito, K., & Uchihara, T. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, 1-12.

Sawyer, M., & Ranta, L. (2001). *Aptitude, individual differences, and instructional design* (pp. 319-353). na.

Scaramucci, M., & Boffi, A. (2013). Assessing English in South America. *The companion to language assessment*, *4*, 1690-1699.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, 357-385.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in second language learning*, *11*, 237-326.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.): *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on individual characteristics and foreign language education*, *6*, 27.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use* (Vol. 9). John Benjamins Publishing.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. An introduction. In N. Schmitt (Ed.), Formulaic Sequences (pp. 2–22).

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching1. *Language Teaching*, *47*(4), 484-503.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & cognition*, *33*(4), 644-653.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. *Handbook of bilingualism: Psycholinguistic approaches*, 371-388.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*(3), 837-842.

Shamsaee, S., & Shams, M. A. (2010). ESP TEACHERS’PEDAGOGICAL AGENDA VS. UNIVERSITY STUDENTS’EDUCATIONAL AMBITIONS: A NEEDS ANALYSIS PROJECT.

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, *62*(4), 339-34

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, *31*(4), 487-512.

Siyanova-Chanturia, A., & Sidtis, D. V. L. (2018). What online processing tells us about formulaic language. *Understanding formulaic language*, 38-61.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). Formulaic language: Setting the scene. In *Understanding Formulaic Language* (pp. 1-15). Routledge.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

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

Skehan, P. (2002). Theorising and updating aptitude. *Individual differences and instructed language learning*, *2*, 69-94.

Skehan, P. (Ed.). (2015). *Processing perspectives on task performance* (Vol. 5). John Benjamins Publishing Company.

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge.

Skehan, P., & Foster, P. (2001). *Cognition and tasks* (pp. 183-205). na.

Skehan, P., Bui, G., Wang, Z., & Shum, S. (2023). Re-Examining Accuracy Measures in Second Language Task-Based Spoken Performance. *Available at SSRN 4648364*.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual review of sociology*, *37*, 57-86.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied psycholinguistics*, *25*(2), 293-321.

Stroud, R. (2021). The effects of strategic planning and rehearsal on second language group discussion task performance. *The Language Learning Journal*, *49*(5), 499-512.

Suzuki, Y. (2023). PRACTICE AND AUTOMATIZATION IN SECOND LANGUAGE RESEARCH.

Storch, N., & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting?. *TESOL quarterly*, *37*(4), 760-770.

Stupar-Rutenfrans, S., Ketelaars, L. E., & van Gisbergen, M. S. (2017). Beat the fear of public speaking: Mobile 360 video virtual reality exposure training in home environment reduces public speaking anxiety. *Cyberpsychology, Behavior, and Social Networking*, *20*(10), 624-633.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, *15*, 165-179.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (Vol. 6, pp. 497-516). Boston, MA: pearson.

Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *61*, 37-72.

Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'neglected'in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, *22*(3), 330-349.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In *Planning and task performance in a second language* (pp. 239-273). John Benjamins.

Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency?. *Language Learning*, *70*(2), 506-547.

Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press.

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. (2020). *Developing fluency with multi-word expressions* (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).

Thomson, H., Boers, F., & Coxhead, A. (2019). 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*, *52*(3), 406-414.

Thornbury, S. (2005). *How to teach speaking*. Longman.

Thornbury, S. (2019). Learning language in chunks. *Part of the Cambridge Papers in ELT Series.* [pdf]. Cambridge University Press.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in second language acquisition*, *16*(2), 183-203.

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, 1362168818799371.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953-968). Academic Press.

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second language research*, *34*(1), 39-65.

VanPatten, B., Williams, J., Rott, S., & Overstreet, M. (Eds.). (2004). *Form-meaning connections in second language acquisition*. Routledge.

Van Os, M., de Jong, N. H., & Bosker, H. R. (2020). Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and nonnative speech. *Language Learning*, *70*(4), 1183-1217.

Van Vu, D., & Peters, E. (2022). The role of formulaic sequences in L2 speaking. In *The Routledge handbook of second language acquisition and speaking* (pp. 285-298). Routledge.

Vieira, G. V. (2017). Lexical access in L2 speech production: A controlled serial search task. *Ilha do Desterro*, *70*, 245-264.

Vilkaitė-Lozdienė, L., & Schmitt, N. (2019). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In *The Routledge handbook of vocabulary studies* (pp. 81-96). Routledge.

Wang, Z. (2014). On-line time pressure manipulations. *Processing perspectives on task performance*, *5*, 27-61.

Wang, Y. (2018). As Hill seems to suggest: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *Journal of English for Academic Purposes*, *33*, 12-23.

Wang, Z., Skehan, P., & Chen, G. (2019). The effects of hybrid online planning and L2 proficiency on video-based speaking task performance. *Instructed Second Language Acquisition*, *3*(1).

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, *44*(3), 263-277.

Webster, S. J. (2015). *Early Career ESOL Teachers' Practical Knowledge of Teaching Speaking* (Doctoral dissertation, University of Leeds).

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*(498), 1.

Wechsler, D. (2014). *WISC-V: Administration and scoring manual*. NCS Pearson, Incorporated.

Wen, Z. E., & Skehan, P. (2021). Stages of Acquisition and the P/E Model of Working Memory: Complementary or contrasting approaches to foreign language aptitude?. *Annual Review of Applied Linguistics*, *41*, 6-24.

Williams, J. N. (2012). Working memory and SLA. In *The Routledge handbook of second language acquisition* (pp. 445-459). Routledge.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

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). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Bloomsbury Publishing.

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

Wood, D. (2020). Classifying and identifying formulaic language. *Routledge handbook of vocabulary studies*, 30-45.

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

Wray, A. (2004). Here's one I prepared earlier': Formulaic language learning on television'in N. Schmitt (ed.): Formulaic Sequences: Acquisition, Processing and Use. *Philadelphia: J. Benjamins.[online](4th December 2011)*.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as. *Phraseology in foreign language learning and teaching*, *123*.

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

Wright, C. (2010). Role of working memory in SLA. Saarbrucken: VDM Publishing House

Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. *Understanding formulaic language: A second language acquisition perspective*, 19-37.

Wulff, S., Swales, J. M., & Keller, K. (2009). "We have about seven minutes for questions": The discussion sessions from a specialized conference. *English for specific purposes*, *28*(2), 79-92.

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, Y., & Liu, L. (2018). Using computer speech recognition technology to evaluate spoken English. *Educational Sciences: Theory & Practice*, *18*(5).

Appendices

Appendix 1: Post-task interview questions

Questions for participants

1. **How did your performance on the task go?**
   Possible follow up questions:
   - Are there comparable tasks that you typically encounter in your classroom activities?
   - Could you share what specific aspects or areas you concentrated on while performing the task?

2. **Let us talk about speaking in English. How often do you practice speaking activities in the classroom?**
   Possible follow up questions:
   - What is the focus of this course?
   - From your perspective, what specific needs or requirements do you have for this ESP course?
   - Why, in your opinion, is it important to direct attention on speaking activities in the classroom?

3. **Tell me more about vocabulary learning in the classroom**
   Possible follow up questions:
   - Do you only learn individual words?
   - How are new words typically introduced and taught in your classroom?
   - What do you think your needs of this ESP course are with regards to vocabulary learning?
   - After learning a new word/expression in the classroom, do your teacher review it, or do you encounter it again during the course?

4. **What, in your opinion, can be done to improve your In your opinion specific English experience?**

Appendix 2: List of formulaic sequences extracted from the students' textbook

- **Good evening!**
- **Let me introduce myself.**
- **I don't think we've met.**
- **Nice to meet you.**
- **I'm very pleased to meet you.**
- **What do you do?**
- **Where are you from?**
- **Do you have a card?**
- **I'd like to keep in contact.**
- **I'm afraid I have to go now.**
- **I look forward to hearing from you.**
- **Can I have your number?**
- **Can you repeat that?**
- **Can you say that again?**

Appendix 3: Monologic task

Speaking task It is your first meeting in a new company for a new job. • Introduce yourself • give your personal details • Talk about your experiences and projects • provide your contact details • end the monologue politely. *Please note that you can make up the information, experiences and projects that you talk about.*

Appendix 4: Sample of the Quick Placement Test

General Proficiency Test

Quick Placement Test – Part One (Questions 1 – 40)

Where can you see these notices?

1. Please leave your room key at Reception.

- ☐ in a shop
- ☐ in a hotel
- ☐ in a taxi

Where can you see these notices?

2. Foreign money changed here

- ○ in a library
- ○ in a bank
- ○ in a police station

Where can you see these notices?

3. AFTERNOON SHOW BEGINS AT 2PM

- ☐ outside a theatre
- ☐ outside a supermarket
- ☐ outside a restaurant

Where can you see these notices?

