Examining the relationship between English oral fluency and working memory capacity in dialogic and monologic performances

Nada Alsheehri

Thesis submitted for the degree of Doctor of Philosophy

The University of Leeds

School of Education

July 2021

Declaration

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others

List of Abbreviations

ANOVA	Analysis of Variance
AR	Articulation Rate
BDS	Backward digit span
CEFR	Common European Framework of Reference
EFL	English as a Foreign Language
ESL	English as a Second Language
FPs	Filled Pauses
L1	First language
L2	Second language
LTM	Long-Term Memory
OST	Operation Span Test
OPI	Oral Proficiency Interview
PSTM	Phonological short-term memory
PTR	Phonation time ratio
RST	Reading Span Test
SST	Speaking Span test
STM	Short-Term Memory
SLA	Second Language Acquisition
SP	Speech Rate
SPs	Silent Pauses
WMC	Working Memory Capacity

Acknowledgements

I would like to show my deepest gratitude and appreciation to many individuals who helped to shape my work into different ways. First and foremost, I would like to thank my supervisors, Professor Deignan and Dr Ahmadian, as without their support, guidance and expertise, I would not have been able to reach this milestone. They have been a very strong and patient team. Their unwavering instructions have helped me at every moment of uncertainty, from my research questions until I was able to stand on my own feet and carry on with the rest of the thesis. My supervisors have been a part of my PhD journey, so I could not ask for more. I thank Professor Deignan for being a challenging and knowledgeable supervisor during my PhD journey. Her comments and feedback on my work have taught me how to be a confident researcher. Dr Ahmadian's advice paved my way through every chapter until I dotted the last *i*, by providing detailed feedback and suggestions to improve the work. I thank Dr. Matt Homer for his statistical advice, which helped me to think about the analysis of my data. He never hesitated to answer my questions with great detail. I am also thankful to Louise Greaves, who works as a doctoral college administrator and student support officer. She is always there for help and guidance for any inquiry that a PhD student has. I dedicate this thesis to my parents, who have waited years to receive an answer to their constant question: "When will you be back?" So I would like to express my heartfelt thanks to my mom and dad for the huge support they gave me during my academic journey. Whenever I go back home to visit for a holiday, they welcome me with smiles, love and appreciation. I thank my family, sisters and brothers for their blessings and prayers and for always being proud of my achievement. I thank forever and ever my lovely sister Nouf for her unwavering support during my postpartum period. I'm so grateful for every single day she spent taking care of me and my baby.

I thank my loving husband for his support and believing in me during this four-year journey. I am so grateful to my son, Meshal, who has lived an expatriate life through no choice of his own. With Meshal I realized the meaning of unconditional love. He is the best thing that happened to me during the difficult time of 2020–2021. I learned how to be more patient and organized than before. I am thankful for my friends who are always around me in good and difficult times. Finally, special thanks go to my sponsor the University of Jeddah, represented by the Saudi Arabian Cultural Mission (SACM), who contributed to my success by supporting me financially in the United Kingdom.

Abstract

Apart from a few exceptions (e.g., Witton-Davies, 2014; Tavakoli, 2016; Peltonen, 2017) that have addressed second language (L2) fluency in a dialogic task, the majority of L2 fluency studies have looked at oral fluency in a monologic task. Dialogue is the more authentic and natural mode of communication, which is reflected in everyday language use. Presently, there is a paucity of research examining dialogue fluency in non-native bilingual interlocutors who share the same L2. Research studies on L2 speech fluency recommend studying fluency in the dialogic context, considering the social aspect of fluency. Responding to the researchers' calls, this study introduced monologic and dialogic tasks to investigate the different aspects of speed, breakdown, and repair fluency in the oral performances of 64 L2 learners of English. The available literature on language learning and processing has highlighted meaningful relationships between individual differences (IDs) in working memory capacity (WMC) and L2 speech production models of first language (L1) as in Levelt (1989) and L2 as in de Bot (1992). However, it is still unknown whether L2 dysfluency in monologue and dialogue is associated with IDs in WMC. Thus, this study aimed also to address this gap in the literature by examining the relationship between utterance fluency in monologue and dialogue, and WMC. 64 undergraduate Saudi students were instructed to do several tasks. A monologic argumentative task and a dialogic discussion task. In a dialogic discussion task, 32 pairs were asked to exchange opinions about a popular topic in their country. They also completed two complex WM tests: Operation Span Test and Backward Digit Span Test. The results were in line with previous research findings showing that L2 participants were more fluent in dialogue than monologue in utterance fluency in terms of speed and breakdown measures. WMC was not a strong predictor for the variations in L2 oral performances between monologue and dialogue. With all the above in mind, it can be concluded that the current study has implications for teaching, testing and literature.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 The Significance of the Study

As a lecturer who works closely with L2 learners in the English Department at the University of Jeddah, a state university in Saudi Arabia, I have noticed that my students' oral fluency in English varies, although they have had the same amount of previous L2 learning and experience (seven years). L2 fluency is one of the L2 proficiency components, along with complexity and accuracy, and it is also referred to as the easiness and smoothness of speech delivery. L2 learners generally have a goal to be fluent in L2, and I have frequently been asked as a teacher "How can I improve my speaking abilities before I travel abroad?" by students wishing to complete their higher education, take summer courses or take IELTS or TOEFL exams. My L2 learners seem to be less confident in speaking than writing. They correct themselves before completing their sentences and they use several repetitions and pauses to produce fully correct sentences. It seems that these dysfluencies are related to hesitations, reformulations and self-corrections when students try to produce complex sentences or to narrate a story or talk about a personal experience. They usually build knowledge about the target language by learning its grammar and vocabulary. Thus, they seem to care more about the grammar and structure of the sentence than the smoothness and flow of the speech delivery.

Today, the majority of EFL students in Middle Eastern countries are still using rote learning, which depends heavily on memorization (Smith and Abouammoh, 2013). It has been argued that the language learning materials and tasks rarely satisfy learners' needs. This possibly leads to many unmotivated leaners who are reluctant to engage

in the pedagogic language tasks inside the classrooms. Their lack of motivation is not helped by the fact that those tasks are not related to real-life situations (Awwad, 2017). Sometimes the lack of exposure to the target language could prevent oral proficiency development, meaning that learners have little or no access to L2 outside the classrooms. As a result, they lack speaking skills and therefore use long pauses, hesitations, and repetitions. Furthermore, teacher-centred instruction, large class size, lack of care about learners' individual differences (IDs), and some traditional/out-dated teaching methods (e.g., grammar translation) are among the main challenges faced by many EFL learners. This type of learning context is called a 'minimal input' condition (Larson-Hall, 2008, p. 36).

Additionally, teaching and practising English speaking in foreign language classroom still follow a non-communicative style, meaning that the speaking is taught and practised by reproducing memorized sentences, repeating texts, and reading aloud (Witton-Davies, 2014). In foreign language classrooms in Saudi Arabia the emphasis remains on written English rather than on spoken English. This leads to lack of oral communication abilities in English outside the classroom because L2 is used in decontextualized mode and learned from written script. Moreover, speaking classes are introduced only in the first year of university education. This year is called the Preparatory Year Program. In this year, English language skills (e.g., reading, writing, speaking and listening) are taught to a large number of L2 students. After the first year, the emphasize shifts to teach English literature, phonology, linguistics, and grammar. Thus, students' speaking abilities are limited, especially in speaking English spontaneously. In contrast, students have considerable knowledge about the grammar, reading, and vocabulary of the language.

Practising L2 speaking in an interactional context is neglected in EFL and ESL (English as a Second Language) classrooms, where dialogue is the natural form of communication (Garrod and Pickering, 2004). Speaking practice should be encouraged in the classroom as students are rarely engaged in monologic settings. Thus, the current study suggests that L2 learners' oral fluency can be improved by engaging in L2 oral interactive activities in and outside the classroom. These oral activities could help to improve conceptualization and formulation processes as measured by pause location and frequency. Researching L2 utterance fluency in the performance of monologic and dialogic tasks is of immense importance because it provides English language instructors and syllabus designers with information about tasks and activities that can be implemented in L2 classrooms to improve L2 learners' oral fluency. An example would be interactive tasks in pairs.

Despite the importance of studying L2 fluency in an interactive task, there remains a paucity of evidence on the differences between monologic and dialogic L2 performances in terms of speed, breakdown, and repair. The findings from previous studies (e.g., Michel, 2011; Witton-Davies, 2014; Tavakoli, 2016; Peltonen, 2017) regarding the differences between monologue and dialogue in terms of breakdown and repair measures are inconsistent and further studies are needed. Other studies (e.g., McCarthy, 2010; Peltonen 2017; Os et al., 2020) examined dialogue fluency alone using some temporal measures, communicative problem-solving tasks, turn taking, and turn pauses with no compariosn with fluency in monologic tasks. In fact, the results of the previous studies are influenced by many variables: measuring fluency with complexity and accuracy components (e.g., Gilabert et al., 2011; Michel, 2011); task types (picture description) (e.g., Derwing et al, 2004); a small number of

participants and type of research (longitudinal) (e.g., Witton-Davies, 2014). Furthermore, the task types and topics that have been used to elicit speech samples do not represent real life communication in terms of everyday language use. For example, picture description tasks were used instead of personal narrative tasks (e.g., Witton-Davies, 2014). According to Kahng (2014), when a speaker shares a true experience with a partner in a conversation, it facilitates L2 spontaneous speech with fewer dysfluencies.

Additionally, previous literature on L2 fluency studies (e.g., Mizera, 2006; Kormos and Trebits, 2011; Michel, 2011; Sato, 2014) did not measure fluency alone as the main object of the study; rather, fluency was measured with other aspects of L2 oral performance, i.e., complexity and accuracy. Such studies are referred to as studies in task-based language teaching. The focus is usually on the effect of different task complexities and conditions on L2 learners' complexity, accuracy, and fluency (CAF) of oral performance. However, study fluency on its own, using a wide range of fluency measures and providing detailed explanation and justification of the choice of measures and decisions in analysis, would be beneficial for other SLA researchers.

As little is known about fluency in dialogue, it is not clear which factors influence L2 performances in monologue and dialogue. Another significance of this study is concerned with the extent to which L2 speed, breakdown, and repair are dependent on working memory capacity (WMC). Although extensive research has been carried out on WM and L2 oral proficiency CAF (e.g., Fortkamp, 2000; Mota, 2003; Guara-Tavares, 2008; Kormos and Safar, 2008; Kormos and Terbits, 2011), no previous study has investigated the extent to which individual differences (IDs) in WMC can explain the variations in L2 fluency performance as a single construct in both task

modes (monologue and dialogue) using two complex WMC tests. The effectiveness of those tasks is examined to ensure their validity when implemented in L2 classrooms. This study also aims to encourage L2 teachers to consider learners'IDs in WMC, which regulates the process of retrieving lexical knowledge to formulate and articulate L2 speech. In fact, learners' attentional resources play an important role in language acquisition, processing and performance (Robinson, 2011). To the best of my knowledge, the relationship between L2 utterance fluency in monologue and dialogue and WM has not been explored before in the SLA context, and this is therefore the research gap that this study aimes to fill. To this end, there is still a need for more studies to look at the broader aspect of L2 utterance fluency which could provide significant information to SLA researchers about the relationship between WMC and all aspects of L2 utterance fluency.

1.2 Organization of the Thesis

This section introduces an outline for the current thesis's structure and content. This thesis has six chapters as follows: introduction, literature review, methodology, results, and discussion and conclusion. Chapter 1 has already introduced the theoretical and practical significance of the study. Theoretical significance lies in the importance of the topic L2 oral fluency in the literature on speech production processing, whereas the practical significance lies in the importance of L2 oral fluency in teaching, learning, and testing.

Chapter 2 introduces the literature review and the most common empirical studies on L2 fluency. It also provides background information that is essential for the reader to understand the purpose of the current study. First, the models of first (Levelt, 1989) and second (de Bot, 1990) language speech production processing and how these models are related to oral fluency are discussed, as are studies that have been carried out into different aspects of fluency. Next, the most common definitions of L2 fluency and the measures of utterance fluency are reviewed. Furthermore, the gaps in this literature are identified. The most important factors that influence the development of oral fluency are reviewed and the differences between monologic and dialogic performances are discussed. This is followed by an evaluation of fluency in dialogue and monologue oral tasks. Then the topic of working memory capacity (WMC) is introduced and discussed in relation to language processing and performance, followed by a discussion of WMC in monologue and dialogue. Measures of WMC are reviewed in detail, followed by empirical studies that have been carried out on WMC

and L2 fluency. This chapter concludes by detailing the research gap and the original contribution of the current study.

Chapter 3 presents the aim of the study, the study design, the research questions, methods, participants, tasks, tests, and pilot study. The procedures undertaken to fulfil ethical requirements are also described. Detailed descriptions of the data collection procedures, and the approaches taken to data coding and data analysis are presented in this chapter. The procedure of checking inter-rater reliability is also discussed in this chapter.

Chapter 4 introduces the results of the research questions in relation to the hypotheses that were presented in the methodology chapter. Statistical analyses of utterance fluency scores are presented in detail, starting with the descriptive statistics of each utterance fluency aspect (speech, breakdown, and repair) in monologue and dialogue. The results obtained by the non-parametric tests of the Wilcoxon signed rank test and Friedman's ANOVA are introduced to show the differences in the participants' performances in monologic and dialogic tasks. Regression analysis is carried out to examine the predictive power of the WM tests Backward Digit Span Test and Operation Span Test in predicting oral fluency in monologic and dialogic performances.

Chapter 5 presents a detailed discussion of the current study findings. It begins with a summary of the key findings, followed by a discussion of the results from the data analysis related to the four research questions. In this chapter the findings are linked and interpreted in relation to extant L2 fluency studies on monologue, dialogue, and WMC.

Chapter 6 has three sections: the first section starts with a summary of the conclusion of the study findings. The second section discusses the implications of this study for learning, language, and testing, and the third section discusses the limitations of the current study and introduces suggestions for further research.

Chapter 2: Theoretical and Empirical Background of Speech Fluency

2.1 Introduction

As a starting point, the following sections present theoretical frameworks that influence this study are introduced and discussed, namely speech production models of first and second language. Then, detailed definitions, background information and the empirical studies for fluency and its use in both modes: dialogue and monologue are also introduced. Next, working memory (WM) section introduces WM model which has been used in previous SLA studies. WM represented by operation span test and backward digit span are the independent variables, whereas speed, breakdown and repair are the dependent variables of the current study. Studies on individual differences of WM and second language performance are also introduced to investigate the relationship between WM and L2 fluency. Finally, this chapter concludes with the research gap.

2.2 Speech Production Models of First and Second Languages

2.2.1 Levelt's Blueprint of the L1 Speaker

Speech production theories, such as the L1 model by Levelt (1989) and the adapted versions for L2 by de Bot (1992) and Kormos (2006), are used as psycholinguistic and theoretical frameworks for fluency studies. Levelt (1989) proposes that speech production involves three incremental stages: conceptualization, formulation, and articulation. In conceptualization, all the intended ideas or information have been generated in the form of a preverbal message to be outputted to a language later in the formulation stage. The preverbal message is non-linguistic message and a part of conceptual planning. The generated message is saved in the WM, which contains all

the information that is currently accessible to the speaker (Levelt, 1989). According to Levelt (1989), conscious attention and WM are more important in the conceptualization and monitoring stages than in the formulation and articulation stages because message generating and monitoring are controlled conscious processes that required the speaker's awareness (WMC). On the contrary, formulation and articulation are considered to be automatic processes in native speakers and they are not controlled by central executive memory. Temple (1997) argues that WM in the L2 speech production model is less automated in the formulation stage, meaning that automatic lexical retrieval is not always possible as in native speech – it is partially automatic. Second language speakers depend on the explicit knowledge and L2 grammar that are stored in the short-term WM and this results in the slow serial formulation of messages.