4. CLOSED FOR HOLIDAYS

Lessons start again on the 8 th January

- ☐ at a travel agent's
- ☐ at a music school
- ☐ at a restaurant

Where can you see these notices?

5. Price per night:

£10 a tent £5 a person

- ☐ at a cinema
- ☐ in a hotel
- ☐ on a camp-site

Questions 6 – 10

- In this section you must choose the word which best fits each space in the text below.
- For questions **6** to **10**, mark **one** letter **A**, **B** or **C**.

Scotland

Scotland is the north part of the island of Great Britain. The Atlantic Ocean is on the west and the North Sea on the east. Some people **(6)** .... Scotland speak a different language called Gaelic. There are **(7)** .... five million people in Scotland, and Edinburgh is **(8)** .... most famous city.
Scotland has many mountains; the highest one is called 'Ben Nevis'. In the south of Scotland, there are a lot of sheep. A long time ago, there **(9)** .... many forests, but now there are only a
(10) ....
Scotland is only a small country, but it is quite beautiful.

- ☐ 6) A) on
- ☐ 6) B) in
- ☐ 6) C) at
- ☐ 7) A) about
- ☐ 7) B) between
- ☐ 7) C) among
- ☐ 8) A) his
- ☐ 8) B) your
- ☐ 8) C) its
- ☐ 9) A) is
- ☐ 9) B) were
- ☐ 9) C) was
- ☐ 10) A) few
- ☐ 10) B) little
- ☐ 10) C) lot

Appendix 5: Sample of the New Vocabulary Level Test

Please select the option a, b, c, or d which has the closest meaning to the word in **bold**.

NVLT Part 1

1. time: They have a lot of **time**.
a. money
b. food
c. hours
d. friends

2. stone: She sat on a **stone**.
a. hard thing
b. kind of chair
c. soft thing of the floor
d. part of a tree

3. poor: We are **poor**.
a. have no money
b. happy
c. very interested
d. tall

4. drive: She **drives** fast.
a. swims
b. learns
c. throws balls
d. uses a car

5. jump: She tried to **jump**.
a. lie on top of the water
b. get up off the ground
c. stop the car on the road
d. move very fast

6. shoe: Where is your other **shoe**?
a. the person who looks after you
b. the thing you keep your money in
c. the thing you use for writing
d. the thing you wear on your foot

7. test: We have a **test** in the morning.
a. meeting
b. travelling somewhere
c. a set of questions
d. an idea to do something

8. nothing: He said **nothing** to me.
a. very bad things
b. zero
c. very good things
d. something

9.cross: Don't **cross**.
a. go to the other side
b. push something
c. eat too fast
d. wait for something

10. actual: The **actual** one is larger.
a. real
b. old
c. round
d. other

11. any: Does she have **any** friends?
a. some
b. no
c. good
d. old

12. far: You have walked **far!**
a. for a long time
b. very fast
c. a long way
d. to your house

13. game: I like this **game**.
a. food
b. story
c. group of people
d. way of playing

14. cause: He **caused** the problem.
a. made
b. fixed
c. explained
d. understood

15. many: I have **many**.
a. none
b. enough
c. a few
d. a lot

16. where: **Where** did you go?
a. at what time
b. for what reason
c. to what place
d. in what way

17. school: This is a big **school**.
a. where money is kept
b. sea animal
c. place for learning
d. where people live

18. grow: All the children **grew**.
a. drew pictures
b. spoke
c. became bigger
d. cried a lot

19. flower: He gave me a **flower**.
a. night clothes
b. small clock
c. beautiful plant
d. type of food

20. handle: I can't **handle** it.
a. open
b. remember
c. deal with
d. believe

21. camp: He is in the **camp**.
a. sea
b. place outside where people enjoy nature
c. hospital
d. building where people sleep

22. lake: People like the **lake**.
a. area of water
b. very young child
c. leader
d. quiet place

23. past: It happened in the **past**.
a. before now
b. big surprise
c. night
d. summer

24. round: It is **round**.
a. friendly
b. very big
c. very quick
d. with no corners

Appendix 6: Sample of the Phrasal Vocabulary Size Test

First 1000 [Go 2]

1. GO ON: It will **go on.**
   - a. ○sleep
   - b. ○repeat
   - c. ○be fast
   - d. ○continue

2. LEAD TO: No one knows what it will **lead to.**
   - a. ○want
   - b. ○have inside
   - c. ○cause in the future
   - d. ○find

3. SO THAT: He sat **so that** they could do it.
   - a. ○to make it possible that
   - b. ○because
   - c. ○very slowly and then
   - d. ○before

4. AT ALL: I don't like it **at all.**
   - a. ○all the time
   - b. ○in any way
   - c. ○at first
   - d. ○sometimes

5. I MEAN: Two, **I mean** three.
   - a. ○I am guessing
   - b. ○maybe
   - c. ○then later
   - d. ○I correct myself

6. AT LEAST: **At least** it is warm.
   - a. ○other things may be bad, but
   - b. ○many days have passed and now
   - c. ○I cannot believe that
   - d. ○The least important thing is

7. IS LIKELY TO: He **is likely to** go.
   - a. ○likes
   - b. ○can
   - c. ○wants
   - d. ○probably will

8. IS TO: He **is to** speak this afternoon.
   - a. ○will
   - b. ○can
   - c. ○wants to
   - d. ○may

9. DEAL WITH: I can **deal with** it.
   - a. ○fix
   - b. ○remember
   - c. ○find
   - d. ○see

10. USED TO: I **used to** go.
    - a. ○want to travel now
    - b. ○went there in the past
    - c. ○usually go there
    - d. ○always travel there

Appendix 7: Sample of the Test of Familiarity of FSs

Good evening!

- ☐ صباح الخير
- ☐ مساء الخير
- ☐ تصبح على خير
- ☐ لا اعلم

Let me introduce myself.

- ☐ دعني أعرف بنفسي
- ☐ دعني أراك
- ☐ دعني أتحدث عن نفسي
- ☐ لا اعلم

I don't think we've met.

- ☐ لا تتحدث بهذا الموضوع مجدداً
- ☐ سوف أراك قريباً
- ☐ لا أعتقد أننا التقينا
- ☐ لا اعلم

Nice to meet you.

- ☐ لم نلتقي منذ زمن
- ☐ سررت بلقائك
- ☐ سأراك غداً
- ☐ لا اعلم

I'm very pleased to meet you.

- ☐ تحب ان تشاركني وجبة الغداء
- ☐ اتمنى ان أراك في الاسبوع المقبل
- ☐ يسرني جدا أن ألتقي بك
- ☐ لا اعلم

What do you do?

- ☐ ماذا تسمع؟
- ☐ ماذا تعمل؟
- ☐ ماذا تفعل؟
- ☐ لا اعلم

Appendix 8: Sample of Digit Span Forward trials

Correct Responses

Item	Correct Response
1.	9 – 7
	6 – 3
2.	5 – 8 – 2
	6 – 9 – 4
3.	7 – 2 – 8 – 6
	6 – 4 – 3 – 9
4.	4 – 2 – 7 – 3 – 1
	7 – 5 – 8 – 3 – 6
5.	3 – 9 – 2 – 4 – 8 – 7
	6 – 1 – 9 – 4 – 7 – 3
6.	4 – 1 – 7 – 9 – 3 – 8 – 6
	6 – 9 – 1 – 7 – 4 – 2 – 8
7.	3 – 8 – 2 – 9 – 6 – 1 – 7 – 4
	5 – 8 – 1 – 3 – 2 – 6 – 4 – 7
8.	2 – 7 – 5 – 8 – 6 – 3 – 1 – 9 – 4
	7 – 1 – 3 – 9 – 4 – 2 – 5 – 6 – 8

Appendix 9: Sample of L2 Language History Questionnaire (3.0)

L2 Language History Questionnaire (Version 3.0, 2015)

See http://blclab.org/ for online use and credit

Participant ID: ____________________

1. Age (in years): __________

2. Sex (Circle one): Male / Female

3. Education (your current or most recent educational level, even you have not finished the degree) (Circle one):

- Graduate school (PhD/MD/JD)
- Graduate school (Masters)
- College (BA/BS)
- High school
- Middle school
- Other (specify): ___________

4. Have you ever studied or learned a second language in terms of listening, speaking, reading, or writing? (Circle one):
Yes / No

5. Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using each language.

Language	Listening	Speaking	Reading	Writing	Years of use[a]

a. You may have learned a language, stopped using it, and then started using it again. Please give the total number of years.

6a. Country of residence: ___________

6b. Country of origin: ___________

6c. If 6a and 6b are different, then when did you first move to the country where you currently live? _______________

7. If you have lived or travelled in countries other than your country of residence or country of origin for three or more months, then indicate the name of the country, your length of stay, the language you used, and the frequency of your use of the language for each country.

Country	Length of stay[a] [month(s)]	Language	Frequency of use[b]
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7

a. You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.

b. Please rate according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	2	3	4	5	6	7

8. Indicate the age at which you started using each of the languages you have studied or learned in the following environments.

Language	At home	With friends	At school	At work	Language software	Online games

2

Appendix 10: Sample of L2 Language History Questionnaire (3.0) (in Arabic)

Create Configure Publish Analyze Print Preview

Quiz: Language History Questionnaire **Status:** Open

4) مستوى التعليم

- ○ جامعة
- ○ كلية
- ○ ثانوي
- ○ متوسط
- ○ أخرى

5) تعليم الأب

- ○ دراسات عليا (دكتوراه)
- ○ دراسات عليا (ماجستير)
- ○ جامعة
- ○ كلية
- ○ ثانوي
- ○ متوسط
- ○ أخرى

6) تعليم الأم

- ○ دراسات عليا (دكتوراه)
- ○ دراسات عليا (ماجستير)
- ○ جامعة

7. اليد المسيطرة

- ○ اليمنى
- ○ اليسرى
- ○ أستخدم كلتا اليدين بالتساوي

Edit Copy Move Delete

8. اذكر لغتك الأم واذكر ايضا العمر الذي بدأت تعلم هذه اللغة و يشمل ذلك الاستماع، المحادثة، القراءة و الكتابة، وعدد سنوات ممارستك لهذه اللغة

[ ] اللغة الأم

[ ] الاستماع

[ ] المحادثة

[ ] القراءة

[ ] الكتابة

[ ] *سنوات الممارسة

Appendix 11: Role-play task

Student A	**Student B**
You work at a company and you meet an employee from a different company for the first time. Introduce yourself, ask about personal details, ask to keep in contact and end the conversation politely.	You are an employee. Exchange greetings with your colleague, answer and ask about personal details. Accept the request to keep in contact and end the conversation politely.

Appendix 12: Ethical Approval granted for the study

Memo

University of Reading

School of Literature and Languages
Department of English Language and Applied Linguistics

To **Khuloud Al-Ali**

From **Dr Christiana Themistocleous**

Date **5-7-2021**

Your application for Ethical Approval

Your project entitled **"An investigation into cognitive and linguistic variables that influence the learning and production of FSs by undergraduate students via a speaking task in an English-as-a-second-language context."** has been considered by the School Ethnics Committee, and I am pleased to report that the Committee raised no ethical objections and subject to your undertaking to store the consent forms in the Department Office the normal way, it is accordingly given permission for the project to proceed under the exceptions procedure as outlined in paragraph 6 of the *University's Ethics Guidance to Schools.*

Signed

Christiana Themistocleous
Dr Christiana Themistocleous
On behalf of the School Ethics Committee

Appendix 13: Information Sheet given to the participants (in English)

University of Reading

Khuloud Al-Ali
co841894@student.reading.ac.uk

Supervisor:
Jeanine Treffers-Daller
+44 (0) 118 378 7260
j.c.treffers-daller@reading.ac.uk

Department of English Language and Applied Linguistics School of Literature and Languages Edith Morley Building
PO Box 218
Whiteknights Reading RG6 6AA

Phone 0118378814
Email appling@reading.ac.uk

INFORMATION SHEET

This research is part of my dissertation, which will be submitted as partial fulfilment of the requirements for the PhD degree in Applied Linguistics at the University of Reading. The focus of this study is on speaking in English as a second language and on making recommendations for improving the teaching and learning of speaking in Kuwait.

You have been selected to take part in this study based on your level of education – you should have completed general English classes and are in your final specialised English class. You will be required to perform a speaking task, followed by a post-task interview. Additionally, you will be completing various tests, including a test of language familiarity and other tests such as language proficiency test, vocabulary test, fixed expressions test and working memory test. Afterward, you will be asked to complete a language history questionnaire. You will be contacted directly by the researcher, Khuloud Alali.

The data gathered from you will be securely stored on a password-protected computer, accessible only by the researcher, Khuloud Alali, and her supervisors. Additionally, recordings from the speaking task will be segmented, coded and scored by a blind rater(s) under my supervision in anonymised conditions. The anonymity of the data is guaranteed and will be used for academic purposes only and destroyed immediately after the completion of the thesis. Your privacy and confidentiality will be carefully observed and you have the rights to withdraw at any stage for any or no reason.

This project has been subject to ethical review by the School Ethics Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at j.c.treffers-daller@reading.ac.uk

Signed

Appendix 14: Information Sheet given to the participants (in Arabic)

Researcher:
Khuloud Al-Ali
co841894@student.reading.ac.uk

Supervisor:
Jeanine Treffers-Daller
+44 (0) 118 378 7260
j.c.treffers-daller@reading.ac.uk

Department of English Language and Applied Linguistics School of Literature and Languages Edith Morley Building
PO Box 218
Whiteknights Reading RG6 6AA

Phone 0118378814
Email appling@reading.ac.uk

صفحة المعلومات

هذه الدراسة هي جزءًا من رسالة الدكتوراه والتي سيتم تقديمها استكمالاً لمتطلبات درجة الدكتوراه في اللغويات التطبيقية في جامعة ريدينج .تركز الدراسة على التحدث باللغة الإنجليزية كلغة ثانية وعلى تقديم توصيات لتحسين التدريس وتعلم المحادثة في الكويت

تم اختيارك للمشاركة في هذه الدراسة وفقاً إلى مستوى تعليمك - على ان يكن قد اتممتِ مقررات اللغة الإنجليزية العامة وأنت في صفك التخصصي النهائي . ستكونين مطالبة بأداء تمرين محادثة، تليها مقابلة .بالإضافة إلى ذلك، ستكملين مجموعة من الاختبارات، بما في ذلك اختبار اتقان اللغة، واختبار المفردات، واختبار المصطلحات الثابتة، واختبار الذاكرة العاملة. بعد ذلك، سيُطلب منك أن تكملين استبيان تاريخ اللغة .سيتم التواصل بك مباشرة من قبل الباحثة، خلود العلي

سيتم الاحتفاظ بالبيانات التي يتم جمعها منك بشكل آمن على جهاز الكمبيوتر المحمي بكلمة مرور، ولن يتم الاطلاع عليها الا من قبل الباحثة، خلود العلي، ومشرفيها على الاطروحة .بالإضافة إلى ذلك، سيتم/ استخراج التسجيلات من مهمة التحدث ومن ثم تجزئتها وترميزها وتقييمها من قبل مقيم/ مقيمين بصورة سرية تحت اشراف الباحثة، خلود العلي، وفق شروط محددة ان سرية البيانات والمعلومات مضمونة وسيتم استخدامها لأغراض اكاديمية بحتة، وسيم اتلافها فورا الانتهاء من الاطروحة سيتم مراعاة خصوصيتك وسرية معلوماتك بعناية ولديك حقوق الانسحاب في أي مرحلة لأي سبب او بدون سبب

خضعت الدراسة لمراجعة أخلاقية من قبل لجنة اخلاقيات الجامعة، وعليه تم السماح للباحثة، خلود العلي، في المضي قدما بموجب الإجراءات الاستثنائية على النحو المبين في الفقرة 6 من ملاحظات الجامعة التوجيهية بشأن أخلاقيات البحث

إذا كان لديك أي استفسارات أو ترغبين في استيضاح أي امر يتعلق بالدراسة، أرجو ان لا تترددين في التواصل بمشرفتي على العنوان أعلاه أو عبر البريد الالكتروني للمشرفة
j.c.treffers-daller@reading.ac.uk

التوقيع

Appendix 15: Consent Form (in English)

School of Literature and Languages
Department of English Language and Applied Linguistics

ETHICS COMMITTEE

Consent Form

Project title: An investigation into cognitive and linguistic variables that influence the learning and production of FSs by undergraduate students via a speaking task in an English-as-a-second-language context.

I understand the purpose of this research and understand what is required of me; I have read and understood the Information Sheet relating to this project, which has been explained to me by Khuloud Al-Ali. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 16: Consent Form (in Arabic)

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

نموذج الموافقة

عنوان الاطروحة: التحقيق/ البحث في المتغيرات المعرفية واللغوية التي تؤثر على تعلم واستخدام العبارات الثابتة من قبل طالبات الكلية من خلال تطبيق تمرين المحادثة في سياق اللغة الإنجليزية كلغة ثانية.

أفهم الغرض من هذه الدراسة وافهم ما هو المطلوب مني، لقدر قرأت وفهمت ورقة المعلومات المتعلقة بالدراسة، والتي تم شرحها لي من قبل الباحثة خلود العلي. أوافق على الترتيبات الموضحة في ورقة المعلومات والمتعلقة بمشاركتي في الدراسة.

أفهم ان مشاركتي اختيارية تماما ولي الحق في الانسحاب من المشاركة في الدراسة في أي وقت.

لقد حصلت على نسخة من نموذج الموافقة هذا وورقة المعلومات المرفقة معه (المصاحبة له).