Moving to the second phase of L1 speech production processing, formulation where the preverbal message is activated and ready to receive encoded grammatical and phonological structures. Grammatical encoding is the lexical retrieval of the words from the mental lexicon. Once the lexical features are put in the right order, the phonological encoding prepares the pre-selected lexical items for the articulation stage. At this stage, the output, speakers carry out explicit speech with a high degree of automation. Respectively, the WM stores all of the generated messages, to ensure that the three stages will be available for future processing (Levelt, 1989). Finally, monitoring can take place in all three stages to ensure appropriate and accurate production. Levelt (1989) explains that these three stages of speech production are incremental; however, the processing proceeds in parallel, meaning that when the speech is being articulated, the next segments will be in the planning stage and other segments will be at the conceptualization phase.

2.2.2 Adaptations of Levelt's Model for L2 speech

The modular model of L1 speech production proposed by Levelt (1989) is one way to explain the variations in L2 speech production. Also, due to the increasing interest in the investigation of the psycholinguistic aspects of L2 speech, de Bot (1992) and Kormos (2006) have adapted and modified a bilingual version of Levelt's original model. de Bot (1992) argues that in the conceptualization stage of the preverbal message, bilingual speakers firstly decide which language to use. He argues that:

> In the conceptualizer communicative intentions are given form in the preverbal message, which contains information about the language in which (part of) an utterance is to be produced. Through this information, the relevant language-specific formulator is activated. In formulator, the preverbal message is converted into a speech plan (de Bot, 1992, p.21).

Thus, all L2 researchers (e.g., de Bot, 1992; Tavakoli and Skehan, 2005; Kormos, 2006) agree that L2 speech production involves the same stages as L1 speech production. However, in the bilingual model, a separate formulator and a separate lexical item are required for each language in order to prepare and move the utterances to the third phase, articulation, which is also not language specific (de Bot, 1992). Alternatively, Kormos (2006) argues that L1 phonological and syntactic information are not stored separately, but that both are stored in long-term memory. She points out that the formulation and articulation stages of L1 run automatically and in parallel, unlike L2 speech processing, which demands attention in both the conceptualization and monitoring phases. This means that WM is responsible for the conscious attention paid to manipulating, and monitoring in L2 production. Sometimes, the controlled attention required in L2 acquisition and use imposes an extra load on WMC and it might negatively affect the speed or the quality of L2 language production

(Fortkamp, 2000). In addition, in the bilingual model, all phases are supervised by a monitor located at the conceptualizer to check the appropriateness of the content of the preverbal message and the accuracy of phonological and lexical production. The monitor can reform the output before the actual production or after the production (Kormos, 2006).

According to Kormos (2006), parallel processing in oral production means that a speaker can simultaneously conceptualize and encode the intended message. She argues that "these mechanisms are only partially automatic even in the case of advanced L2 learners" (p.26). Thus, the speaker has the ability to work on two stages at the same time because both lexical retrieval and syntactic encoding are largely automatic in L1 speakers and proficient L2 speakers. In contrast, in speech production, the stages of lexical retrieval and syntactic encoding for low-proficiency L2 speakers are not automatic (Kormos, 2006). Thus, these two stages could not be processed in parallel. In this case, low-proficiency speakers use serial processing for conceptualization, formulation, articulation, and finally monitoring (Kormos, 2006). Apparently, this results in speech with dysfluencies, such as filled/silent pauses, while the messages are being encoded.

In relation to language fluency, Towell et al. (1996), Skehan et al. (2016) and Lambert et al. (2017) argue that the three stages of L2 speech production are related to sub-dimensions of L2 utterance fluency, namely, speed, breakdown, and repair. In other words, the stages of L2 speech production processing are assessed via speed, breakdown, and repair fluency measures (Hanzawa, 2021). Lambert et al. (2017) explain that filed/unfilled pauses between clauses are related to the conceptualization stage. As for the formulation stage, filled/unfilled pauses within clauses are indicators

of syntactic encoding. Repair, reformation, or repetition reflect the monitoring of the speech production. As for speed fluency, it is claimed to evaluate the overall efficiency of the three speech production stages (Hanazwa, 2021).

Taken all together, the above models are widely accepted and well-known theoretical frameworks for L1 and L2 speech processing. Therefore, this study will adopt both the unilingual (Levelt, 1989) and bilingual (de Bot, 1992; Kormos, 2006) speech models since the sample of the study comprises bilingual speakers of L1 Arabic and L2 English. This study adopted this theoretical framework to examine how L2 learners use L2 fluency in monologic and dialogic task modes in terms of the objective measures of L2 fluency (speed, breakdown and repair). It is expected that the findings of this research will advance our understanding of how the stages of L2 speech production processing can be assessed by aspects of utterance fluency (e.g., speed, breakdown, and repair) in monologic and dialogic tasks, taking into consideration individual differences in working memory capacity.

2.3 Defining L2 Fluency

Fluency is to some extent a difficult term to define in the context of foreign language research because it has different meanings; it is not a straightforward measurement (Chambers, 1997; Lahmann et al., 2015). On the one hand, it is important to differentiate between fluency as a descriptor of overall general proficiency, and fluency as communicative competence. In communicative language teaching “the notion of fluency is used to assess how well learners use their knowledge to achieve their linguistic and communicative purpose” (Chambers, 1997, p.537). On the other hand, fluency as a general proficiency is usually used when people are commenting about a

person who speaks a language fluently, often referring to a foreign language speaker. For example, someone might say she is a fluent speaker of English (Chambers, 1997).

In a later study, Lennon (2000) described the narrow/broad senses of fluency as lower order and higher order fluency. Lennon (1990) explains the broad/higher order fluency in the English language as the overall speaking proficiency in English language testing and teaching. It is the ability to speak a language with the correct grammar, syntax, large vocabulary size, and nativelike accent. Thus, fluency is considered as one component of oral proficiency, along with accuracy and complexity.

Fillmore (1979) proposed four broad definitions for the concept of fluency that are used to make judgements about the oral fluency of others, and they are also based on temporal aspects of the language. First, fluency is speaking accurately with the ability to fill the whole time with speech and fewer pauses. Second, it is the ability to deliver the message with a solid semantic base without too many fillers such as, “you know,” “like,” and “I think.” Third, it is producing speech that fits different social contexts by using communicative skills. Fourth, speakers should be able to utilize the language to express feelings, such as sadness and happiness, fun, imagination, or creativity. Thus, when taken as a combination, the above definitions represent the everyday meaning of fluency that can be used by the average person.

Moving on to more technical definitions of fluency, we begin with the one proposed by Lennon (1990). Lennon defines the narrow/lower order meaning of fluency as the listener’s perception of the smoothness and the easiness of the speaker’s speech. She further defines fluency as one of the independent components of speech production, for example when someone speaks fluently and smoothly but without accurate grammar. Derwing et al. (2009) refer to fluency as a procedural automatic

skill in the speakers of the target language. In other words, fluency is speakers' ability to incorporate the fundamental cognitive process that is responsible for producing utterances in a suitable time and manner to maintain the fluidity of the speech. Similar to Lennon, Derwing and Munro (2015) define fluency as the smoothness and ease of L2 speech – speech that is free from pauses, hesitations and other dysfluency markers. Lennon (1990) and Derwing and Munro's (2015) definitions of fluency are akin to the concept that Segalowitz (2010) calls utterance fluency. Segalowitz (2010, 2016) defines fluency from a cognitive perspective in terms of three aspects: utterance fluency, perceived fluency, and cognitive fluency. These will be defined in turn in the sections that follow below.

2.3.1 Utterance fluency

In language testing, L2 fluency is frequently measured to infer which temporal measures affect the listeners' judgement of L2 speech (Kormos and Dénes, 2004). Researchers such as Skehan (2003) and Tavakoli and Skehan (2005) were the first to formulate the inclusive framework of fluency aspects. According to Skehan (2003), temporal measures are speed, breakdown, and repair. He argues that it is important to measure what causes disturbance to the flow and smoothness of the speech, for example pauses, repairs and hesitations. It is also important to measure what slows down the speed of the speech, for example the articulation rate, length of run, and speech rate. Thus, in the past 10–15 years these three aspects have become the standard for selecting oral fluency measures. For example, Skehan (2003) and Segalowitz (2010) adopted three speech components as a framework to define utterance fluency. Segalowitz (2010) suggests that utterance fluency consists of the temporal features of speech that can be acoustically measured, namely (1) speed,

such as syllable duration, articulation rate, and the length/number of syllables; (2) breakdown, such as silent pauses, the length of silent pauses, the number of filled/unfilled pauses, the length of filled pauses, mid/end silent pauses, and mid/end final pauses; and (3) repair, such as repetitions, reformations, hesitation, and false starts.

2.3.2 Perceived fluency

Perceived fluency is listeners' perceptions of speakers' speech. In L2 studies, perceived fluency is measured by human ratings of L2 speakers for utterance fluency and other objective aspects of overall oral fluency (Segalowitz, 2010). For example, Derwing et al. (2004) investigated how listeners' judgement of L2 participants' speech correlated with their oral performances as measured by speech rates, silent pauses, and the mean length of runs between two silent pauses. Data were collected from twenty Mandarin speakers of English as their L2 across three speech elicitation tasks: picture elicitation, monologue, and conversation. The results indicated that L2 speakers' fluency was lower in the picture description task than in the monologue and conversation tasks, whereas their fluency was similar in both the monologue and conversation tasks. Additionally, native speakers' judgement is a popular factor in fluency studies because it has been identified to correlate with some aspects of L2 utterance fluency, such as speech rate, within-clause pauses, and pauses at clause boundaries. Thus, perceived fluency captures what we are unable to see in utterance fluency measurements.

2.3.3 Cognitive fluency

Segalowitz (2010) defines L2 cognitive fluency as speakers' ability to carry out fundamental cognitive process responsible for producing utterances in a suitable time

and manner to maintain the fluidity of the speech. The cognitive process is explained as “the lexical search, for packaging the information into a grammatically appropriate form, for generating an articulatory script for speaking utterance” (Segalowitz, 2010, p.48). There are two aspects of cognitive fluency: attentional control and access fluidity. Attentional control has to do with a speaker’s ability to efficiently process speaking that is not limited by the capacity of short-term memory, whereas access fluidity is the speed and flexibility with which lexical and grammatical information are retrieved and accessed. Segalowitz (2010) also argues that sociolinguistic (social interaction), psychological (aptitude, age, working memory, experience), and psycholinguistic factors (speech perception, speech processing) contribute to developing the level of oral fluency.

Recently, Tavakoli and Hunter (2018) introduced four approaches to differentiate between the broad and narrow senses of fluency: (1) the very broad sense refers to general L2 proficiency (including speaking skills); (2) the broad sense of fluency refers to someone being able to speak fluently, communicate confidently, and take turns in the conversation; (3) the very narrow sense of fluency refers to three sub-components of utterance fluency, speed, breakdown, and repair; and (4) the narrow sense of fluency refers to fluency as one of the components of oral production, apart from accuracy and complexity.

The two perspectives of broad and very broad fluency have been used by raters to assess oral fluency, especially when they are asked to rate the speech in a broad sense only (e.g., Riggenbach, 1991; Kormos and Dénes, 2004). In the context of L2 fluency, the narrow sense is usually examined and considered as one component of

oral fluency, together with the temporal measures of fluency: speed, pausing, and repair (Lennon, 1990).

A range of fluency definitions has been surveyed above. Although there is still little agreement on what constitutes fluency, previous studies are similar in describing fluency as the ease and smoothness of speakers' speech. Furthermore, studies of L2 fluency, such as those by de Jong (2012), Kahng (2014), Tavakoli (2016), and Peltonen (2017, 2018) used Segalowitz's (2010) utterance fluency components as reliable measures for speech fluency across dialogue and monologic tasks, and L1 and L2 data. This study mainly focused on the measurable quality of oral fluency in the context of foreign language teaching and learning. That means producing speech smoothly with few hesitations, false starts, or pauses. The current study focused on what Lennon (1990) called the narrow sense of fluency, what Segalowitz (2010) called temporal features of utterance fluency and what Tavakoli and Hunter (2018) called a very narrow perspective of fluency (objective measures). Perceived and cognitive fluency are not the focus of this study, as examining both would require different procedures and approaches in data collection, which was not possible in the current study.

2.4 Measuring Fluency

Temporal measures that are related to the speed of speech and the extent of pausing have been used as standard measures for L2 fluency studies, such as the early study of Lennon (1990). The repair measures of L2 fluency were added later by Skehan (2003) and Tavakoli and Skehan (2005). Skehan (2003) was the first researcher to create a comprehensive fluency framework that measured three aspects of fluency: speed, breakdown, and repair. Since then, these three fluency dimensions have been

the standard for choosing the most reliable fluency measures (e.g., Kormos, 2006; de Jong, 2012; Bosker et al., 2013; Tavakoli et al., 2020).

Although a considerable amount of literature has been published on utterance fluency, there is still a lack of consensus on the best fluency measures to employ because there is no general agreement about choosing specific measures and ignoring others (Guz, 2015; Foster, 2020). It is important to mention that there is no straightforward measure or task for fluency (Leonard, 2015). Thus, second language acquisition (SLA) researchers have used different measurements at different times for different reasons.

This section introduces and evaluates in detail the common measures of oral fluency, mainly utterance fluency measures (Segalowtiz, 2010), because they are the focus of the current study. SLA researchers such as Skehan (2003; 2014), Kormos and Dénes (2004), Kormos (2006), de Jong et al. (2012; 2015) and Kahng (2015) have generally agreed on the most reliable measures for different aspects of utterance fluency under the three major sub-components that form the basis of fluency research (speed, breakdown, and repair). The first subcomponent includes speech rate, mean length of run, phonation time ratio, and articulation rate. The second subcomponent includes filled/unfilled pause location, frequency, and length. The third subcomponent includes repetition, self-correction, reformation, and false starts. These are defined in turn in the following section.

2.4.1 Speed Fluency

Speech rate (SR) is a composite measure that combines speed and pausing aspects. The SR measure is one of the most important aspects of fluency and it is widely used to detect changes in the fluency of oral production (Huensch and Tracy–Ventura, 2017). It estimates the speed of the speech, and it varies from one speaker to another.

Some people speak quickly while others speak slowly (de Jong et al., 2012). SR is a temporal variable and an essential measure in fluency because the more a speaker pauses, the slower their SR will be (de Jong, 2016). Measuring a speaker's speed rate is calculated by counting the total numbers of syllables divided by the total spoken time, including the number and length of both silent and filled pauses (Skehan, 2003; de Jong et al., 2012).

SR is a widely used measure in task-based research studies (e.g., Gilabert et al., 2011; Michel, 2011) that focus on complexity, accuracy, and fluency. It provides an overall understanding of fluency. Other researchers (e.g., Witton-Davies, 2014; de Jong, 2016) argue that the articulation rate is a more specialized measure that should be used to avoid the overlap across complexity, accuracy and fluency (CAF) dimensions and to measure fluency more precisely.

The syllable is the basic unit used in psycholinguistic studies to calculate temporal measures in fluency such as SR, phonation time ratio (PTR), and articulation rate (AR). Influential studies by language teaching researchers (e.g., Lennon, 1990; Kormos and Dénes, 2004; Skehan and Tavakoli, 2005; de Jong et al., 2015; Bui and Huang, 2016; Peltonen, 2017; 2018) preferred to use the number of syllables per minute as the unit of reference for measuring speed over words per minute because counting words per minute is imprecise and some words are longer than others. Additionally, the online speech analysis tool PRAAT, which is used to measure SR, is based on syllables rather than words.