الاسم:

التوقيع:

التاريخ:

Appendix 17: A Sample of The PHRASE List

Phrasal expressions divided to match 1K frequency bands of the most common word families in the BNC. The Integrated List Rank represents where each item falls when both lists (individual and phrase lists) are merged together.

There are three categories of genre with frequency information to help the list user discern the appropriateness/usefulness of each phrase. The frequency information breaks down as follows:

* * * = phrase most common in this genre (or as common)

* * = phrase less common in this genre

* = phrase infrequent in this genre

X = phrase rare or non-existent in this genre

1k phrases

Integrated List Rank	Phrase	Frequency (per 100 million)	Spoken general	Written general	Written academic	Example
107	**HAVE TO**	83092	* * *	* *	*	I exercise because I **have to**.
165	**THERE IS/ARE**	59833	* * *	* * *	* *	**There are** some problems.
415	**SUCH AS**	30857	*	* * *	* * *	We have questions, **such as** how it happened.
463	**GOING TO (FUTURE)**	28259	* * *	* *	**x**	I?m **going to** think about it.
483	**OF COURSE**	26966	* * *	* *	*	He said he?d come **of course**.
489	**A FEW**	26451	* * *	* *	*	After **a few** drinks, she started to dance.
518	**AT LEAST**	25034	* * *	* *	* *	Well, you could email me **at least**.
551	**SUCH A(N)**	23894	* * *	* *	*	She had **such a** strange sense of humor.
556	**I MEAN**	23616	* * *	**x**	**x**	It?s fine, but, **I mean**, is it worth the price?
598	**A LOT**	22332	* * *	*	**x**	They go camping **a lot** in the summer.
631	**RATHER THAN**	21085	* * *	* * *	* * *	Children, **rather than** adults, tend to learn quickly.
635	**SO THAT**	20966	* * *	* *	*	Park it **so that** the wheels are curbed.
655	**A LITTLE**	20296	* * *	* *	*	I like to work out **a little** before dinner.
674	**A BIT (OF)**	19618	* * *	* *	**x**	There was **a bit** of drama today at the office.

717	**AS WELL AS**	18041	* * *	* * *	* * *	She jogs **as well as** swims.
803	**IN FACT**	15983	* * *	* * *	* *	The researchers tried several approaches, **in fact**.
807	**(BE) LIKELY TO**	15854	* * *	* * *	* * *	To be honest, I?m **likely to** forget.
825	**GO ON**	15610	* * *	* * *	*	He **went on** for a while before stopping for lunch.
845	**IS TO**	15232	**x**	* * *	**x**	Obama **is to** address the media this afternoon.
854	**A NUMBER OF**	15090	* *	* * *	* * *	**A number of** concerns were raised.
879	**AT ALL**	14650	* * *	* * *	* *	Do you have any kids **at all**?
888	**AS IF**	14470	* *	* * *	**x**	They walked together **as if** no time had passed.
892	**USED TO (PAST)**	14411	* * *	*	**x**	It **used to** snow much more often.
894	**WAS TO**	14366	**x**	* * *	* *	The message **was to** be transmitted worldwide.
908	**NOT ONLY**	14110	* * *	* * *	* * *	**Not only** was it cheap, it was delicious.
913	**THOSE WHO**	13951	**x**	* * *	* * *	He would defend **those who** had no voice.
934	**DEAL WITH**	13634	* * *	* * *	* *	The police had several issues to **deal with**.
939	**LEAD TO ('CAUSE')**	13555	*	* * *	* *	Excessive smoking can **lead to** heart disease.
951	**SORT OF**	13361	* * *	**x**	**x**	It?s **sort of** why I?m here.
974	**THE FOLLOWING**	12963	**x**	*	* * *	He made **the following** remarks.
984	**IN ORDER TO**	12762	* *	* * *	* * *	We shared a room **in order to** reduce costs
988	**HAVE GOT (+NP)**	12734	* * *	*	**x**	I don?t know what he **has got** planned.

Appendix 18: Univariate regression analysis of number of FSs used in speaking task

Table 1. Univariate regression analysis associating Number of FSs used in speaking task with several predictor variables separately.

Predictor	Number of FSs used in speech samples				
	B	*p*	S.E	95% CI	*β*
Complexity of speech	.709	<.001	.101	.507–.912	.709
Accuracy of speech	.817	<.001	.082	.651–.982	.817
Rate A	.726	<.001	.726	.528–.923	.726
Rate B	.720	<.001	.099	.521–.919	.720

Appendix 19: Univariate regression analysis of CAF of speech

Table 1. Univariate regression analysis associating Complexity of entire Speech with several predictor variables separately.

Predictors	Complexity of entire speech sample			
	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.568	<0.001	0.332–0.805	0.568
Vocabulary test	0.608	<0.001	0.380–0.836	0.608
FS test	0.522	<0.001	0.277–0.767	0.522
Test of familiarity of FS	0.668	<0.001	0.454–0.881	0.668
Number of FSs used in speech samples	0.709	<0.001	0.507–0.912	0.709
DSS	0.352	0.011	0.084–0.621	0.352
Working memory	0.369	0.008	0.102–0.636	0.369
Age of onset of learning English	-0.257	0.069	-0.534–0.021	-0.257

Table 2. Univariate regression analysis associating Accuracy of entire Speech with several predictor variables separately.

Predictors	Accuracy of entire speech sample			
	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.556	<0.001	0.317–0.794	0.556
Vocabulary test	0.543	<0.001	0.302–0.784	0.543
FS test	0.429	0.002	0.170–0.688	0.429
Test of familiarity of FS	0.729	<0.001	0.532–0.925	0.729
Number of FSs used in speech samples	0.817	<0.001	0.651–0.982	0.817

Table 3. Univariate regression analysis associating Fluency of entire Speech with several predictor variables separately.

Predictors	Fluency (Rate A)				Fluency (Rate B)			
	B1 coef.	P value	95% CI	Stand. Beta coef.	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.543	<0.001	0.302–0.784	0.543	0.526	<0.001	0.282–0.770	0.526
Vocabulary test	0.589	<0.001	0.356–0.821	0.589	0.545	<0.001	0.304–0.785	0.545
FS test	0.450	<0.001	0.194–0.706	0.450	0.410	0.003	0.148–0.672	0.410
Test of familiarity of FS	0.643	<0.001	0.423–0.863	0.643	0.646	<0.001	0.427–0.865	0.646
Number of FSs used in speech samples	0.726	<0.001	0.528–0.923	0.726	0.720	<0.001	0.521–0.919	0.720

Appendix 20: Multivariate regression analysis of CAF of speech

Table 1. Multivariate regression analysis associating C-SA with several predictor variables.

Predictors	C-SA					
	B	*p*	S.E	95% CI	*β*	*F*
GP test	.173	.225	.141	-.110–.456	.173	F(4, 46) = 17.861, *p* <.001
FS test	.118	.381	.133	-.150–.386	.118	
TF-FS	.510	<.001	.110	.289–.731	.510	
WM	.293	.003	.095	.102–.484	.293	

*Significant at *p* = .012 after Bonferroni adjustment.

Table 2. Multivariate regression analysis associating A-SA with several predictor variables.

Predictors	A-SA					
	B	*p*	S.E	95% CI	*β*	*F*
GP test	.212	.153	.146	-.082–.505	.212	F(3, 47) = 20.698, *p*<.001
FS test	.027	.844	.135	-.245–.299	.027	
TF-FSs	.604	<.001	.114	.375–.833	.604	

*Significant at *p* = .016 after Bonferroni adjustment.

Table 3. Multivariate regression analysis associating F-SA with several predictor variables.

Predictors	Fluency (Rate A)					Fluency (Rate B)				
	B	*p*	S.E	95% CI	*β*	*B*	*p*	S.E	95% CI	*β*
GP test	.213	.194	.161	-.112–.538	.213	.224	.175	.163	-.103–.552	.224
FS test	.095	.527	.150	-.206–.397	.095	.039	.800	.151	-.265–.343	.039
TF-FSs	.488	<.001	.126	.234–.741	.488	.509	<.001	.127	.253–.765	.509

*Significant at *p* = .017 after Bonferroni adjustment.

Appendix 21: Univariate regression analysis of CAF of FSs

Table 1. Univariate regression analysis associating Complexity of FSs with several predictor variables separately.

Predictors	Complexity of FSs			
	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.473	<0.001	0.220–0.726	0.473
Vocabulary test	0.412	0.003	0.150–0.674	0.412
FS test	0.317	0.023	0.045–0.590	0.317
Test of familiarity of FS	0.668	<0.001	0.455–0.882	0.668
Number of FSs used in speech samples	0.990	<0.001	0.950–1.030	0.990
Age of onset of learning English via online games	-0.001	0.049	-0.001–0.000	-0.001

Table 2. Univariate regression analysis associating Accuracy of FSs with several predictor variables separately.