Tauroza and Allison (1990) recommended this method for calculating the SR by using the number of syllables per minute, instead of the words per minute. The number of syllables per word differs by genres; for example, the SR in dialogue will be faster than

the SR in monologue tasks if syllables per word are used as a speed measure. This is because speakers are likely to speak more in a dialogic task than in a monologic task. Furthermore, dialogue is a shared responsibility between two speakers who exchange talk to maintain the flow of the speech. Tauroza and Allison (1990) found that SRs in conversation were higher than SRs in lectures. They also suggested that the disadvantage of counting the number of syllables rather than the number of words is that the number of syllables in the transcripts might not be the same as the number of syllables in actual speech production. For example, some participants might say 'istudent' by inserting 'I' in the beginning of the word. Therefore, the researcher should decide what to count as a syllable in the speaker's oral production.

Articulation rate is another temporal variable for speed fluency that excludes pausing behaviour in speech. It is included in this study to detect the speed of the speakers' speech regardless of their filled/silent pauses. AR is considered to be a pure measure of the speed of delivery, and it is unrelated to other fluency measures mathematically because AR does not count filled/silent pauses. AR can be measured by dividing the total number of syllables by the spoken time, excluding silent pauses (Bosker et al., 2012; de Jong, 2016). According to de Jong (2016), AR is a precise way of examining SR because it is related to the articulation stage of speech production mentioned in Levelt's (1989) model. Witton-Davies (2014) claims that AR is a useful research tool that differentiates between pausing and rate, unlike SR which combines both. When SR is used alone, it is not possible to know whether the number of pauses or the AR causes a higher SR in one speaker compared to another. Thus, both SR and AR are used in this study to measure speech units (syllables) per speaking time, but AR measures the speed of the speech without counting the amount of pausing and SR measures the speed of the speech taking pausing into account.

Composite measures are measures that combine aspects of speed and breakdown. Skehan (2003) recommends measuring what causes disturbance to the flow of the speech (e.g., speed fluency) and what slows down the speed of the speech fluency (e.g., breakdown fluency). The mean length of run (MLR), the mean length of syllables (MLS), and the PTR are defined as composite measures that show the relationship between speed and pausing. They are salient measures that have been used by several studies, such as Derwing et al. (2004), Kormos and Dénes (2004), and Bosker et al. (2014).

MLR is a measure of the degree of automaticity (Towell et al., 1996; Skehan, 2014a). PTR shows the amount of pausing that is related to speech (Towell et al., 1996). The present study included the SR, AR, MLR, and PTR of L2 speed fluency to gain an overall view of the participants' oral fluency in monologue and dialogue tasks.

2.4.2 Breakdown Fluency

Speed fluency is an essential temporal measure in oral language, but it cannot measure other dysfluency markers like filled/silent pauses. The multidimensional nature of fluency requires a set of measures that can fully examine aspects of speech performance. Thus, the breakdown fluency measure is related to the continuing flow of the speech (de Jong et al., 2012). It is important to explain some of the reasons for pauses. Pauses in speech, whether they are filled or silent, could be produced to ask for clarification, show empathy or assess the listener's understanding (Foster, 2020). According to Tavakoli and Wright (2020), pauses are produced in speech to plan for the next utterances, to retrieve lexical and grammatical information, or to monitor overall speech processing.

L2 research on breakdown fluency has focused on pause frequency, length and location to examine the utterance fluency of learners with different L1s (e.g., Tavakoli, 2016), or learners with different proficiency levels (e.g., Peltonen, 2018). Examining filled/silent pauses in these three regards (location, number, and duration) is common when measuring breakdown fluency because these measures distinguish between native and non-native speakers. The number of pauses is known as an indicator of disfluency, meaning that low- proficiency learners tend to pause more often than high-proficiency learners (e.g., Kormos and Dénes, 2004; Tavakoli et al., 2020). Non-native speakers tend to pause more frequently than native speakers (e.g., Kahng, 2014). In addition, measure of pause frequency has been used as an indicator for disfluency linked to raters' judgements of a speaker's fluency (Cucchiarini et al., 2002; Iwashita et al., 2008; Sato, 2014).

The location of pauses is another important measure of breakdown fluency. Pause location refers to whether the pause occurs in the middle of a clause or at the end of a complete clause (Kahng, 2017; Shea and Leonard, 2019). These two distributions of the pauses tend to capture how often the L2 speaker pauses within clauses and at clause boundaries. Pause location distinguishes between different fluency levels (Kahng, 2014). Several previous studies, for example, Tavakoli, (2010), Wood, (2010), Bui and Huang (2016), and Kahng (2017) found that L2 speakers paused more often within clauses due to difficulties with encoding or planning their speech, whereas L1 speakers (native speakers) paused at the end of clauses or grammar junctures.

Additionally, pause location is found to be related to Levelt's (1989) stages of speech production processing. In other words, a pause within a clause is related to the formulation stage, whereas a pause at the clause boundary reflects the

conceptualization stage of speech production (Kahng, 2017). According to de Jong (2018), L2 learners are likely to produce frequent and long pauses within clauses. Frequent pausing, for example pauses in the middle/within or at the beginning of clauses, does not always mean disfluency. It might occur because the L2 learner needs more time to retrieve syntactic and linguistic knowledge, which is related to the formulation stage in speech production processing (Tavakoli, 2011; de Jong, 2018).

As for pause duration, the results are mixed: the pausing phenomenon can discriminate between learners' levels of fluency. According to Kormos and Dénes (2004), the length of pauses differs between fluent and non-fluent speakers, while Cucchiarini et al. (2002) and de Jong et al. (2015) found that average pause duration was not a strong predictor for L2 fluency. Quantitative studies, such as those by Lennon (1990), Riggenbach (1991), Freed (1995) and Hasselgreen (2005) indicated that fluent L2 speakers are likely to have short pause length (0.025<ms), fewer filled/silent pauses within clauses, longer run, and higher phonation time ratio, and SR. Furthermore, these measures were found to be correlated with raters' judgements of L2 oral fluency performance across studies (e.g., Kahng, 2014; 2018). More recently, Tavakoli et al. (2020) argued that pause duration is related to L2 learners' proficiency level. The duration of pauses in higher L2 proficiency learners was shorter than in lower L2 proficiency learners. In addition, de Jong et al. (2012) found that L2 fluency measures of English predicted L1 fluency measures of Dutch learners. For example, pause length behaviour in L2 English could be related to L1 Dutch pausing behaviour regardless of the speaker's proficiency level. Thus, these mixed results regarding pause length could be explained by individual differences across languages (L1 versus L2) (de Jong et al., 2012; Shea and Leonard, 2019).

The duration and number of filled pauses (FPs) is a common measure of breakdown fluency in many L2 oral fluency studies, such as Cucchiarini et al. (2002), Kormos and Dénes (2004), Kahng (2014), and Tavakoli et al. (2020), whereas their location is less commonly used (e.g., end-clause FPs). Filled pauses have been founded to be linked to individual speaking styles as some speakers consider FPs to be part of their natural speech (e.g., Cucchiarini et al., 2002; Kahng, 2014; Shea and Leonard, 2019). Due to the mixed results and the scarcity of the studies that have included FPs as a measure of L2 fluency, the present study utilises the location, frequency, and duration both FPs and SPs.

2.4.3 Repair Fluency

Repair fluency refers to signs of dysfluencies such as false starts, repetitions, reformations, or replacements (Skehan, 2003; de Jong et al., 2012). False start means an utterance or idea that is discarded or eliminated without completing it. Repair can be a repetition, meaning any partial or complete repetition of a word, phrase, or clause without any modification (de Jong et al., 2012). Another example of repair is reformulation – that is, any phrases or clauses that are repeated with modification to the words order, syntax, or morphology (de Jong, 2016). Replacement means words or phrases that are immediately substituted for others (Ellis et al., 2005). These dysfluencies are often found in the speech of native speakers as well as that of L2 speakers. Repairs have been investigated by several studies (e.g., Riggenbach, 1991; Tavakoli and Skehan, 2005; Bosker et al., 2012; Kahng, 2014). On the one hand, some of repair measures have been shown to affect the smoothness and speed of oral fluency performance. For example, in Bosker et al. (2013), FPs and repetitions

were frequently used by less fluent speakers. On the other hand, repairs were used by advanced speakers to maintain the flow of their oral speech.

Finally, repair is claimed to be related to the speech processing stages (Levelt, 1989; Kormos, 2006; Witton-Davies, 2014). For example, L2 learners might use repetitions and reformations to have more time for online planning. Thus, repetitions and reformations could be related to the conceptualization and formulation stages of Levelt's (1989) L1 speaking model, while reformations, replacements, and false starts are assumed to be related to the third and fourth stages of L2 speech processing: articulation and monitoring (de Bot, 1992; Kormos, 2006).

Towell et al. (1996) argued that although the previously mentioned repair measures are good predictors of fluent speech, the length and number of pauses are not because they might be characteristics of an individual's speech. Also, online planning and task types are among those variables that can affect the speech fluency measures.

Finally, Foster (2020) argues that social, physical, cognitive, and linguistic variables are independent variables that can affect L2 fluency in speech. For example, it is possible that a speaker spends time preparing the preverbal message and modifying it before, during and after the formulation phase of putting the intended message into words. This dysfluency can be related to the individual speaking style, the social context, or the subject of the speaking task. Sometimes a speaker deliberately uses disfluencies to aid the listener's understanding, to ask for attention, or to show empathy. As for the linguistic ability variable, it is related to L2 knowledge that is not fully automated because it may require robust paraphrasing with the available resources at hand.

In addition, the individual differences in WM may cause variation in the ability to retrieve and produce the structure of the target language (Wright, 2013). As for utterance fluency, the lack of access fluidity (Segalowitz, 2016) to syntactic knowledge and the limitation of WMC (Baddeley, 2007) slow down the speed of L2 speech by producing pauses, hesitations, or repetitions. Thus, this does not always mean that the L2 speech is wrong, but that it is slow as a result of monitoring the meaning and form of the speech (de Bot, 1992; Kormos, 2006).

Having discussed the most influential frameworks of first and second language speech production processing and defined what is meant by fluency, and what are measures of fluency, I'll now move on to discuss the empirical studies of L2 fluency.

2.5 Empirical Studies on L2 Fluency

This section introduces studies on fluency from two aspects: the methodological perspectives and findings. Both aspects are important in showing how fluency is examined in the literature. Furthermore, the focus of this thesis is to introduce new findings to the literature of L2 fluency, and language learning and teaching. First, this section starts with studies that examined the factors affecting oral fluency, such as L2 proficiency (e.g., Iwashita et al., 2008; de Jong et al., 2013), first language, and WM because to some extent these factors influence L2 oral fluency performance.

A range of studies have studied the oral production features in L2 fluency by employing three common strategies (Kahng, 2014): comparing L1 speech samples with L2 speech samples in terms of utterance fluency measures (e.g., de Jong, 2016, Derwing et al., 2009); using a longitudinal approach to study the development of L2 fluency (e.g., Lennon,1990; Towell et al., 1996; Segalowitz and Freed, 2004; Leonard, 2015); and using native speakers' judgements by relating utterance fluency to perceived

fluency across different tasks (e.g., narrative or descriptive) (e.g., Derwing et al., 2004; Kormos and Denese, 2004; Tavakoli, 2011; Bosker et al., 2014). Finally, a summary of the most common L2 utterance fluency studies is provided in Table 1 (pp. 53-58).

2.5.1 First Language Fluency and Longitudinal Studies

Research that investigates the relationship between L1 and L2 fluency is of interest because some L2 researchers (e.g., de Jong et al., 2013; Leonard, 2015; de Jong, 2016; Peltonen, 2018) have found that fluency is a personal trait while others, such as Derwing et al. (2009), state that fluency is a language-specific trait. They have also published significant findings through their investigations of different aspects of L2 production. Some of these studies used temporal features, language typologies of L1 vs. L2, fluency measurements across modes, and developments in cognitive ability that have implications for the study of fluency. The hypothesized relationship between L1 fluency and L2 fluency has drawn the attention of some L2 researchers and prompted them to investigate this relationship in different contexts. Using L1 and L2 data to examine the influence of personal speaking styles and differences between the two languages in terms of utterance fluency may provide a new understanding of L2 speech fluency.

For example, Derwing et al. (2009) compared the scores of fluency temporal measures of L1 speakers (Slavic and Mandarin) and L2 speakers (English) with native speakers' judgements of the same participants' speech samples. The study was conducted to examine whether the temporal characteristics of the participants' L1 predicted the temporal characteristics of their L2. The L1 and L2 speech performances of the same participants were collected over two years while they studied abroad. The participants conducted a picture narrative task and a picture description task. Oral production data

were measured by number of syllables, filled/unfilled pauses, repetitions, self-corrections, and false starts. The results of the native raters' judgements indicated that there was a weak relationship between L1 fluency and L2 fluency. The correlation was stronger for the Slavic than for the Mandarin speakers of English. Derwing et al. (2009) claimed that fluency could not be transferred from L1 to L2. However, the authors argued that several factors, such as proficiency level, amount of instruction, properties of L1, experience, and cognitive factors might prevent the development of fluency. Moreover, the ESL program abroad did not focus heavily on speaking skills and activities, but rather the focus was on teaching grammar and academic English. This led to a high increase in declarative knowledge (explicit knowledge) but a slower development of procedural knowledge (implicit knowledge), which is to some extent responsible for oral fluency.

In contrast to Derwing et al.'s (2009) study, de Jong (2016) compared the speech fluency of native speakers of Turkish, Dutch, and English who spoke Dutch as their L2. Utterance fluency was measured by filled and silent pause locations and duration, either within clauses or between clauses. She also investigated the pause locations before both low- and high-frequency words. The results indicated that L2 speakers paused within utterances more frequently and for longer than L1 speakers, who tended to pause between clauses, and at utterance boundaries. However, the study showed that when L1/L2 speakers produced longer constituents/clauses, there was more opportunity to pause within them than when L1/L2 speakers produced shorter constituents/clauses.

The data also showed that both L1 and L2 speakers paused before lower frequency words more often than before higher frequency ones. L1 and L2 speakers often

paused at utterance boundaries, which is what Levelt (1989) called the stage of conceptual planning and lexical retrieval in speech processing. Overall, the findings from Derwing et al. (2009) and de Jong (2016) showed that some measures of L2 fluency are related to L1 fluency, such as SR, silent pauses (SPs), FPs, and speakers' speech monitoring style.

Peltonen (2018) argues that L1 fluency measures are one of a significant factor that should always be included when measuring oral fluency. She explored the relationship between L1 and L2 fluency in the monologue speech of forty-two Finnish students. A mixed-method approach was used to first examine the relationship between L1 and L2 fluency measures. Second, a qualitative analysis of a picture description task was conducted to compare students' L1 fluency and L2 fluency through the use of stalling mechanisms (drawls, fillers, and repetitions).

The results indicated a high correlation between most L1 and L2 fluency measures, such as temporal measures (SR, MLR, mid-clause SPs, FPs, and mean length of end-clause SPs). For stalling mechanisms, the results were mixed across L1 and L2 because both L1 and L2 participants differed in how they dealt with time pressure. They tended to prefer certain types of stalling mechanisms over others. Some L1 participants used more fillers while the L2 participants used more drawls and repetitions to buy time to form the intended message and recall syntactic knowledge. Thus, it can be said that stalling mechanisms are related to individual speakers' styles. As for the regression analyses, Peltonen (2018) confirmed that participants' style in their L1 is an important individual factor that could predict L2 fluency performance and it should therefore be included in future L2 fluency research and assessment.