Predictors	Accuracy of FSs			
	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.469	<0.001	0.215–0.722	0.469
Vocabulary test	0.398	0.004	0.135–0.662	0.398
FS test	0.315	0.024	0.042–0.587	0.315
Test of familiarity of FS	0.648	<0.001	0.429–0.866	0.648
Number of FSs used in speech samples	0.989	<0.001	0.947–1.031	0.989

Table 3. Univariate regression analysis associating Fluency of FSs with several predictor variables separately.

Predictors	Fluency of FSs			
	B1 coef.	P value	95% CI	Stand. Beta coef.
General Proficiency test	0.433	0.002	0.174–0.692	0.433
Vocabulary test	0.378	0.006	0.112–0.643	0.378
FS test	0.277	0.049	0.001–0.553	0.277
Test of familiarity of FS	0.633	<0.001	0.411–0.856	0.633
Number of FSs used in speech samples	0.983	<0.001	0.931–1.036	0.983

Appendix 22: Multivariate regression analysis of CAF of FSs

Table 1. Multivariate regression analysis associating C-FSs with several predictor variables.

Predictors	Complexity of FSs					
	B	*p*	S.E	95% CI	*β*	*F*
GP test	.316	.056	.161	-.008–.641	.316	*F* (4,46) = 12.666, *p*<.001
FS test	-.125	.392	.145	-.418–.167	-.125	
TF-FS	.516	<.001	.125	.265–.767	.516	
Age of using L2 via online games	-.001	.024	.000	-.001–.000	-.250	

*Significant at *p* = .016 after Bonferroni adjustment.

Table 2. Multivariate regression analysis associating A-FSs with several predictor variables.

Predictors	A-FSs					
	B	*p*	S.E	95% CI	*β*	*F*
GP test	.219	.192	.166	-.114–.553	.219	*F* (3,47) = 12.436, *p*<.001
FS test	-.075	.629	.154	-.385–.235	-.075	
TF-FS	.561	<.001	.130	.300–.821	.561	

*Significant at *p* = .017 after Bonferroni adjustment.

Table 3. Multivariate regression analysis associating F-FSs with several predictor variables.

Predictors	F-FSs					
	B	*p*	S.E	95% CI	*β*	*F*
GP test	.198	.250	.169	.250–-.143	.198	*F* (3,47) = 11.262, *p*<.001
FS test	-.101	.523	.157	.523–-.417	-.101	
TF-FS	.569	<.001	.132	.303–.836	.569	

*Significant at $p = .017$ after Bonferroni adjustment.

Appendix 23: Hierarchical regression analysis of CAF of speech in the opposite direction

Table 1. Hierarchical regression analysis of predictors of C-SA conducted in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* Change	Model sig.
Model 1					.493	.503	<.001	*F* (1, 49) = 49.583, *p*<.001
FSs used by part.	.709	.101	<.001	.507–.912				
Model 2					.567	.082	.003	*F* (2, 48) = 33.782, *p*<.001
FSs used by part.	.479	.119	<.001	.239–.719				
TF-FSs	.367	.119	.003	.127–.607				
Model 3					.628	.066	.005	*F* (3, 47) = 29.167, *p*<.001
FSs used by part.	.470	.111	<.001	.247–.693				
TF-FSs	.254	.117	.035	.019–.490				
FS test	.283	.095	.005	.092–.474				
Model 4					.622	.001	.686	*F* (4, 46) = 21.528, *p*<.001
FSs used by part.	.461	.114	<.001	.232–.690				
TF-FSs	.243	.121	.050	.000–.487				
FS test	.251	.124	.048	.002–.500				
GP test	.055	.135	.686	-.217–.327				
Model 5					.669	.050	.008	*F* (5, 45) = 21.215, *p*<.001

FSs used by part.	.407	.108	<.001	.189–.625				
TF-FSs	.283	.114	.017	.053–.512				
FS test	.178	.119	.140	-.061–.417				
GP test	.079	.127	.535	-.176–.334				
WM test	.235	.085	.008	.063–.407				

*Significant at $p = .01$ after Bonferroni adjustment.

Table 2. Hierarchical regression analysis of predictors of A-SA in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig.
Model 1					.660	.667	<.001	F (1,49) = 98.121, $p<.001$
FSs used in speech samples	.817	.082	<.001	.651–.982				
Model 2					.734	.077	<.001	F (2, 48) = 69.843, $p<.001$
FSs used in speech samples	.593	.094	<.001	.404–.781				
TF-FSs	.357	.094	<.001	.169–.546				
Model 3					.744	.015	.090	F (3, 47) = 49.499, $p<.001$
FSs used in speech samples	.588	.092	<.001	.403–.773				
TF-FSs	.303	.097	.003	.107–.498				
FS test	.136	.079	.090	-.022–.295				

Model 4					.742	.003	.460	*F* (4, 46) = 36.912, *p*<.001
FSs used in speech samples	.575	.094	<.001	.386–.764				
TF-FSs	.286	.100	.006	.085–.488				
FS test	.088	.102	.391	-.117–.294				
GP test	.083	.111	.460	-.141–.307				

*Significant at *p* = .12 after Bonferroni adjustment.

Table 3. Hierarchical regression analysis of predictors of fluency (Rate A) of entire speech in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig
Model 1					.517	.527	<.001	*F* (1, 49) = 54.555, *p*<.001
FSs used in speech samples	.726	.098	<.001	.528–.923				
Model 2					.567	.058	.013	*F* (2, 48) = 33.767, *p*<.001
FSs used in speech samples	.532	.119	<.001	.292–.772				
TF-FSs	.309	.119	.013	.068–.549				
Model 3					.596	.035	.042	*F* (3, 47) = 25.554, *p*<.001
FSs used in speech samples	.526	.116	<.001	.293–.758				
TF-FSs	.226	.122	.071	-.020–.471				

FS test	.207	.099	.042	.008–.407				
Model 4					.591	.004	.484	*F* (4, 46) = 19.085, *p*<.001
FSs used in speech samples	.510	.118	<.001	.272–.748				
TF-FSs	.206	.126	.108	-.047–.460				
FS test	.150	.128	.248	-.108–.409				
GP test	.099	.140	.484	-.183–.381				

*Significant at p = .12 after Bonferroni adjustment.

Table 4. Hierarchical regression Analysis of predictors of fluency (Rate B) of entire speech in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* change	Model sig
Model 1					.509	.519	<.001	*F* (1,49) = 52.806, *p*<.001
FSs used in speech samples	.720	.099	<.001	.521–.919				
Model 2					.563	.062	.010	*F* (2, 48) = 33.251, *p*<.001
FSs used in speech samples	.519	.120	<.001	.278–.761				
TF-FSs	.320	.120	.010	.079–.561				
Model 3					.576	.020	.128	*F* (3, 47) = 23.618, *p*<.001
FSs used in speech samples	.514	.118	<.001	.276–.753				
TF-FSs	.257	.125	.045	.006–.509				

FS test	.157	.101	.128	-.047–.362				
Model 4					.572	.005	.433	*F* (4, 46) = 17.728, *p*<.001
FSs used in speech samples	.496	.121	<.001	.253–.740				
TF-FSs	.235	.129	.075	-.024–.494				
FS test	.092	.131	.488	-.173–.356				
GP test	.113	.143	.433	-.175–.402				

*Significant at *p* = .12 after Bonferroni adjustment.

Appendix 24: Hierarchical regression analysis of CAF of FSs in the opposite direction

Table 1. Hierarchical regression analysis of predictors of C-FS conducted in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* Change	Model sig.
Model 1					.435	.446	<.001	*F* (1, 49) = 39.523, *p*<.001
TF-FSs	.668	.106	<.001	.455–.882				
Model 2					.425	.002	.703	*F* (2, 48) = 19.491, *p*<.001
TF-FSs	.649	.118	<.001	.412–.887				
FS test	.045	.118	.703	-.192–.283				
Model 3					.434	.019	.197	*F* (3, 47) = 13.757, *p*<.001
TF-FSs	.587	.127	<.001	.332–.841				
FS test	-.078	.150	.606	-.381–.225				
GP test	.212	.162	.197	-.114–.538				
Model 4					.483	.057	.024	*F* (4, 46) = 12.666, *p*<.001
TF-FSs	.516	.125	<.001	.265–.767				
FS test	-.125	.145	.392	-.418–.167				
GP test	.316	.161	.056	-.008–.641				
Age of using English via online games	-.250	.000	.024	-.001–.000				

*Significant at *p* = .012 after Bonferroni adjustment.

Table 2. Hierarchical regression analysis of predictors of A-FS conducted in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* Change	Model sig.
Model 1					.408	.419	<.001	*F* (1, 49) = 35.409, *p*<.001
TF-FSs	.648	.109	<.001	.429–.866				
Model 2					.398	.002	.665	*F* (2, 48) = 17.507, *p*<.001
TF-FSs	.626	.121	<.001	.383–.869				
FS test	.053	.121	.665	-190.–.296				
Model 3					.407	.021	.192	*F* (3, 47) = 12.436, *p*<.001
TF-FSs	.561	.130	<.001	.300–.821				
FS test	-.075	.154	.629	-.385–.235				
GP test	.219	.166	.192	-.114–.553				

*Significant at *p* = .016 after Bonferroni adjustment.

Table 3. Hierarchical regression analysis of predictors of F-FS conducted in the opposite direction

	β	*SE*	*p*	95% CI	R^2	R^2 change	Sig. *F* Change	Model sig.
Model 1					.389	.401	<.001	*F* (1, 49) = 32.835, *p*<.001
TF-FSs	.633	.111	<.001	.411–.856				
Model 2					.376	.000	.911	*F* (2, 48) = 16.093, *p*<.001
TF-FSs	.628	.123	<.001	.380–.875				
FS test	.014	.123	.911	-.233–.261				
Model 3					.381	.017	.250	*F* (3, 47) = 11.262, *p*<.001
TF-FSs	.569	.132	<.001	.303–.836				
FS test	-.101	.157	.523	-.417–.215				
GP test	.198	.169	.250	-.143–.539				

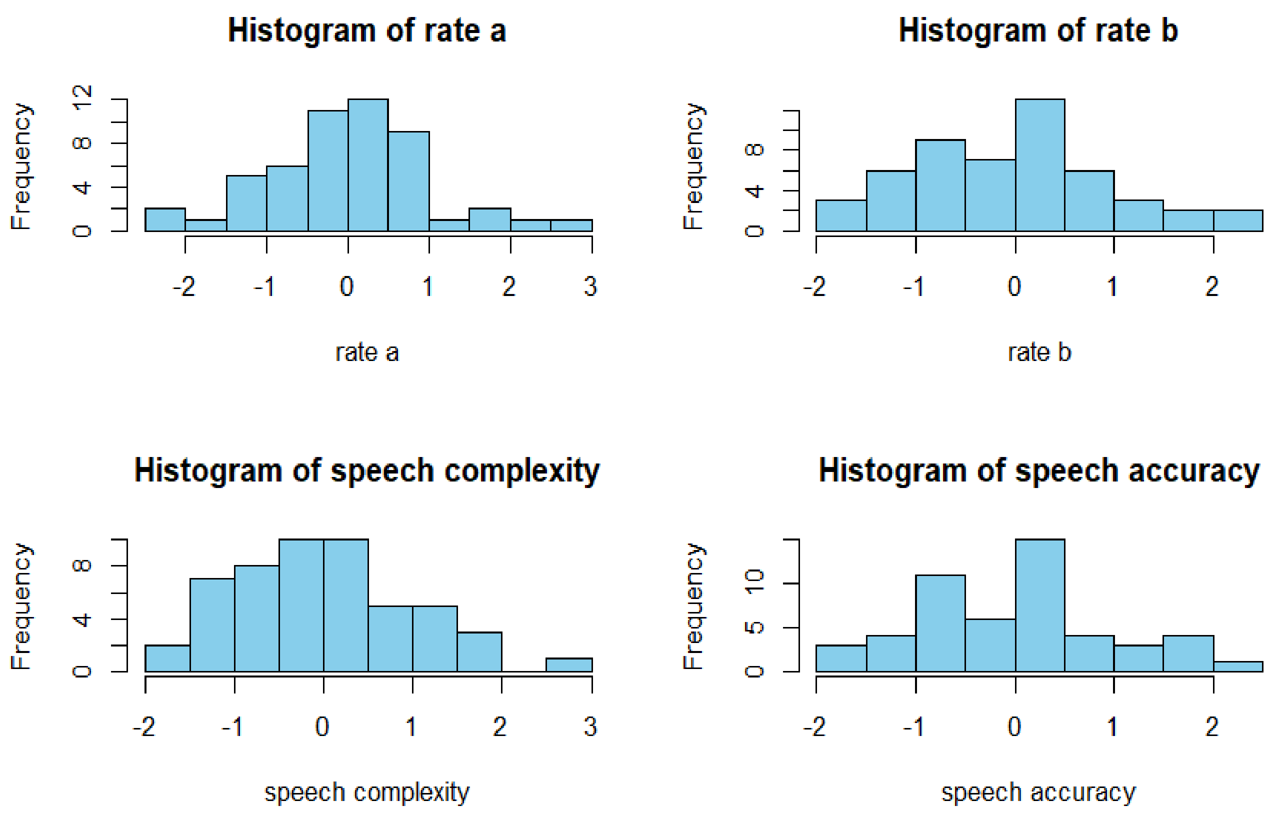
*Significant at *p* = .016 after Bonferroni adjustment.

Appendix 25: Tests of Normality – CAF of Speech

Table 1. Shapiro-Wilk Test for Normality: CAF of Speech

	Statistics	df	Sig.
Fluency (Rate A)	0.986	51	0.804
Fluency (Rate B)	0.977	51	0.417
Complexity of speech	0.973	51	0.288
Accuracy of speech	0.970	51	0.219

Figure 1: Normality Histogram of CAF of Speech



Appendix 26: Collinearity Statistics of dependent variables

Table 1. Collinearity statistics of complexity of speech and predictors

Collinearity Statistics/ Dependent Variable: Complexity of Speech		
	Tolerance	VIF
General Prof test	0.410	2.438
FSs test	0.448	2.229
Test of familiarity of FSs	0.447	2.235
WM test	0.779	1.282
Number of FSs elicited from speech samples	0.429	2.326

Table 2. Collinearity statistics of accuracy of speech and predictors

Collinearity Statistics/ Dependent Variable: Accuracy of Speech		
	Tolerance	VIF
General Prof test	0.410	2.434
FSs test	0.487	2.050
Test of familiarity of FSs	0.438	2.281
Number of FSs elicited from speech samples	0.322	3.104

Table 3. Collinearity statistics of Rate A and predictors

Collinearity Statistics/ Dependent Variable: Rate A		
	Tolerance	VIF
General Prof test	0.411	2.431
FSs test	0.481	2.077
Test of familiarity of FSs	0.487	2.049
Number of FSs elicited from speech samples	0.416	2.403

Table 4. Collinearity statistics of Rate B and predictors

Collinearity Statistics/ Dependent Variable: Rate B		
	Tolerance	VIF
General Prof test	0.410	2.438
FSs test	0.490	2.038
Test of familiarity of FSs	0.481	2.076
Number of FSs elicited from speech samples	0.427	2.337

Table 5. Collinearity statistics of complexity of FSs and predictors

Collinearity Statistics/ Dependent Variable: Complexity of FSs		
	Tolerance	VIF
General Prof test	0.367	2.723
FSs test	0.482	2.070
Test of familiarity of FSs	0.484	2.062
Age of using L2 via online games	0.807	1.237

Table 6. Collinearity statistics of accuracy of FSs and predictors

Collinearity Statistics/ Dependent Variable: Accuracy of FSs		
	Tolerance	VIF
General Prof test	0.415	2.406
FSs test	0.498	2.007
Test of familiarity of FSs	0.505	1.979

Table 7. Collinearity statistics of fluency of FSs and predictors

Collinearity Statistics/ Dependent Variable: Fluency of FSs		
	Tolerance	VIF
General Prof test	0.418	2.387
FSs test	0.496	2.015
Test of familiarity of FSs	0.507	1.972

Table 8. Collinearity Statistics of total number of FSs used by participants and predictors

Collinearity Statistics/ Dependent Variable: Total number of FSs used by participants		
	Tolerance	VIF
Complexity of speech	0.287	3.480
Accuracy of speech	0.170	5.853
Fluency of speech (Rate B)	0.276	3.594
General prof test	0.410	2.440
FSs test	0.474	2.110
Test of familiarity of FSs	0.430	2.325

Appendix 27: Durbin-Watson statistics

Table 1. Durbin-Watson test between C-SA and predictors

Durbin-Watson	
General Prof test	1.962
FSs test	
Test of familiarity of FSs	
WM test	
Number of FSs elicited from speech samples	
Dependent Variable: C-SA	

Table 2. *Durbin-Watson test between A-SA and predictors*

Durbin-Watson	
General Prof test	2.229
FSs test	
Test of familiarity of FSs	
WM test	
Number of FSs elicited from speech samples	
Dependent Variable: A-SA	

Table 3. *Durbin-Watson test between Rate A and predictors*

Durbin-Watson	
General Prof test	1.993
FSs test	
Test of familiarity of FSs	
Number of FSs elicited from speech samples	
Dependent Variable: Rate A (F-SA)	

Table 4. Durbin-Watson test between Rate B and predictors

Durbin-Watson	
General Prof test	2.114
FSs test	
Test of familiarity of FSs	
Number of FSs elicited from speech samples	
Dependent Variable: Rate B (F-SA)	