Zuniga and Simard (2019) examined self-repair (a temporal feature of repair fluency) in fifty-eight participants who spoke French as their L1 and English as their L2. The participants' self-repair was examined by asking them to carry out two picture narrative tasks, one in L2 English and the other in L1 French. The tasks were similar in terms of vocabulary, plot, and length. The participants' L1 self-repair, English proficiency and attentional control were tested to investigate the effects of these factors on L2 self-repair. The results indicated a strong correlation between L1 and L2 production in terms of self-repair. In addition, the regression analyses showed that L2 self-repair behaviour explained 40% of the variance, whereas L2 proficiency made very little contribution to L2 self-repair behaviour. Zuniga and Simard (2019) concluded that repair fluency is an underlying trait that is not related to proficiency. This means that L1 self-repair is linked to L2 self-repair. Even if the learner's L2 proficiency increases, L2 self-repair will not change because it is a stable personality trait.

More recently, Duran-Karaoz and Tavakoli (2020) investigated the relationship between L1 (Turkish) fluency, L2 (English) fluency, and fluency behaviour in terms of speed, breakdown, and repair across different proficiency levels. The study also examined whether language proficiency mediates the relationship between L1 and L2 oral fluency. Two oral narrative tasks were collected from 44 L2 Turkish participants and coded for speech rate, articulation rate, repair, and SP location, and FP location. Results suggested that there was a relationship between L1 and L2 fluency in terms of mid-clause FPs and mid-clause SPs, speech rate, and total number of repairs except for articulation rate. Regression analyses suggested that some measures of L2 speed and breakdown fluency measures could not be predicted by L1 speed and breakdown fluency.

2.5.2 Study Abroad and L2 Fluency Development

Investigating development in speaking skills is always the goal for longitudinal research conducted in a study-abroad (SA) context. As well as L1 fluency has been considered when investigating the predictable variables of L2 fluency gain. An important finding in the SA field was published by Lennon (1990). He examined the oral fluency of 4 EFL students who had studied English for 6 months in the UK. Data were collected twice; the first time was after the students' arrival into the UK and the second was before the students' departure for Germany. Lennon (1990) used trained EFL native speakers' judgements to rate the participants' spontaneous speech using global fluency measures such as speed delivery (speech rate), MLR, filled/unfilled pauses, self-corrections, and repetitions. The results indicated that words per minutes, position of filled pauses, and SR were strong indicators of L2 students' oral fluency development, while self-corrections or reformations were poor indicators of dysfluency. Lennon explained that the reformation of some sentences in speech is a common feature, even in native speakers' speech, and repair tends to increase in the oral performance of advanced L2 speakers.

Towell et al. (1996) longitudinally investigated the fluency development of advanced French students studying English as their L2 in the UK for four years. Towell and his colleagues used Levelt's model of speech production as a general base to track the L2 temporal development process from conceptualization to formulation and then to articulation. They believed that fluent L2 oral production happened when declarative knowledge was converted to procedural knowledge within the three speech production phases. Therefore, Towell et al. carried out this study to prove the above claim and to investigate how and when this proceduralization happened. The results of the L1 and L2 speech tasks showed that students developed because of syntactic complexity not

because of the speed or numbers of the produced pauses. Towell et al. (1996) suggested that “the component of the production model in which proceduralization takes place is the formulator, which is centrally concerned with assigning grammatical representations to messages and passing these grammatically-encoded messages to the articulator” (p.113)

According to de Jong et al. (2013), there are two potential directions that L2 researchers (e.g., Segalowitz and Freed, 2004; de Jong, 2013; Leonard, 2015) use to examine the relationship between L2 cognitive fluency and L2 utterance fluency. The first is through a longitudinal study *within participants* to examine the development of some utterance fluency measures over a period of time (e.g., Segalowitz and Freed, 2004; Leonard, 2015). On the one hand, if the participants develop over time in a specific measure of utterance fluency, it can be assumed that their L2 cognitive fluency has improved over time. On the other hand, if there has been no change in any aspect of utterance fluency measures over time, despite an improvement in the overall proficiency of the L2 participants, it can be assumed that this measure of utterance fluency is not related to L2 cognitive fluency. Instead, it is probably related to the participants’ general cognitive skills or to their individual speaking style because these have not changed over time. The second direction for studying the relationship between L2 cognitive fluency and utterance fluency is to examine these relationships directly (e.g., de Jong et al., 2013).

For example, Segalowitz and Freed (2004) used the first direction, a longitudinal study to track L2 fluency gain in study abroad (SA) vs. at home (AH) contexts. This study compared utterance fluency and oral proficiency data in two groups of American participants learning Spanish as their L2 in both contexts, AH and SA. Segalowitz and

Freed (2004) examined the interactions between three variables: oral production abilities, learning contexts and L2-related cognitive abilities. Utterance fluency was measured by analyzing oral data from an Oral Proficiency Interview (OPI) in terms of temporal features (e.g., speech rate and mean length of run). Cognitive fluency was measured by a word recognition test to examine the speed of lexical access and a repeat and shift task to examine attentional control. The results showed a complex relationship between cognitive measures of fluency, L2 fluency gain, and language contact. Study-abroad participants developed more temporal aspects of L2 utterance fluency than AH participants. Additionally, there is also a strong correlation between fluency measures, and cognitive aspects tasks: speed efficiency of lexical access and attentional control.

Leonard (2015) longitudinally tracked the L2 fluency development of thirty-nine English participants who were studying L2 Spanish abroad for one semester. By employing a pre-test/post-test design, Leonard (2015) investigated the relationship between different variables: (1) L2 linguistics knowledge as measured by grammar and vocabulary; (2) L2 cognitive processing speed as measured by a picture naming task and a sentence completion task; (3) L2 utterance fluency as measured by speed, breakdown, and repair; (4) the influence of the participants' L1 English; and (5) their previous knowledge of L2 Spanish. In addition, a number of fluency measures were examined: MLR, AR, repair, and pause location and frequency.

Leonard (2015) found that the L2 participants who started the semester abroad with low fluency scores showed a remarkable development in L2 fluency performance at the end of the semester, for example in pausing phenomena and AR. In contrast, other participants with high fluency scores prior to their SA showed higher L2 fluency gain

at the end of the semester abroad. Additionally, he argues that having previous linguistics knowledge before studying abroad does not always give L2 learners significant gain in L2 fluency but it encourages them to improve their fluency either at home or abroad. Overall, the findings showed no significant gains regarding the measurements of L1/L2 filled/unfilled pauses or hesitations. It would seem that they are not related to L2 proficiency but to speakers' speaking styles (Segalowitz, 2010) because some learners use either FPs or SPs and sometimes both when they face problems with formulating the intended message.

Regarding the second direction of measuring the relationship between L2 cognitive fluency and utterance fluency, de Jong et al. (2013) examined whether L2 cognitive fluency measured by linguistic knowledge and processing speed can explain the variations in L2 utterance fluency performances as measured by speed, breakdown, and repair. L2 participants' linguistic knowledge was measured by sentence completion tasks to examine their grammar and vocabulary knowledge. A lexical retrieval task was also used to measure the participants' processing speed. The results showed that some measures of utterance fluency were good predictors of cognitive fluency. For example, the findings indicated that linguistics skills were strongly correlated with the duration of syllables, whereas pause duration was weakly correlated with linguistic knowledge.

2.5.3 Native Speakers' Judgement

The literature on L2 fluency has considered native speakers' judgements in order to measure L2 utterance fluency by comparing the speech components of native speakers with those of non-native speakers and then correlating the temporal variables of utterance fluency with perceived fluency. In a significant study by Kormos

and Dénes (2004) set out to examine the perceptions of fluency of native and non-native teachers in twelve Hungarian L2 students' speech performance. The accuracy and lexical diversity of the L2 participants' speech were examined and rated in terms of temporal variables of oral fluency. These were PTR, MLR, and AR, SR, and stressed words/60. The temporal variables affected both native and non-native teachers' judgements about L2 fluency. Teachers also had different judgements about the influences of the following features on their perceptions of L2 fluency: lexical richness, length of pauses, and grammatical accuracy. Some disfluency aspects such as FPs and SPs did not affect the fluency perceptions of either group of raters.

To determine the effect of native speakers' judgement, Derwing et al. (2004) investigated the fluency performance of L2 participants across three oral tasks. Trained and untrained native-speaker judges rated the speech samples of low-proficiency L2 speakers for comprehensibility, fluency, and accentedness. Data for Derwing et al. (2004) showed that there was no significant difference between trained and untrained native-judges in prosody. Additionally, there was a strong correlation between fluency and comprehensibility and to lesser extent accentedness.

In a study by Tavakoli (2011), the locations of mid-clause silent pauses and end-clause silent pauses were analyzed in native and non-native speeches to compare the oral fluency of both groups. The results indicated that native speakers paused at the end of utterances while non-native speakers paused in the middle of utterances. Non-native speakers paused to possibly think of suitable words or to revise their clauses before completing their oral production.

Bosker et al. (2014) conducted four experimental studies to investigate the relationships between the three measures of utterance fluency (speed, breakdown,

and repair) and the relationship of each measure to perceived L2 fluency. The purpose of these four experiments was to investigate which speech characteristics contributed to the general listeners' sensitivity. In the first experiment Bosker et al. (2014) used untrained judges to rate L2 Dutch speakers' speech performance. He argued that other studies such as those by Rossiter (2009) and Derwing et al. (2004) did not find differences between trained and untrained judges' ratings of fluency. The results showed that two aspects of utterance fluency (breakdown and speed) predicted the perceived fluency, while the sensitivity of the listeners did not contribute to the aspects of speed, breakdown or repair. It has become clear that listeners' judgements contributed the most to overall L2 speech rating.

2.5.4 Oral Fluency and L2 Proficiency

Another area of L2 fluency development is conducting cross-sectional studies in EFL classrooms to study the role of L2 proficiency in L2 oral fluency (e.g., Kormos and Dénes, 2004; Iwashita et al., 2008; Hilton, 2008, 2014; Tavakoli et al., 2020). Iwashita et al. (2008) investigated the relationship between L2 test takers with different proficiency levels and their performance across five speaking tasks in TOEFL (IBT). Students' overall proficiency performance was examined in three features: linguistics resources (grammar and vocabulary), fluency, and phonology. The cross-sectional investigation showed that knowledge of vocabulary and oral production features (phonology and fluency) were strong predictors that influenced the raters' scoring decisions across all levels of the speaking tasks. In contrast, the grammar complexity and accuracy were less significant predictors for overall speaking proficiency. Similarly, Iwashita (2010) examined four features of oral proficiency in two language groups: English as EFL and Japanese as JFL. The participants' oral productions were

rated across five narrative tasks to see which speech features distinguished the two levels of proficiency (low, high). Overall, the results of the quantitative analyses of four measures (grammar accuracy, fluency, vocabulary, and syntactic complexity) were in line with the previous study of Iwashita et al. (2008), which confirmed that vocabulary and oral features determined the overall oral proficiency of L2 learners.

In addition, owing to the complexity of the construct of fluency, different approaches have been taken by researchers, but many have concluded that developments in utterance fluency are linked to developments in proficiency levels (de Jong et al., 2013), cognitive fluency or sometimes speakers' speaking styles. de Jong et al. (2013) argue that it is important to establish which aspects of utterance fluency are indicators of automatic L2 speaking in order to identify the characteristics that learners must develop. de Jong et al. (2013) examined the extent to which L2 linguistic knowledge and processing speed predicted L2 utterance fluency as measured by breakdown, repair, and speed. They found that linguistic skills correlated with mean syllable duration and explained 50% of participants' variation in speaking fluency. Furthermore, they claimed that the negative correlation between the mean length of silent pauses and the measures of L2 cognitive fluency could have been explained by L1 fluency or personal speaking characteristics.

Most of the studies reviewed in the previous sections have primarily focused on L2 fluency in a monologue. This leaves dialogue as the neglected area of investigation in fluency studies. Therefore, the next section will look at the differences between the performances of both modes. Furthermore, a few studies will be reviewed that examined the significant role of L2 utterance fluency in dialogic tasks.

2.6 The Differences between Monologic and Dialogic Performances

Monologue is the production of speech by a one speaker who depends on his/her own resources to complete the speech, while in dialogue speakers take turns and thus constantly change roles – the speaker changes from being a speaker to a listener (McCarthy, 2010; Tavakoli, 2016). In dialogue, both the speaker and the listener try to avoid overlapping turns by following a universal tendency to reduce pauses between turns (Garrod and Pickering, 2004). The goal of dialogue is for both the speaker and the interlocutor to reach a common conception or understanding of what they are talking about, or the conversation will fail (Garrod and Pickering, 2004). Achieving a common perception in dialogue requires speakers to accommodate and relate their choices of syntactic and phonological structures to the ongoing dialogue (Kootstra et al., 2009). This leads speakers to use each other's vocabulary and linguistic structures and limit the interlocutor's selections of language production (e.g., changing the topic) (Kootstra et al., 2009). As for monologue, the goal of speaking is to encode the preverbal message into grammatical and phonological structures (Levelt, 1989).This stage of conceptualization is a direct fixed procedure from the intended message to the articulation output.

Additionally, in language processing and speech production, it is traditionally believed that a monologue is a more explicit task than a dialogue. In other words, de Jong and Perfetti (2011) and Tavakoli (2011) argue that in a monologue task it is easy to control the speakers' performances (e.g., in terms of the topic choice), and there are more predictable outcomes and less demanding pragmatic planning. Alternatively, it has been argued that measuring L2 fluency in dialogue is less controlled and has less predictable outcomes. Despite all of this, it does seem that dialogue is the more authentic situation owing to its interactive nature, reflected in everyday communication

such as talking to a friend, a colleague, or a classmate (Garrod and Pickering, 2004; Tavakoli, 2016). Arguably, a monologue also lacks several speech features that usually exist in dialogue, such as: (1) clarification requests (e.g., excuse me, can you repeat that please?); (2) interruptions by the interlocutor; (3) between-turn pauses (e.g., gaps); (4) overlapping speaking time (Edwards, 2008); (5) collaborative completion (Peltonen, 2017); and (6) interactive alignment (Garrod and Pickering, 2004).

The 'between-turn pauses' are developed by speakers during dialogue when one speaker remains silent while the other speaker takes the turn to respond. Contrastingly, an overlap occurs when two speakers speak simultaneously, and one of them dominates the floor while the other remains silent (Wilson and Don, 1986). Collaborative completion happens when one speaker completes a sentence or a thought that has been started by the other speaker.

Thus, it can be concluded that dialogue is easier than monologue because in the latter the speaker starts every single talk from scratch, whereas in the former, the dialogue partners are collaborating to alternate each other's speech and facilitate the process of their oral production (Garrod and Pickering, 2004).

2.6.1 Dialogue Fluency

L2 dialogue fluency could be explained in terms of interactive alignment model (Garrod and Pickering, 2004). Fluency in dialogue has been also defined by researchers (e.g., McCarthy, 2010; Sato, 2014; Peltonen, 2017) into several terms such as confluence, interactional fluency, and collaborative fluency.

Starting with interactive alignment model, L2 fluency can be conceptualized in the context of dialogue through the interactive alignment model, as it explains how interlocutors might affect each other's performances. Garrod and Pickering (2004) argue that conversation is a simple task because the interactive alignment between partners automatically links behaviour and perception. Alignment in the conversation means that people take advantage "of each other's choices of words, sounds, grammatical forms, and meanings" (Garrod and Pickering, 2004, p.9). The interactive nature of the dialogue predicts that a person tends to unconsciously mimic his/her partner's behaviour, such as facial expressions and the accent or tone of the speaker's voice. Thus, it can be argued that the linguistic representations of the alignment support and distribute the processing load between the partners during the conversation (Garrod and Pickering, 2004).

Alignment as a socio-cognitive phenomenon can also happen in language classroom conversation because the interlocutors share similar knowledge, backgrounds, or language processing skills. It is a useful framework used by researchers to draw predictions about language processes in dialogue and to help L2 teachers refine or create pronunciation activities in classrooms (Pickering and Garrod, 2004). This does not mean that the process of alignment is entirely repetitive, but it makes the listener update his/her model so that it is similar to the speaker's model (Trofimovich, 2013).