Table 5. Durbin-Watson test between C-FS and predictors

Durbin-Watson	
General Prof test	2.499
FSs test	
Test of familiarity of FSs	
Age of using L2 via online games	
Dependent Variable: C-FS	

Table 6. Durbin-Watson test between A-FS and predictors

Durbin-Watson	
General Prof test	2.534
FSs test	
Test of familiarity of FSs	
Dependent Variable: A-FS	

Table 7. Durbin-Watson test between F-FS and predictors

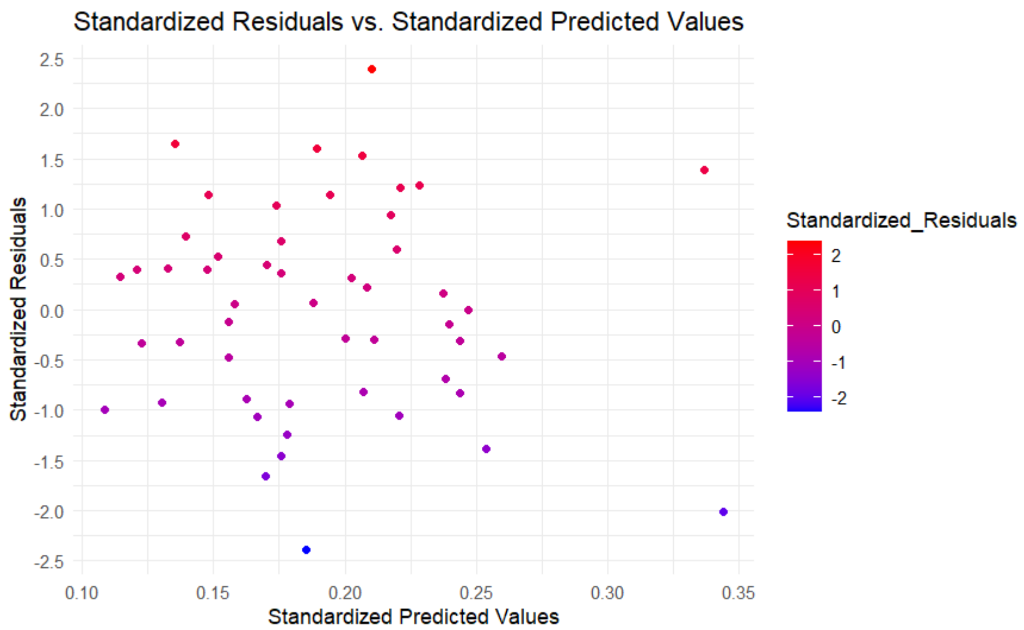
Durbin-Watson	
General Prof test	2.491
FSs test	
Test of familiarity of FSs	
Dependent Variable: F-FS	

Table 8. Durbin-Watson test between number of FSs elicited from speech samples and predictors

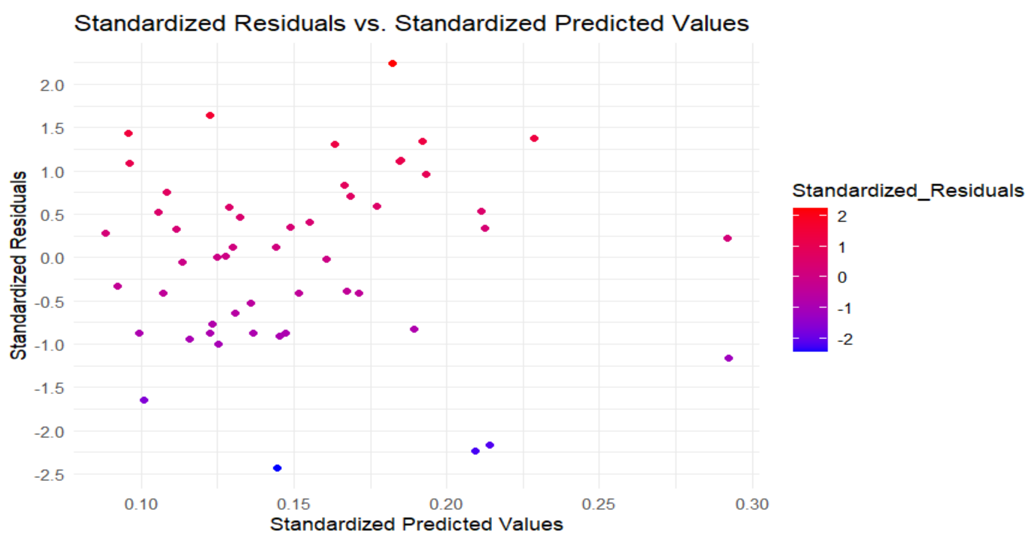
Durbin-Watson	
Complexity of Speech	2.337
Accuracy of Speech	
Fluency of Speech (Rate B)	
General Prof test	
FSs test	
Test of familiarity of FSs	
Dependent Variable: Number of FSs elicited from speech samples	

Appendix 28: Scatterplots of standardised residuals versus standardised predicted values

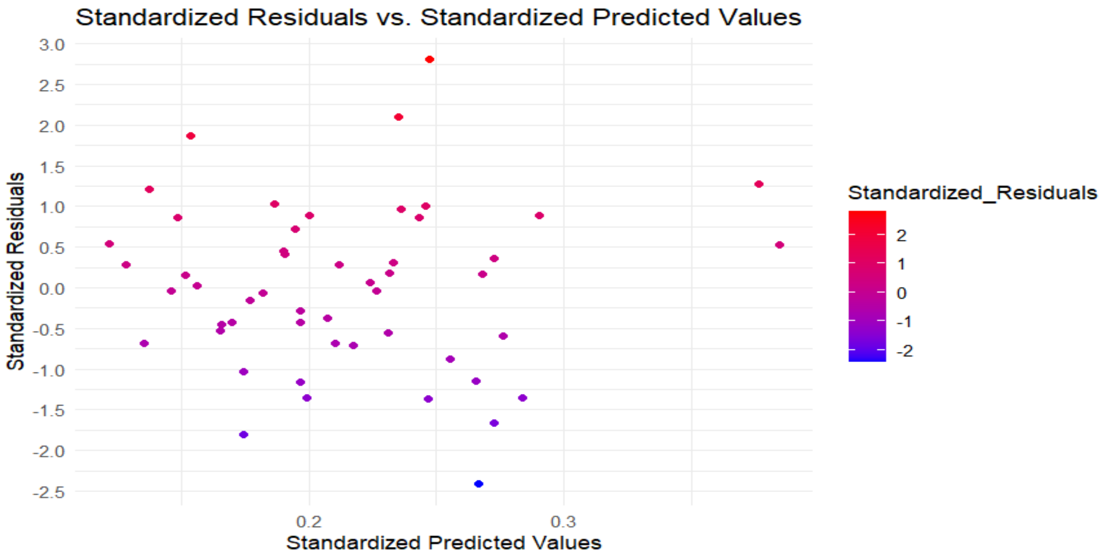
Scatterplot graph 1. C-SA standardised residuals versus standardised predicted values



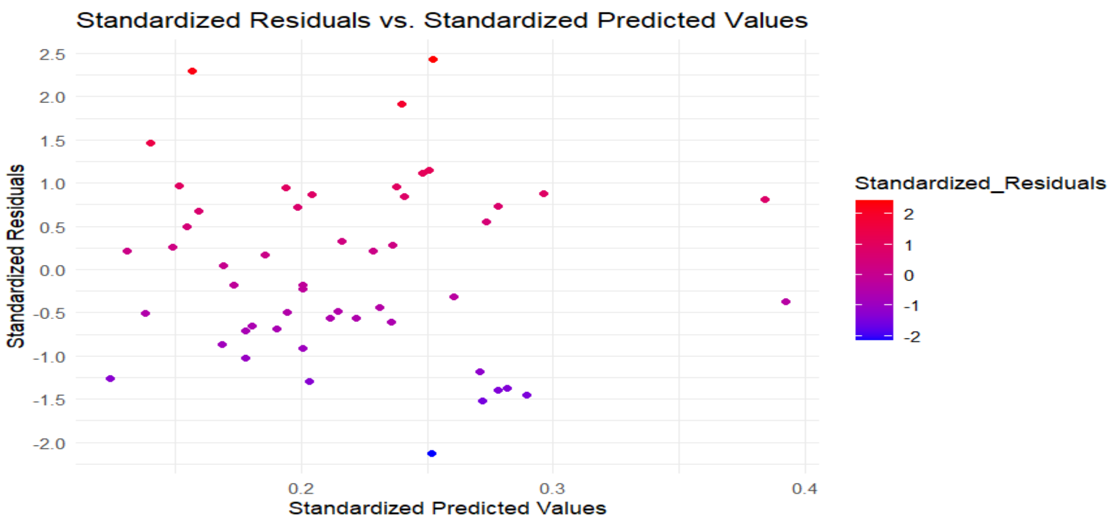
Scatterplot graph 2. A-SA standardised residuals versus standardised predicted values