McCarthy (2010) introduces the term confluence as a definition for fluency in a dialogue. It means that two speakers cooperate in a joint processing to create a fluent interaction. Fluency in a conversation is created interactively between two speakers. Thus, the responsibility is shared between the speakers in a dialogue to maintain and create a flowing dialogue and fill the silences. McCarthy (2010) notes that confluence

can be measured by turn-taking and subjects' impressions of the smoothness of the conversation.

Sato (2014) used the term interactional fluency. He studied interactional oral fluency by focusing on peer interaction tasks and comparing performance in these tasks with individuals' performances. A total of fifty-six L2 Japanese learners received paired decision-making tasks and their perceived fluency was investigated by four native English raters who based their evaluation on the learners' oral production. The results revealed that L2 fluent speakers used backchannels and turn-takings in natural patterns during peer interaction speech, whereas less fluent speakers were hesitant to initiate turns. Raters' perceptions were compared with individuals' performance to differentiate fluent from disfluent speakers based on temporal aspects (e.g., pruned/unpruned speech and pauses) and interactional features (e.g., turn-takings). The results of the raters' judgements indicated that the participants' SRs were higher in the pair-interaction task than in the monologue task. Regarding the temporal measures, the raters considered pauses as a dysfluency marker that slowed the speech in both performances. The regression analysis revealed that individual learners' performance was weak and not important in terms of predicting their performance in an interactional context.

More recently, Peltonen (2017) introduced the term dialogue fluency to indicate the individual speakers' contributions to the collaborative forms of fluency in dialogue. Participants' performance in dialogue was measured by objectively calculating the number of repetitions, the mean length of turn pauses, and the number of collaborative completions. The present study uses this definition for measuring L2 dialogue fluency excluding the measure of collaborative completion and turn pauses.

2.6.2 Oral Fluency Studies on Dialogic and Monologic Tasks

Kirk (2016) argues that there are two basic approaches adopted in the research on L2 fluency. One is looking at the monologic construct of fluency from psycholinguistic perspectives (see Table 1, p. 53-58), for example cognitive fluency that taps into attentional control and access fluidity. Attentional control is based on organizing the language in LTM, whereas access fluidity is explained as fast access to long-term WM. This results in a higher SR, fewer pauses and longer speech run (ibid). Thus, Kirk (2016) argues that monologic fluency is considered a traditional approach in fluency research, and it belongs to an individual speaker. The second approach to fluency is the dialogic construct where fluency in dialogue depends on both speakers' cognitive fluency. A conversation is co-constructed by both the speaker and the interlocutor with smooth turn takings (ibid). However, sometimes fluency in dialogue is affected by unfamiliar grammatical sentences or vocabulary choices. As a result, co-construction in dialogue could become more difficult between speakers and this might result in hesitations and long silent or filled pauses between turns (Kirk, 2016).

When reviewing the literature, it was noted that dialogue studies have been used frequently for L2 testing and assessment. Examples of these are a structured interview, IELTS (long turn speaking test), TOEFL (computer-based interview), and OPI (oral proficiency interview). IELTS involves face-to-face interaction with the examiner and the examinee may have some time to prepare his/her speech to answer the question. The examiner's role can be open, guided, or both. OPI is used to examine the participants' overall oral proficiency (Freed et al., 2004). OPI employs several questions that give more flexibility to the questioner and to the candidate. In this test, the questioner can select and change the questions to make them suitable for the candidate's level (Witton-Davies, 2014).

Addtionally, some of the previous studies fluency focused on dialogue tasks only, for example Bortfeld et al. (2001), Davis (2009), McCarthy (2010), and Peltonen (2017). Bortfeld et al. (2001) examined the disfluency rate in conversation tasks and other variables that might affect L2 fluency performance, such as age, gender, the relationship between speakers, task role, and topic difficulty. The results showed that older participants appeared to produce more disfluent speech than younger learners. The participants' disfluency also increased when a task contained a high planning demand. For example, a diagram task was more complex than a picture description task. Regarding gender, men produced more fillers and repetitions than women. However, fillers in speech, such as 'mm' or 'uh' were used to maintain the speech fluidity and to tell the interlocutor that the speaker was about to talk.

In a similar study, Davis (2009) argues that speech assessment in a pair test is complex and challenging because different factors may affect the oral production, for example, interaction with the partner, the interlocutor's proficiency, task type, and raters' judgement. Davis examined the influence of the interlocutor's proficiency level on the speaker's score and the number of words in the opinion paired task. The results showed that the interlocutor's proficiency level had little influence on the score of the examinee's speech. However, a higher proficiency level partner maintained the flow of the conversation and produced more words than a lower proficiency level partner. In addition, the discussion of a group of pictures task required little interaction or exchange of information, meaning that the interlocutor's role had little influence on the examinee's score.

McCarthy (2010) argues that although the majority of fluency studies have found SR and pausing phenomena to be reliable measures for a monologic task, recently there

have been an increasing focus on fluency conceptualization in interactive performance. McCarthy (2010) discusses two features of turn-taking in an ongoing conversation that may potentially affect both the creation and maintenance of fluency, namely, turn-opening and turn-closing. On the one hand, turn-openers "link and provide continuity with the immediately previous talk and can be seen as creating smooth transitions" (McCarthy, 2010, p.5) by using connections (e.g., but, and), reactions (e.g., laughter, aha, oh), and management items (e.g., well, ok, right). On the other hand, turn-closing "usually transfers to a new speaker at points" (McCarthy, 2010, p.7). An example of this is completing previous syntactic elements or intonation using adjectives such as *lovely, interesting, wonderful* in order to invite the other speaker to take their turn.

L2 researchers (e.g., Tauroza and Allison, 1990; Derwing et al., 2004; Sato, 2014; Witton-Davies, 2014; Nitta and Nakatsuhara, 2014; Tavakoli, 2016; Peltonen, 2017; Os et al., 2020) studied utterance fluency in dialogue tasks (see Table 1, p. 53-58) and they have indicated that L2 speakers are more fluent in dialogue than in monologues. For example, Tauroza and Allison (1990) confirmed the appropriateness of a dialogic task compared to a monologic task for measuring L2 fluency. Their findings indicated that measures of pruned SR (total number of syllables per minute excluding self-corrections, repetitions, false starts, or repairs) and unpruned SR (total number of syllables/60 including self-corrections, repetitions, false starts, or repairs) were faster in dialogue than in monologue. Tauroza and Allison (1990) compared the SR per minute for different monologic and dialogic tasks to assess listening difficulty in the following situations: conversations, interviews, radio programmes, and academic lectures of native British English speakers. The analysis of the recorded listening samples was akin to the previous findings which indicated that the SR of

native speakers was faster in dialogue and interview samples than in lecture and radio samples.

In Derwing et al.'s (2004) study, three tasks were used – narrative picture description, dialogue, and monologue – to assess the L2 fluency of low-proficiency participants. In the dialogue task, the students were given time to think and respond to the following question: "Talk about the happiest moment in your life." They were also directed to ask the researcher the same question to make the conversation more natural. This type of discussion is called instructed conversation. The results of the native speakers' judgements indicated that the participants were more fluent in monologue and dialogue tasks than in the narrative description task. The ratings for the monologue and dialogue tasks were the same and this was because the same topic was used in both modes (practice effect).

Nitta and Nakatsuhara (2014) claimed that traditionally, pre-task planning was frequently used in monologue and neglected in dialogue. They carried out a study to examine the effect of pre-task planning (planned and unplanned conditions) on two decision-making tasks. Speech samples of thirty-two Japanese students were analyzed in term of conversation analysis and global fluency as measured by speed, breakdown, and repair. The students' opinions were collected in the form of post-test questionnaires to understand the relationship between pre-task planning/unplanning conditions and oral interactions. The data from Nitta and Nakatsuhara's (2014) study indicated that the planning conditions had little effect on the participants' interactions. The results of the questionnaires did not show any significant differences between the two planning conditions. Interestingly, these results showed that the participants preferred to speak without pre-planning time because long pre-task planning may

prevent students from interacting collaboratively in the task and decrease their SR. This is because they might focus on trying to remember their previously planned speech, which causes dysfluency. As for breakdown fluency and turn length in planned speech, there was a small improvement in the form of longer turns and fewer pauses. In contrast, unplanned speech, the test takers' performances contained more pauses and short turns at the beginning of their talk to gather information or to plan the interlocutor role. However, their interaction was more balanced and co-constructed than the test takers in the planned condition.

In his longitudinal study, Witton-Davies (2014) investigated the development of English oral fluency in narrative monologue and discussion dialogue for a group of L1 Taiwanese students over four years. Several fluency measures were found to be effective in developing students' oral proficiency, such as pruned SR and AR. In addition, these two measures were considered to be better predictors of fluency than composite measures. In addition, pause location, pause frequency, and mid-clause pauses contributed more to L2 fluency than the pause time. Repairs in this study were analyzed separately because each measure behaved differently. The results from the utterance fluency measures across the two modes revealed that dialogue performance outperformed monologue performance, seen in smoother and less complex production, over the four years abroad.

Similarly, Tavakoli (2016) examined the differences in L2 oral fluency performance in two modes: monologue and dialogue. Thirty-five L2 participants with a variety of L1s performed two tasks: a discussion dialogue and a narrative retelling of a personal experience. A number of common fluency measures were used, such as the length of pauses, number of repairs, phonation time and ratio, AR, and SR. In addition, two

dialogue-only measures were used, namely the number of interruptions and turns. The findings of the study were in line with Witton-Davies's (2014) study, which indicated that fluency in dialogue was higher than in monologue; dialogue was found to be faster in ARs, longer in runs, and shorter in pauses than monologue. In addition, Tavakoli reported that there were few interruptions (e.g., yeah, mmm) between turns and if any interruption was found, it was because of the participants' involvement in the interaction during the dialogue. However, in Tavakoli's study, unclaimed pauses between turns were divided equally between two speakers because the qualitative analysis of the between-turn pauses indicated that the speaker either paused for a long time, after completing the clause, or the other speaker remained silent because s/he was not sure of whether or not the other partner had finished his/her turn. It could be argued that between-turn pauses were because both speakers were not yet ready to speak.

Peltonen (2017) used quantitative and qualitative data (mixed methods) to examine L2 oral dialogue fluency in a problem-solving task for forty-two Finnish learners. The participants were lower secondary school (Group 1) and upper secondary school (Group 2). The study aimed to investigate how well the participants worked together to manage their own turns as a speaker and an interlocutor in order to avoid long silences between turns. The results from comparing the L2 fluency of each group showed that the Group 2 participants' SRs, SPs and mean length of silent pauses were more statically different from those of the Group 1 participants. The overall dialogic performances of the Group 2 participants were more fluent with fewer and shorter turn pauses than the performance of the Group 1 participants. Peltonen (2017) explained that during the interaction in the oral tasks, high-proficiency students were found to use fillers, repetitions, and FPs (e.g., stalling mechanisms) to maintain their

turns and these were correlated significantly with other temporal fluency measures. Thus, the qualitative analysis indicated that the fluent speakers used stalling mechanisms (e.g., repetitions) and communication strategies to compensate for speech dysfluencies in dialogue.

More recently Os et al., (2020) investigated whether turn-taking behaviour as a characteristic of dialogue influenced the perceived fluency ratings of native and non-native speech. Also, the effects of SR (fast or slow), dialogue gaps and overlapping between the question and answer turns of native and non-native speakers of Dutch were examined. The raters listened to short dialogues and rated the fluency of the speaker's answer on a 9-point scale. The results showed that in native speech, fast or too eager answers and slow or too reluctant answers negatively affected fluency perception. Whereas, in non-native speech, only slow or too reluctant answers affected the fluency rating. Furthermore, the speech of the fast speakers was rated as more fluent than the speech of the slow speakers. As for overlaps and gaps, in the fast speech of the native speakers, overlapping was rated as less fluent than gaps, while in the slow speech of both native and non-native speakers, overlapping was rated as more fluent than gaps.

Additionally, monologue and dialogue have been studied in task-based language research in terms of complexity, accuracy and fluency (CAF) (see Table 1, p.53-58). For example, Michel et al. (2007) studied the effect of the Robinson's cognition hypothesis (2005) on task CAF of L2 performance in monologue and dialogue. Forty-four L2 Dutch learners conducted simple and complex monologue and dialogue tasks. Complexity was measured by lexical complexity and accuracy was measured by repairs (reformations, repetitions, and false starts). Fluency was measured by

un/pruned speech, SR, and the number of FPs. The results suggest that in the light of the cognition hypothesis, task complexity and task conditions have a positive effect on CAF, meaning that the L2 complex dialogues and monologues were more accurate but less fluent than the simple dialogues and monologues. In terms of accuracy and fluency, L2 simple dialogues were more accurate and fluent than monologues. Thus, task complexity and task conditions only affected the accuracy measures in monologues. The results support to some extent the cognitive hypothesis theory.

Gilabert et al. (2011) examined the tasks manipulation of cognitive demands across three simple and three complex dialogic and monologic tasks. The oral performance of forty L1 Spanish learners of L2 English was examined in terms of CAF. For fluency, speed measures were used, such as syllables/60 pruned/unpruned speech/60, and FPs. Complexity was measured by structural and lexical complexity, whereas accuracy was measured by the number of errors/100 words. Then, all of the participants received a questionnaire about proficiency, motivation, task difficulty, stress, and confidence. Each participant performed three complex tasks and three simple tasks in monologue and dialogue. Correlation was used to examine the impact of CAF and proficiency on oral performance in both modes.

The results showed a strong correlation between L2 learners' CAF and their proficiency in monologue, whereas a weak correlation was found in the dialogue task. In terms of fluency, there was a correlation between L2 speed fluency and L2 proficiency. This means that participants with higher L2 proficiency were faster speakers. As for task complexity, the findings were in line with those of Michel et al. (2007) – increasing cognitive complexity resulted in increasing complexity and accuracy in monologue and decreasing fluency. In dialogue, the trade-off effect of

increasing cognitive complexity did not affect CAF. This might be because in the monologic task one person controlled the whole speech, whereas in the dialogic task each participant was given a portion of the story and both speakers had to work as a pair to reconstruct the whole story. Also, fluency was studied along with other accuracy and complexity dimensions and not as a single construct. In the present study, fluency is considered alone as a one dimension in order to examine the oral production of monologue and dialogue.

In another study by Michel (2011), the global measures of CAF of L1 and L2 speakers were studied in monologue and dialogue. Forty-six Turkish speakers learning Dutch as their L2 conducted +/- simple and complex monolgic and dialogic tasks. Their performances were also compared to 44 native speakers of Dutch and dnalyzed in terms of CAF. The results showed that L1 and L2 oral dialogues were more accurate, complex and fluent than L1 and L2 oral monolgues. Thus, the results showed that dialogues samples were faster than monologues, and the number of pauses and repairs were lower in dialogues than in monologues.

According to Michel (2011), the opportunities to improve language learning in monologue are fewer than in dialogue. She argues that there is less intake of new information and feedback in monologue compared to dialogue. This may cause a lack of attention to form and meaning and high pressure on attentional resources. In monologue, for example, the speaker depends on his/her own cognitive resources and knowledge. Furthermore, modified speech in monologue is only produced by self-monitoring. This means that during the speech production processing in monologue, the preverbal message is planned and conceptualized by a single speaker.

On the other hand, in dialogue, the speaker's own speech is planned and conceptualized during the other interlocutor's turn (e.g., Tavakoli and Foster, 2008; Michel, 2011). Furthermore, a dialogue task can be eased and simplified by aligment. Michel (2011) argues that the alignment process seems to free up more attentional resources for L2 speakers because they have more time to carry out the online planning of the speech. They therefore have more resources available for their actual talk when it is their turn. In dialogue, speakers could interrupt each other's talk, asking for clarifications. They may also copy each other syntactically and lexically during pair work. Considering all the above, Michel (2011) argues that this leads to more fluent and less complex speech in dialogue compared to monologue.

In another study, Tavakoli (2018) examined the development of L2 proficiency by comparing the performance of forty L2 participants in monologue and dialogue during their study abroad. Their oral development was measured through complexity, accuracy, lexis, and fluency (CALF). The comparison of oral development in the two modes revealed that L2 oral development may differ by task condition. The L2 participants in the dialogic task produced longer AS units, subordinate verbs, and accurate verbs, whereas in the monologic task the participants produced more accurate, longer, and less complex clauses. In monologue, the participants were able to control their production at a clause level, whereas in dialogue more complex verbs were used. As for fluency, SR and MLR in both modes showed statistically significant improvement over a short period abroad, whereas repair measures were not significant in either task condition (Kormos, 2006; Tavakoli, 2016). This might suggest that repairs are a matter of L1 personal speaking styles.

Table 1 below summarized fluency studies in four topics. First section of the table summarizes studies that used oral fluency in Task Based Language Teaching and working memory capacity. Second section introduces fluency studies that used monologic tasks. Third section introduces fluency studies that used dialogic tasks only. Fourth section introduces fluency studies in both monologic and dialogic tasks.

Table 1. Summary of the Different Utterance Fluency Measures Used in Previous Studies

Monologic Task Based Language Teaching (CAF)		
Mota (2003)	**Tasks:** 13 students conducted WM test and two oral tasks (picture and narrative) **Measures:** Speaking Span Test (SST) and complexity, accuracy, and fluency (CAF)	**Results:** high WMC was significantly correlated with CAF. Regression analysis shows that the scores of L2 participants' operation SST predicted the proportion of variations in the performance of CAF
Guara-Tavares (2008)	**Tasks:** 50 participants conducted WM test, two narrative tasks, and a retrospective interview. **Measures:** the relationship between pre-task planning, Reading Span Task and CAF was measured.	**Results:** high correlations between WMC Reading Span Task and L2 CAF. There was a statistical significant effect of pre-task planning on L2 speech complexity and accuracy, but not on fluency.
Kormos and Trebits (2011)	**Tasks:** 44 participants conducted two oral tasks (simple and complex), WM test. **Measures:** L1 Backward Digit Span task (BDS), and CAF measures.	**Results:** BDS has a small effect on L2 oral performance in both tasks. Scores of high WMC students correlated with the scores of the complex narrative task only.
Monologic Fluency Studies		
Lennon (1990)	**Tasks:** 4 EFL learners conducted 6 picture sequence tasks. **Measures:** trained EFL native speakers' judgements. Objective fluency measures: speed delivery (speech rate), MLR, filled/unfilled pauses, self-corrections, and repetitions	**Results:** words per minutes, position of filled pauses, and SR were strong indicators of L2 students' oral fluency development, while self-corrections or reformations were poor indicators of dysfluency.

Towell et al. (1996)	**Tasks:** 12 French learners conducted narrative tasks (retelling a story) in L1 French and L2 English. **Measures:** L2 temporal fluency measures (speech rate, phonation time and ratio, articulation rate, and mean length of run).	**Results:** students' L2 developed abroad because of syntactic complexity not because of the speed or numbers of the produced pauses.
Segalowitz and Freed (2004)	**Tasks:** 40 L1 English L2 Spanish participants conducted oral proficiency interview test and word recognition test. **Measures:** interactions between three variables: oral production abilities, learning contexts and L2-related cognitive abilities were examined in two L2 groups.	**Results:** a complex relationship between cognitive fluency, L2 fluency gain, and language contact. Study-abroad participants developed more temporal aspects of L2 utterance fluency than at home participants. There is also a strong correlation between fluency measures, and cognitive aspects tasks.
Iwashita et al. (2008)	**Tasks:** 5 speaking tasks from TOEFL test. **Measures:** 200 oral performance were measured by linguistics resources (grammar and vocabulary), fluency, and phonology measures. Rater judgment.	**Results:** the knowledge of vocabulary and oral production features (phonology and fluency) were strong predictors that influenced the raters' scoring decisions across all levels of the speaking tasks. The grammar complexity and accuracy were less significant predictors for overall speaking proficiency.
Derwing et al. (2009)	**Tasks:** 32 L1 (Slavic and Mandarin) and L2 (English) conducted a picture narrative task and a picture description task. **Measures:** temporal measures of fluency. Native speakers' judgments (8 raters).	**Results:** native raters' judgements indicated that there was a weak relationship between L1 fluency and L2 fluency. The correlation was stronger for the Slavic than for the Mandarin speakers of English. Fluency could not be transferred from L1 to L2.
de Jong et al. (2013)	**Tasks:** 179 L1 Dutch participants performed eight speaking tasks (e.g., lexical retrieval task and sentence completion task. **Measures:** L2 cognitive fluency measured by linguistic knowledge and processing speed. L2 utterance fluency performances as measured by speed, breakdown, and repair	**Results:** some measures of utterance fluency were good predictors of cognitive fluency. Linguistics skills were strongly correlated with the duration of syllables, whereas pause duration was weakly correlated with linguistic knowledge.

Leonard (2015)	**Tasks:** 39 participants received a pre-test/post-test design. A picture naming task and a sentence completion task. Grammar and vocabulary task. **Measures:** investigated the relationship between (1) L2 linguistics knowledge; (2) L2 cognitive processing speed; (3) L1/L2 utterance fluency as measured by speed, breakdown, and repair.	**Results:** no significant gains regarding the measurements of L1/L2 filled/unfilled pauses or hesitations. These measures are not related to L2 proficiency but to speakers' speaking styles because some learners use either FPs or SPs and sometimes both when they face problems with formulating the intended message.
de Jong (2016)	**Task:** oral performance of 18 L1 Dutch speakers and 52 L2 English speakers. **Measures:** Silent and filled pauses location and duration in L1 and L2.	**Results:** L2 speakers paused within utterances more frequently and for longer than L1 speakers, who tended to pause between clauses, and at utterance boundaries.
Peltonen (2018)	**Task:** 42 L1 Finnish and L2 English students conducted a picture description task. **Measures:** relationship between L1 and L2 fluency in the monologic speech. Stalling mechanisms.	**Results:** a high correlation between most L1 and L2 fluency measures (e.g., SR, MLR, mid-clause SPs, FPs, and mean length of end-clause SPs). Stalling mechanisms are related to individual speakers' styles
Zuniga and Simard (2019)	**Tasks:** 58 L1 French participants spoke L2 English carried out two narrative picture tasks (English and French). **Measures:** self-repair, English proficiency and attentional control were tested.	**Results:** a strong correlation between L1 and L2 production in terms of self-repair. Regression analyses showed that L2 self-repair behaviour explained 40% of the variance, whereas L2 proficiency made very little contribution to L2 self-repair behaviour. Repair fluency is an underlying trait.
Duran-Karaoz and Tavakoli (2020)	**Tasks:** 44 L1 Turkish participants spoke L2 English carried out two oral narrative tasks. **Measures:** Examining relationship between L1 and L2 fluency, and fluency behaviour in terms of speed, breakdown, and repair across different proficiency levels.	**Results:** there was a relationship between L1 and L2 fluency in terms of mid-clause FPs and mid-clause SPs, speech rate, and total number of repairs. Regression analyses suggested that some measures of L2 speed and breakdown fluency measures could not be predicted by L1 speed and breakdown fluency.
Dialogic Fluency Studies		
Nitta and Nakatsuhara (2014)	**Tasks:** 32 L2 Japanese students conducted two decision-making dialogic tasks.	**Results:** the planning conditions had little effect on the participants' interactions. Questionnaires did not show any significant differences

	Measures: pre-task planning (planned and unplanned conditions). Global fluency measures (speed, breakdown, and repair).	between the two planning conditions. Participants preferred to speak without pre-planning time because long pre-task planning may prevent students from interacting collaboratively in the task and decrease their SR.
Peltonen (2017)	**Tasks:** 42 Finnish carried out L2 oral dialogue fluency in a problem-solving task. **Measures:** quantitative (fillers, repetitions, FPs, SR) and qualitative data (stalling mechanisms, communication strategies) were collected from all 42 participants.	**Results:** higher proficiency (Group 1) participants' SRs, SPs and mean length of silent pauses were more statically different from those of the lower proficiency participants (Group 2). The overall dialogic performances of the Group 2 participants were more fluent with fewer and shorter turn pauses than the performance of the Group 1 participants
Os et al., (2020)	**Task:** Conversations data were used in two experiments: (1) recordings of 10 native Dutch speakers, (2) 10 non-native Dutch speakers. **Measures:** perceived fluency ratings, turn-taking behaviour, effects of SR, dialogue gaps and overlaps.	**Results:** In native speech, fast or too eager answers and slow or too reluctant answers negatively affected fluency perception. Whereas, in non-native speech, only slow or too reluctant answers affected the fluency rating. In the fast speech of the native speakers, overlapping speech was rated as less fluent than gaps, while in the slow speech of both native and non-native speakers, overlapping was rated as more fluent than gaps.
Dialogic and Monologic Fluency Studies		
Tauroza and Allison (1990)	**Tasks:** 30 minutes of native English speakers' speech in conversations, interviews, radio programmes, and academic lectures were analysed separately. **Measures:** all speech categories were analysed in terms of pruned SR, repetitions, false starts, repairs, and unpruned SR	**Results:** measures of pruned SR (total number of syllables per minute excluding self-corrections, repetitions, false starts, or repairs) and unpruned SR (total number of syllables/60 including self-corrections, repetitions, false starts, or repairs) were faster in dialogue than in monologue
Derwing et al (2004)	**Task:** narrative picture description, dialogue, and monologue tasks. **Measures:** L2 utterance fluency measures and native speakers' judgments.	**Results:** native speakers' judgements indicated that the participants were more fluent in monologue and dialogue tasks than in the narrative description task. The ratings for the monologue and dialogue tasks were the same and this was because the same topic was used in both modes (practice effect).

Michel et al. (2007)	**Tasks**: 44 L2 Dutch learners conducted L2 simple and complex monologue and dialogue tasks. **Measures:** number of CAF were used (e.g., SR, false starts, and lexical complexity)	**Results:** the cognition hypothesis, task complexity and task conditions have a positive effect on CAF. L2 complex dialogues and monologues were more accurate but less fluent than the simple dialogues and monologues. L2 simple dialogues were more accurate and fluent than monologues.
Gilabert et al. (2011)	**Tasks:** 40 learners carried out three simple and three complex dialogic and monologic tasks. **Measures:** measures of complexity, accuracy and fluency. L2 proficiency test and questionnaire.	**Results:** a strong correlation between L2 learners' CAF and their proficiency in monologue, whereas a weak correlation was found in the dialogue task. A positive correlation between L2 speed fluency and L2 proficiency. Increasing cognitive complexity resulted in increasing complexity and accuracy in monologue and decreasing fluency.
Michel (2011)	**Tasks:** 40 L1 Turkish L2 Dutch conducted +/- simple and complex L1 and L2 monologue and dialogue tasks. **Measures:** L1 and L2 global measures of CAF	**Results:** L1 and L2 oral dialogues were more accurate, complex and fluent than L1 and L2 oral monolgues. Dialogues samples were faster than monologues, and the number of pauses and repairs were fewer in dialogues than in monologues.
Sato (2014)	**Tasks**: 56 participants conducted tasks in pairs and in individuals' performances. **Measures:** perceived fluency (raters' judgements) was measured in terms of SR, pruned/unpruned speech, pauses, turn-takings, and backchannels.	**Results:** L2 fluent speakers used backchannels and turn-takings in natural patterns during peer interaction speech, whereas less fluent speakers were hesitant to initiate turns. Raters' judgements indicated the participants' SRs were higher in the pair-interaction task than in the monologue task.
Witton-Davies (2014)	**Tasks:** 17 participants carried out narrative monologue and discussion dialogue. **Measures:** SR, AR, pause location, pause frequency, mid-clause pauses, and turn pauses.	**Results:** fluency measures across the two modes revealed that dialogue performance outperformed monologue performance, seen in smoother and less complex production, over the four years abroad.
Kirk (2016)	**Tasks:** two L2 Japanese learners conducted monologic and dialogic tasks. **Measures:** mixed methods: quantitative (SR, AR, length of SPs) and qualitative (interactive	**Results:** dialogue was more fluent than monologue in SR, AR, and length of SPs. Qualitative analysis showed that speech in dialogue was facilitated by alignment between speaker and listener. This can be seen in gaze, gestures, and head movements.

	alignment model) analyses of dialogue and monologue fluency.	
Tavakoli (2016)	**Tasks:** Thirty-five L2 participants performed two tasks: a discussion dialogue and a narrative retelling of a personal experience **Measures:** the length of pauses, number of repairs, phonation time and ratio, AR, SR, and dialogue-only measures.	**Results:** fluency in dialogue was higher than in monologue; dialogue was found to be faster in ARs, longer in runs, and shorter in pauses than monologue.
Tavakoli (2018)	**Tasks:** 40 L2 participants conducted monologue and dialogue tasks. **Measures:** L2 proficiency development measured by complexity, accuracy, lexis, and fluency.	**Results:** the L2 participants in the dialogic task produced longer AS units, subordinate verbs, and accurate verbs, whereas in the monologic task the participants produced more accurate, longer, and less complex clauses. SR and MLR in both modes showed statistically significant improvement over a short period abroad.

2.7 Working Memory

2.7.1 Working Memory Capacity (WMC)

Working memory is a limited capacity mental system or workplace that is responsible for temporarily storing, maintaining, and manipulating information that is necessary for the completion of complex activities (Baddeley, 2001). Kormos and Sáfár (2008) indicated that there were individual differences (IDs) in the WM system and that this may lead to variation in learning skills, comprehension, and production abilities, namely L2 writing, reading, speaking, and vocabulary. WM is assumed to be the link between humans' actions and their perceptions and between long-term memory (LTM) and semantics (Segalowitz, 2010). Wen (2012) defines WM as "the limited capacity of multiple mechanisms and processes in the service of complex L2 activities or tasks"

(p.10). To further understand the role of WM, the next section introduces the WM model, and discusses its hypothesized relationship with L2 fluency.

2.7.2 WM Model

Baddeley and Hitch (1974) and Baddeley (2003) developed the most influential theoretical model for WM and it contains four elements (see Figure 1): (1) the central executive, which is the system responsible for attentional capacity control; (2) the visuospatial sketchpad, which is the storage of visual and spatial information; (3) the phonological loop, which stores verbal materials; and (4) a recently added element, episodic buffer, which "allows the various components to interact and enables their content to become available to conscious awareness" (Baddeley, 2017, p.299). This means that the episodic buffer links or binds all components of the WM to be stored in the LTM. Some of the WM model components (e.g., the central executive) are associated with the acquisition and learning of both first and second languages (Baddeley, 2012). Thus, the four WM elements explained below highlight the important role that WM plays in improving L2 processing and performance.