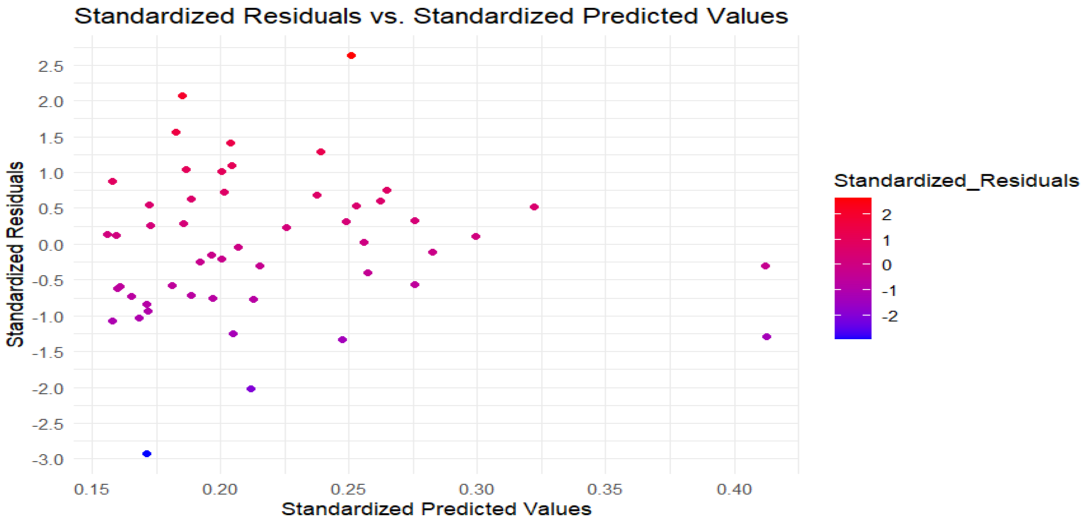
Scatterplot graph 3. Rate A standardised residuals versus standardised predicted values



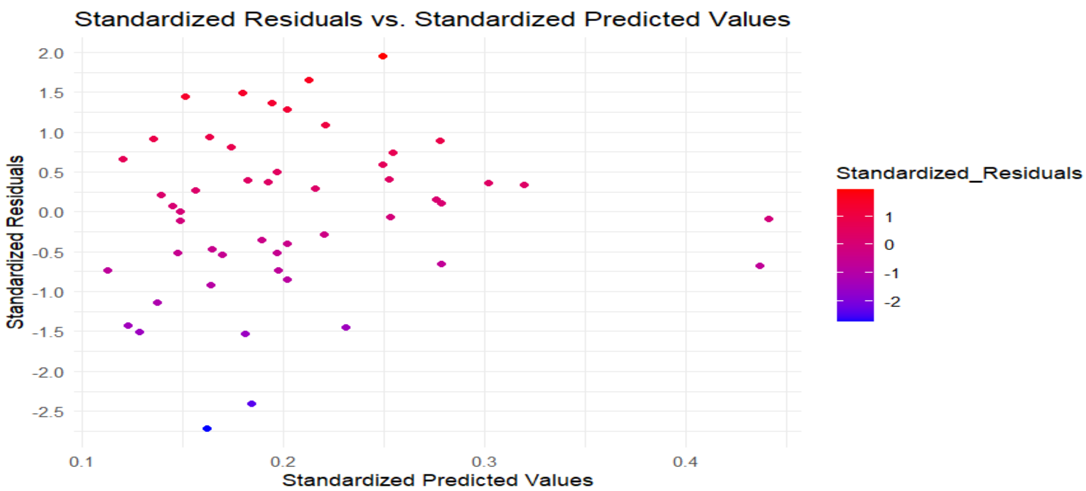
Scatterplot graph 4. Rate B standardised residuals versus standardised predicted values



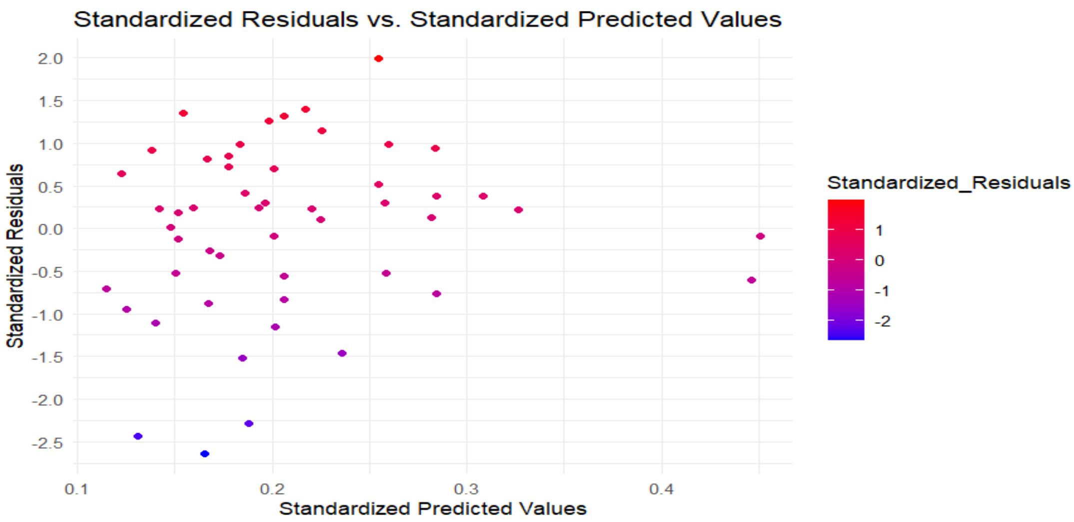
Scatterplot graph 5. C-FS standardised residuals versus standardised predicted values



Scatterplot graph 6. A-FS standardised residuals versus standardised predicted values



Scatterplot graph 7. F-FS standardised residuals versus standardised predicted values



Appendix 29: Examples of participants' speech samples

Table 1. Examples of participants' speech samples elicited from three different proficiency levels.

Proficiency level A1/ **Participant 14**	**Proficiency level A2/** **Participant 34**	**Proficiency level B1/** **Participant 25**
Ahh **My name is** X ahh my ahh I am ahh I ah I I live in X city ahh my hobby is **fashion designer** ahh ahh **my birthday** is ahh ahh twenty four ahh March mmm **I am** twenty one mmm I I am stud studying at the business studies my maj major in matreles matreles management mmm *speaking in Arabic* ahh in my **spare time** mm I like to watch ahh serious <u>(she means series)</u> ahh and **read books** ahh my personality is calm and I love ahh cooperation and I like to give ahh what I have to my work and avoid problems ummm ahhh *asking for confirmation*	Ahh Hello **my name is** X I'm twenty **years old** I study for three years in college I I work I work ahh for for one year in bank I have experience with this ahh you **can count on** me and ahh this is my number for if you like to PAUSE ahh this is my number this is **my personal** number and this is my email ahh and thank you for this opportunity ahh and thank you *Khalast I encouraged her to speak more by giving ideas She was hesitant* hello my name is X I'm twenty years old I work I study three years in a college **my experience** with ahh it was	Ahh **my name is** X I just become a professor in I have a I had a job meeting at the college of business and I have the job I got it and I'm the like I'm **going to** teach English there umm so ahh my **email address is** like X@hotmail.com my number is like five triple six five two one four and and I **live near** the college so **it's easy for me** to go and I'm not gonna be late for like a meeting or a lecture I have there and ahh before become at like a professor there I used to work at a a what's it called yeah I used to work at a hospital for the HR and I got **a lot of** experience from that job like ah I get to know **how to manage** like if there is a problem I can manage it ahh if there a complaint or something I know **how to handle** people so that's why I I I become a professor to teach like student that's like for my other job they got me to know how to act with people and customers and everything so I know that I know how to handle like student cause they're people and ahh so I got to know them and ahh I like teaching as it **make me feel like**

ahh ahh ahh ahh greeting to you and I hope you acc accept *(she said I finished and I encouraged her to continue)* ahh accept accept for the job and **thank you** *I encouraged her to continue*	so difficult in the college I work for one year in a bank and PAUSE I would like to open my shop ahhh PAUSE I **would like to** open my store and designing clothes and and **this is my** number and this is my personal email and thank you for this opportunity ahh and I would like to **hear from you**	giving knowledge to people and I'm passing the knowledge that I used to know and ahh I just really like teaching it just make me like happy when they know a new stuff or a new thing umm umm but I'm I'm Im'm not just like this not the end I'm not just looking to be a professor at the college ahh like I wana **learn more** so I can give more and the more I get the more I give and yeah **I think** that **will be** all so thank you for having me **thank you for** listening to me and thank you **for giving me** the **opportunity to talk about** me and about my business and about my teaching and ahh job experience I think that will be all

- Writing in Bold: indicate FSs uttered by participants during speech samples
- Writing in Italics: indicates L1 use during speaking task