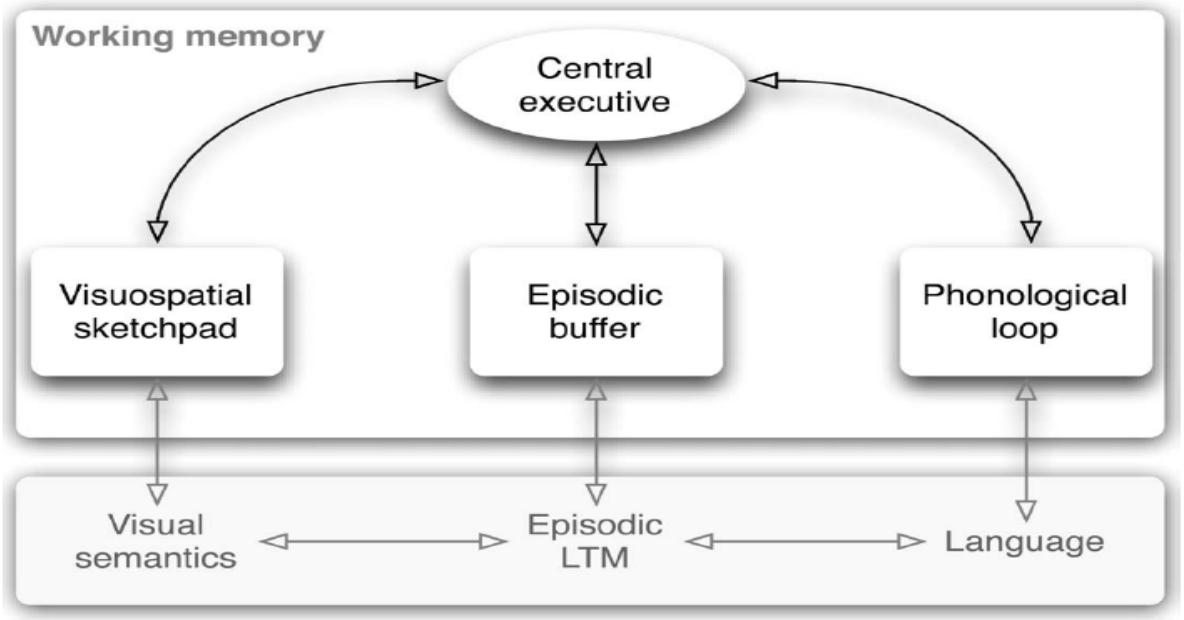


Figure 1. The Four Components of WM Model Presented by Baddeley (2003, p.835).

I. Central Executive

The central executive is the core of WM and the most complex system to be examined in WM model (Baddeley, 2003). It is also known as a limited capacity attentional system able to operate, select, regulate, and control mental processes, including the cognitive processes of L1 mental activities and behaviours. During cognitive processing, the central executive is given the role of restraining unrelated irrelevant information. It is also able to maintain the flow of information to long-term WM (Baddeley, 1993; Baddeley, 2003). The central executive is in control of two systems: the visuospatial sketchpad and the phonological loop (Figure 1). In the older version of the WM model, the central executive was responsible for attentional cognitive processing and for storing information (Baddeley, 1993). Later empirical evidence suggested removing the storage function from the central executive system so that the fourth WM component, the episodic buffer, was in charge of storing information (Baddeley, 2015). There is still a need for further research on the central executive as it is the least understood WM component (Baddeley, 2003).

II. The Phonological Loop

The phonological loop is the most researched articulatory component in this WM model. It is the initial phonological component when one acquires a language for the first time (Baddeley, 2017). The phonological loop deals with verbal data and it has two stores, one for holding phonemes for a few seconds and the other for the articulatory rehearsal process, which smooths rehearsal of the saved phonemes (Baddeley, 2003). Furthermore, the phonological loop can maintain visual and auditory

information, encode them as oral data, and store them in the phonological store temporarily. The capacity of the phonological loop is measured by WM tests that involve the instant serial recall of items, digits, or words.

III. Visuospatial Sketchpad

The visuospatial sketchpad has two separate sub-components: one is for visual information and the other is for dealing with spatial information. The visuospatial sketchpad is able to receive and store spatial and visual information (Baddeley, 2003). It has limited capacity and can store only three to four elements such as characteristics related to shape, colour, and location. For example, the visuospatial sketchpad can store visual data and manipulate its locations, characteristics, and movements within the related context (Baddeley, 2000). In SLA research, the visuospatial sketchpad is assumed to be responsible for learning and acquiring the syntactic features of new words by connecting the meanings of these new words with their related visual stimulus (Baddeley, 2003).

IV. Episodic Buffer

The episodic buffer is the latest component to be added to the WM model in order to compensate for the deficiency of removing the storing function from the central executive component (Baddeley, 2000). It is controlled by the central executive system. The episodic buffer is like a hub that receives and combines all the information from the visuospatial sketchpad and phonological loop (Baddeley, 2015) and associate it with further information from LTM and perception. Then, the episodic buffer combines all the information into unified episodes that can be accessed consciously. The episodic buffer is considered to be a passive store that cannot carry out any further processing on the episodes, only hold them (ibid).

Thus, it can be seen from Baddeley's model (Figure 1) that WM components are to a large extent related to language learning and processing. Therefore, the current study adopted Baddeley's model as the theoretical framework for examining WMC. The next section will discuss studies that examined the relationship between WM and L2 learning and processing.

2.7 Working Memory and L2 Language Processing and Performance

Ellis (2015) believes that WM plays a significant role in L2 acquisition as a mental construct that is responsible for the regulation of attention, restoration, and perception. The variation between declarative knowledge and procedural knowledge is widely known in the field of cognitive psychology and SLA (Towell et al., 1996). de Jong and Perfetti (2011) discuss the fact that fluency development encompasses two constructs: automatic processing and procedural knowledge. They also argue that procedural knowledge differs from declarative knowledge in its fast equal processing and lower dependence on WM resources; therefore, it is important in fluent speech. In contrast, declarative knowledge needs comprehensive, syntactic, phonological, and semantic knowledge. It also depends on cognitive resources and attention. This is in line with the previously mentioned models of speech production by Kormos (2006) and Levelt (1989), i.e., that the procedural knowledge of both L1 and advanced L2 speakers is automatically employed in the speech production phases of formulation and articulation. In contrast, lower level of L2 proficiency speakers need an additional store of L2 knowledge that depends on declarative knowledge in the conceptualization and formulation stage (i.e., the lexical retrieval of syntactic information and grammar rules) (Kormos, 2006).

It is important to examine how Levelt's (1989) model of speech production explains the fluent speech of native speakers. Also, it is an essential theoretical topic that helps us to understand the roles assigned to the components of WM by L1 and L2 speech production models (Mizera, 2006). Starting with Levelt's (1989) model of first language production, the second phase of speech production processing is formulation, where L2 linguistic and syntactic information is retrieved from the LTM. This access to linguistic information is a controlled process that is regulated by the central executive WM (Baddeley, 2003). The WM allocates and directs speakers' attentional resources to prepare form and meaning through the central executive.

Moving on to the L2 production model of de Bot (1992), the WM construct can serve as a basis from which to understand L2 processing. It can be argued that individual differences in WMC account for the variations in L2 production tasks (Payne and Whitney, 2002). As mentioned previously, in de Bot's (1992) L2 speech model, L2 production is a controlled process in the most important processes: the formulation (lexical access) and articulation stages. This controlled processing implies the use of WM, meaning that limited WMC has a large effect on L2 acquisition and therefore L2 performance. According to Kormos (2006), L2 speech production is a challenging and a sequential processing due to the limited access to the mental lexicon. The WM is expected to be heavily loaded by the demands to distribute attentional and memory resources to the particular stages of speech production (Skehan, 2015). Individual differences in cognitive capacity include WM, attentional resources, and processing abilities (Cho, 2017).

Working memory is a limited capacity cognitive system that is not expansible, meaning that more challenging tasks are believed to consume both memory and attentional

resources. This affects the storage capacity and consequently the speed of processing (Skehan, 2014b). The limited capacity of WMC has an implication for oral language performance as it limits the amount of data available for processing (Kormos, 2006). Speaking is a complex cognitive task that requires storage and processing in the various stages of speech production (Fortkamp, 2000). In the conceptualization stage, learners must temporarily hold their conceptual chunks in the WM while diverting attention to other cognitive operations, such as accessing the mental lexicon, providing phonological and morphological forms, monitoring the whole operation, and activating the required lemma for future processing (Payne and Ross, 2005). Therefore, in relation to this study, it is hypothesized that a high WMC allows more attentional resources to be freed up to retrieve materials from the long-term working memory and to process L2 speaking (Weissheimer, 2011; Baddeley, 2012).

2.8 WM Measures

How to measure the capacity of WM is still open to debate due to the lack of reliable and valid WMC tests (Mitchell et al., 2015). There is no straightforward test for WMC and researchers have not yet reached a consensus on a single test that can be used to assess the impact of WM on different aspects of L2 learning, including L2 utterance fluency. Furthermore, the tests used to measure WMC are varied, and each test has a function that taps into one aspect of the broad construct of WM.

There are two test complexities when measuring WM. First, simple WM tests are used to measure short-term memory (STM) while complex tests are used for long-term memory (LTM). Simple WM tests measure phonological loop capacity such as the simple span task (SST), which requires instant recall of series of words, digits, or visuals after a short written or oral presentation. In SST, the test takers should have

sufficient WM storage and language knowledge (Kormos and Trebits, 2011; Juffs and Harrington, 2011).

Several WM span tasks have been designed to assess different capacities of the WM components (i.e., storage and processing) (Daneman and Carpenter, 1980). On the one hand, in order to examine test takers' central executive of WM, storage and attentional control should both be measured by one of the complex WM tests, such as the Backward Digit Span, Operation Span, Counting Span, or Reading Span tasks. These are commonly used measures of WM in SLA and cognitive psychology (Conway et al., 2005). On the other hand, Listening Span and Speaking Span tasks are commonly used by cognitive psychology. Each task is explained briefly below:

The Operation Span Test (OST) (Turner and Engle, 1989) is a computer-based task that includes increasing sets of items. The test takers are required to read each arithmetic problem and decide if it is correct or incorrect. Then, the test takers are asked by the examiner to recall the final word after each mathematical problem. The OST focuses on the storage and processing of items. The OST does not require language comprehension. Detailed explanation of the OST is found in section 3.6.2.2.

The Counting Span Test (CST) (Case et al., 1982) is another computer-based task. The participants are introduced to different kinds of coloured shapes that are shown all together on a computer screen with other shapes of the same colours. The test takers are required to say out loud the number of similar coloured shapes in the right order. Visual distractors, such as similar shapes that have different colours, are introduced in this task.

With Digit Span tests there are two types: Forward-Digit Span (FDS) (Botwinick and Storandt, 1974) that measures only memory storage, and Backward Digit Span (BDS)

(Kormos and Safar, 2008) that taps into WM's temporary storage and processing. In BDS, the participants must recall the reversed version of the numbers that have been presented randomly in ascending order. Digit or non-word repetition task can avoid the prior language knowledge effect, which can happen when some participants lack knowledge of or familiarity with some L2 vocabulary. Alternatively, L1 can be used in word or number to avoid this confounding problem (Juffs and Harrington, 2011). Detailed explanation of the BDST is found in section 3.6.2.1.

The Reading Span Test (RST) (Daneman and Carpenter, 1980) is a computer-based task that includes increasing sets of items. The first set starts with two sentences and the final set ends with six sentences. The test takers are required to read each sentence out loud and decide if each sentence is grammatically correct or incorrect. Then, they are asked to recall the final word that appears after each sentence. WM span is the maximum number of items that are recalled in the right order.

The Listening Span Test (LST) (Daneman and Carpenter, 1980) is similar to the RST, but it is an oral version that requires listening instead of reading. Both the LST and RST require language comprehension to complete the task (Juffs and Harrington, 2011). LST is a computer-based task that includes increasing sets of items. The first set starts with two sentences and the final set ends with six sentences. The test takers are required to listen to each sentence and complete a written verification question that is related to the content of the sentence the participants have just heard. Then, they recall the final word after each sentence. The participants' listening span is the maximum number of items that they can correctly recall.

In the Speaking Span Test (SST) (Daneman and Green, 1986), the participants read silently increasing sets of words that appear one at a time on a computer screen. The

participants are then required to produce a grammatically correct sentence for each vocabulary that appears at the end of each sentence.

From the above, it can be concluded that all verbal WM span tests contain both WM processing and storage. The tasks are completed individually by introducing random gradually increasing sets of items in what is known as a processing task. This may take the form of arithmetic problems, problem solving, or sentence reading. Then the processing task is followed by spoken or written items that are to be recalled at the end of each set. This is called a stimulus and can take the form of words, numbers, or letters. Finally, the test takers' WM spans are the maximum number of items that are recalled in the right order.

Results are mixed and inconsistent across all WM studies with regard to the relationship between WMC and improvement in L2 oral fluency. For instance, L2 researchers used a number of common L1 and L2 WM tasks, such as the RST and LST (Daneman and Carpenter, 1980). In this test, the participants read or listen to an increasing number of sentences and they have to recall the final words that appear after each sentence. Studies like those by de Bot (1992), Kormos (2006) and Kormos and Safar (2008) used RST and they found a high correlation between high WMC scores and fast accurate speech rate (SR).

On the other hand, other studies like those by Mizera (2006) and Gilabert and Muñoz (2010) found no relationship between WMC and oral fluency. For example, Kormos and Dénes (2004 examined whether the individual differences in WMC Reading Span Test of fifty-nine Spanish participants explained their different oral performances in fluency, accuracy, and complexity. Fluency was measured by unpruned SR and the number of syllables per minute. The results revealed a low positive correlation

between the participants' L1 RST, lexical complexity ($r = .24$, $p < .05$), and fluency ($r = .23$, $p < .05$) but no correlation was found between RST and the participants' accuracy. Next section will shed light on more results on WMC and fluency studies.

2.9 Working Memory and Second Language Fluency Studies

Working memory is a widely studied cognitive variable among L2 learners because it is capable to store, process, and retrieve linguistics information (Cho, 2017). Temple (1997) explains that WM is more important for beginner and intermediate level L2 proficiency speakers and less important for L2 advanced speakers because fluent advanced speakers depend less on the WM store. In addition, WM resources need to be used by both fluent and less fluent speakers. For example, less fluent speakers use the attentional WM resources while speaking to choose the appropriate lexical and phonological items to complete their utterances. These attentional resources are also used to check the right syntax or the suitable words in the speakers' mental lexicon (Payne and Whitney, 2002). Working memory is even more important in the processing of L2 than L1 because of the lack of automatization and the conscious controlled nature of L2 processing (Kormos, 2006; Weissheimer and Mota, 2009).

Segalowitz (2010, 2016) recommends that IDs like L1, WMC, aptitude, age, and motivation should be considered along with cognitive processing to uncover possible dysfluency reasons in L2 speech. There is a general tendency in SLA research to believe that individuals differ in their cognitive processes, especially in how they produce fluent, smooth speech. This means that there are differences in the attention they use to fill the WM with the information they should remember (Cowan, 2010). It may be the case that these variations are related to the limitations in the WM that help to regulate the flow of information (Weissheimer and Mota, 2009). Segalowitz (2016)

claims that investigating a person's characteristics from cognitive perspectives alone is not enough; L2 utterance fluency measures should also be considered, meaning that sometimes dysfluency can be L2 specific. For example, some people are slower in their L2 lexical retrieval than in their L1.

Daneman (1991) and Baddeley (2017) argue that WM can differentiate between high/low WMC individuals in L2 acquisition, L2 performance, and the performance of tasks that involve measurements of fluency. Furthermore, WMC plays an important role in establishing word-meaning links. It is also one of the IDs that is hypothesized to affect L2 learners' speech fluency because insufficient word-meaning links could naturally slow down speech production processing, increase short memory workload, and reduce speakers' SR (Baddeley, 2017). While fluent L2 learners have enough mental resources to enable them to draw small chunks from declarative knowledge or to plan for alternative procedures to avoid errors or dysfluencies (Kormos, 2006), L2 users need not only word-meaning links but also lexical access to create a message and to deal with grammatical and phonological encodings of the sentences for final processing (Segalowitz, 2010).

In the SLA literature, WMC is usually tested in relation to pre-task planning and its effect on the complexity, accuracy and fluency (CAF) of L2 speech production. For instance, a study was conducted by Mota (2003) to prove that high WMC facilitates L2 speech production in complex tasks. The WMC of thirteen L2 students was examined by conducting SST. In addition, the participants' L2 oral speech was elicited in relation to three aspects: complexity, accuracy, and fluency across two oral tasks (picture and narrative). The results showed that high WMC was significantly correlated with CAF. In addition, the results of the regression analyses indicated that the scores

of L2 participants' operation SST predicted the proportion of variations in the performance of CAF. Thus, WMC is found to be a significant predictor of CAF.

A similar study was conducted by Guara-Tavares (2008), found high correlations between the measures of the WMC reading span task and L2 CAF. The study confirmed the significant role of planning time in helping students to retrieve complex speech and providing more time for L2 self-monitoring.

Furthermore, Kormos and Trebits (2011) investigated the performances of forty-four L1 Hungarians in two L2 oral English tasks. Furthermore, a correlation test between the participants' scores in the L1 WMC test and their L2 oral performances was performed. WMC was measured by an L1 backward digit span task (BDS), and the L2 oral performances in both simple and narrative tasks were examined in terms of fluency, complexity, and accuracy. The participants were divided into a low WMC group and a high WM group based on their scores in the BDS. The results revealed that WMC has a small effect on L2 oral performance. The results further indicated that the scores of high WMC students correlated with the scores of the complex narrative task only. In addition, participants with high WMC produced longer complex clauses than lower WMC participants, who produced shorter clauses.

In addition, several correlational studies were conducted to explore the relationship between WM attentional capacity in processing language and students' cognitive processes in L2 production and development. For example, Fortkamp (1999) examined the relationship between WM as measured by the RST and SST, and oral fluency as measured by an objective rating (the number of words per speaking time) and a subjective rating (native speakers' judgements). Sixteen L2 students conducted WM tests in their L1 Portuguese and L2 English. To assess their L2 fluency, the

students completed three oral tasks, an oral slip task, a picture description, and an oral reading task. The first two tasks aimed to assess fluency at the discourse level while the third task aimed to assess fluency at the articulatory level.

The results showed that there was no relationship between L1 SST and all three oral fluency tasks, whereas there was a significant correlation between L2 SST and L2 picture description task. Another interesting finding was that L1 and L2 RST scores were correlated with the scores of the oral reading task, meaning that WM supported fluency at the articulatory level only. Taken together, the findings suggested that WMC here was task specific and it was related to each student's processing efficiency during the performance of the task. In addition, it was found that students' scores in L2 SST were more related to their L2 proficiency than to their WM.

In another study, Fortkamp (2000) examined the relationship between WMC as measured by the SST (Daneman, 1991), the operation word span test OWST (Turner and Engle, 1989), and L2 speech production. The objective of Fortkamp's (2000) study was to investigate whether this relationship is task specific or a general capacity that stays the same in all tasks. L2 speech production was measured by temporal variables of fluency, namely, speed, breakdown, and repair; and complexity, accuracy and lexical density. The results showed significant correlations between SST and fluency, accuracy, complexity and lexical density in narrative and descriptive tasks. The results further showed no correlation between the OWST and L2 speech production measures, again suggesting that WM here is task specific. The ceiling effects might be a reason for reaching this conclusion because the correlation test is sensitive to restricted or close-to-maximum scores and small sample size (thirteen participants).

Unlike Fortkamp (2000), Mizera (2006) considered an independent proficiency test to examine the proficiency level of forty-four L2 (Spanish) participants. The participants conducted three WMC tests, namely the SST, a non-word repetition task and a math span task. Mizera (2006) aimed to explore the relationship between participants' WMC and their L2 fluency performances in three speech tasks: word translation, comic strip, and an imitation test. In addition, seven fluency measures were used to tap into participants' abilities in lexical retrieval and monitoring. Overall, the results showed weak correlations between SST and two measures of fluency: repair and speed. Also, no correlations were found between any of the WMC tests and other measures of L2 fluency. This might be due to the small number of participants and their low proficiency level.

Kormos and Sáfár (2008) conducted a longitudinal study to examine the relationship between phonological short-term memory (PSTM), WM and developing L2 English use in listening, speaking, reading, and writing among 121 L1 Hungarian students abroad. Phonological short term-memory (PSTM) is measured by BDS and non-word span test. Both measures of WM were correlated with the scores of the independent proficiency test. Only the intermediate students' scores for PSTM were correlated with their scores for L2 English use in writing, and their overall proficiency. In contrast, a moderate correlation was found between beginner students' WM scores and their L2 speaking task scores. Kormos and Sáfár (2008) suggested that intermediate students had an advantage from learning abroad and that was revealed by the results of the non-word span task. In contrary, beginner students experienced more explicit learning, as illustrated by BDST.

As seen in the above WM studies, such as Fortkamp (2000) and Mizera (2006), there are two drawbacks to the use of WM tests: (1) the lack of an independent measure for the proficiency test; (2) WMC tests are used for students' L2 only, without considering their L1. Additionally, it can be argued that the individual differences in participants' L2 proficiency are to a large extent influenced by their scores in the WM tests. Thus, the current study considered conducting an independent proficiency test to indicate the participants' L2 level of proficiency.

More recently, Georgiadou and Roehr-Brackin (2017) studied the relationship between (1) utterance fluency as measured by speed and hesitations; (2) the behaviour of self-repair to PSTM and executive WM. The participants' PSTM was measured by L2 English word recall test (WRT), whereas their executive WM was measured by L1 Arabic BDST and L2 English Listening Span Test (LST). Seventy-seven L1 Arabic speakers learning English as their L2 completed interview tasks to examine their SR, hesitations, pauses, and self-repairs behaviour. The results showed that the intermediate participants with higher executive WM scores produced fewer pauses because their conceptualization and formulation stages were improved. In addition, there was no significant relationship between any of the WM tasks and self-repair, repetition, or SR measures of fluency.

Finally, Zalbidea (2017) examined the interactive and independent effects of task complexity (−/+) and task modality (speaking and writing) on L2 accuracy and the complexity of L2 production. She also investigated how the individual differences in WMC, as measured by the OST, were related to linguistic measures in L2 speaking and writing tasks. Thirty-two L2 Spanish participants completed two versions of less and more complex argumentative tasks in two modalities: speaking and writing. The

tasks were analyzed in terms of complexity (e.g., the perceived complexity questionnaires), accuracy (e.g., error-free sentences), and linguistics measures (e.g., conjunctions).

The results of the quantitative analyses revealed that task modality played a greater role than the complexity of the task in improving the linguistic performance of lower-intermediate participants. In the speaking tasks, the thirty-two L2 participants produced language that was more syntactically complex than in the output of the writing tasks, in which the language produced was lexically more complex and accurate. Finally, WMC had a robust influence on the linguistic measures in both writing and speaking tasks, but only when the tasks' cognitive demands were enhanced. Thus, a large WMC improved L2 learners' processing of the retrieved information, and as a result the quality of the L2 oral production improved. For example, they were able to produce subordinating conjunctions in argumentative speaking tasks, whereas the same L2 learners produced less subject agreement errors in the written tasks. This is because WM is related to accuracy and monitoring.

2.10 Contribution of the Study

Presently, there is a paucity of research examining dialogue fluency in non-native bilingual interlocutors who share the same L2. Research studies (e.g., Segalowitz, 2010) on L2 speech fluency recommend studying fluency in a dialogic context, considering the social aspect of fluency; however, only few empirical studies have considered the concept of dialogue fluency in their data (e.g., Sato, 2014; Witton-Davies, 2014; Tavakoli, 2016; Peltonen, 2017). Whereas previous literature on L2 fluency studies in monologue tasks and in peer interaction tasks (e.g., Mizera, 2006; Michel et al., 2007; Michel, 2011; Sato, 2014; Tavakoli, 2018) did not examine fluency

alone as the main object of the study, but instead fluency was analyzed with other components of L2 oral performance i.e., complexity and accuracy. The focus usually is on the effect of different task complexities and conditions on L2 learners' complexity, accuracy, and fluency (CAF) of oral performance. However, study multidimensional nature of fluency on its own, using a wide range of fluency measures and giving detailed explanations and justifications of choices of measures and decisions of analysis, would be beneficial for other SLA researchers. This is what the current study aims to achieve.

Additionally, L2 fluency has been studied in monologue and dialogue in several ways: (1) in different task types such as interviews, radio programmes, and academic lectures (e.g., Tauroza and Allison, 1990); (2) in all dimensions of complexity, accuracy and fluency CAF (e.g., Kormos and Trebits, 2011; Gilabert et al., 2011; Tavakoli, 2018); (3) in relation to native speakers' judgements (perceived fluency) (e.g., Derwing et al.,2004; Sato, 2014); (4) in L2 learners with different proficiency levels (e.g., Gilabert et al., 2011); (5) in L2 fluency development during study abroad period (e.g., Witton-Davies, 2014); and (6) in an interactive alignment model (e.g., Kirk, 2016). However, studying L2 fluency alone in two task conditions (e.g., monologue and dialogue) in terms of length and location of mid-clause silent/filled pauses and end clause silent/filled pauses, all repair measures, and all speed measures to our knowledge was not the focus of previous studies. Although Witton-Davies (2014) and Tavakoli (2016) used variety of fluency measures to compare dialogic and monologic performances, the number of participants was not large enough to reach statiscal reliable results. For example, Witton-Davies (2014) investigated the oral performance of only 17 participants abroad, and Tavakoli (2016) used the data of 35 L2 participants

with different L1s. Additionally, Witton-Davis used same topic in both task conditions. This might cause practice effect in the performance of both tasks. Thus, the aim of this study is to address these shortcomings by using large number of participants (64) with the same L1, and different topics for both task conditions to offer focused and reliable findings.

Furthermore, other studies have explored L2 utterance fluency as a single construct in a dialogic task only (1) in different planning conditions: planned and unplanned tasks) (e.g., Nitta and Nakatsuhara, 2014); (2) in relation to stalling mechanisms and communication strategies (e.g., Peltonen, 2017); and (3) in relation to L2 learners' perceived fluency in dialogue (e.g., Os et al., 2020). The above-mentioned studies have focused on a few aspects of L2 utterance fluency measures (i.e., speech rate, pause frequency) and their findings are inconsistent, meaning that further investigations are needed. Thus, there is still a need for research on the operationalization of utterance fluency alone in both task modes (monologue and dialogue) in ESL and EFL classrooms to have a broad picture of L2 oral performance in dialogic and monologic tasks. The lack of research in this area is because the measures of the temporal variables of fluency are time consuming when they are calculated in seconds using PRAAT software.

There is also a lack of systematic research examining the other factors that could contribute to dysfluent L2 speech performance, learning and processing in oral tasks (Skehan, 2015b). For example, working memory capacity (WMC), first language, task type, task familiarity/content, L2 proficiency, aptitude, personality, age, and interlocutor influence are among the factors that could influence L2 oral performance. For instance, individual differences (IDs) in WMC play a significant role in L2 learning

and processing. These IDs also influence L2 development and performance (Skehan, 2015). L2 speech production processes generally used to communicate between people are not always effortless or smooth (Felker et al., 2019). When speakers have difficulties with formulating or articulating the stream of words, the speech production is disrupted and punctuated by disfluency markers such as fillers, repairs, or repetitions (Felker et al., 2019). This could be associated with limited WMC, which is responsible for allocating and directing speakers' attentional resources in L2 processing stages such as conceptualization, formulation, monitoring, and articulation (Skehan, 2014b). Moreover, L2 speech processes are not automatized yet, especially in less advanced L2 proficiency speakers. L2 speech is a conscious processing that depends on WMC. Therefore, it can be argued that speakers' cognitive resources may become exhausted and limited when L2 learners' oral performance goes through several processes such as lexical retrieval, and grammatical and phonological encoding (Kormos, 2006). This can lead L2 speakers to use more pauses or repairs during the processes of formulation or monitoring to keep the flow of their speech.

L2 researchers (e.g., Tavares, 2008; Kormos and Sáfár, 2008; Gilabert and Munoz, 2010; Michel, 2011; Mojavezi and Ahmadian, 2014) have conducted different studies to examine the influence of WMC on L2 oral processing and production. However, there are still inconsistent results and a lack of agreement regarding the impact of WMC on L2 production and processing, especially in monologic and dialogic performances. Most of the above mentioned WM studies focused on all aspects of CAF in monologic tasks only not on oral fluency as a single construct in monologic and dialogic tasks. Moreover, the above mentioned studies have not dealt with L2 fluency in much detail in terms of all aspects of utterance fluency such as speed breakdown and repair. The second aim of the current study is to fill the gap of the

previous research by investigating the relationship between the individual differences in WMC and several aspects of L2 fluency performance in monologue and dialogue.

Thus, studying IDs in L2 learners' WMC and the differences in the performance of L2 monologue and dialogue is expected to offer data that could predict the outcomes of L2 oral monologic and dialogic tasks. The analysis of monologic and dialogic fluency measures aims to provide original contributions to the research of SLA and testing. Also, this study can be used as an example for future research where participants' L1 data should be considered to make the study more robust.Therefore, the aim of this study is to address the methodological shortcomings found in previous research by presenting a methodological approach to oral fluency in an individual and an interactional contexts.

Participants' speech performances are elicited via monologic and dialogic tasks where different topics are used to answer each task mode and to prevent practice effect. Monologues and dialogues are analysed quantitatively for a wide number of temporal fluency measures such as pause frequency, location and length, speech rate, articulation rate, reformation, false starts, and total repairs to study the differences between the two task modes (see Table 10, p.110). Two complex WM tests are used to examine the productive power of WMC in monologue and dialogue. The instructions of WMC tests and oral tasks are introduced in Arabic and English to ensure the participants' understanding of the tasks.

2.11 Conclusion

The above section has summarized the literature on L2 fluency, particularly looking at the theories of L1 and L2 speech production models, common definitions and aspects of L2 fluency, empirical studies of L2 fluency in task-based language performance, and monologic and dialogic tasks. Longitudinal studies were reviewed to investigate fluency development over a period of time. This chapter has highlighted the definitions of WM and the relationship between WMC and different aspects of L2 speaking performance, such as complexity, accuracy, and fluency (e.g., Mota, 2003; Payne and Ross, 2005). Finally, the chapter concluded with the original contribution of the current study.

Chapter 3: Methodology

3.1 Introduction

This chapter introduces the aims of the current research study, the research questions, the hypotheses, the study design, and the methods. First, the two main research questions and sub-questions are presented, along with the relevant hypotheses. Second, the employed study design and the chosen procedures are explained in detail with justifications for their choice. Third, a detailed description of the participants, instruments, tests, oral tasks and pilot study is provided. Fourth, the ethical considerations are reported, then the data collection procedure and data analysis are explained, followed by a discussion of the measures and the tools chosen for data coding and transcribing. Lastly, there is a brief explanation of the statistical analyses employed in the study.

3.2 Aim of the Study

This study aimed to examine the differences in L2 learners' oral performance in monologic and dialogic tasks in terms of L2 utterance fluency. Language oral performance was operationalised through measures of L2 utterance fluency: speed, breakdown, and repair (see Table 8). Section (2.6.3) demonstrated the importance of WMC in L2 speech processing because attentional resources depend on WMC. Thus, this study also aimed in an original way to examine the extent to which the predictive power of WMC, as measured by Backward Digit Span and Operation Span tests, can explain the variations in L2 utterance fluency performance in monologic and dialogic tasks.

3.3 Research Questions (RQ) and Hypotheses (H)

Research question 1: *Are there any differences between L2 learners' monologue and dialogue in terms of utterance fluency?*

This question was divided into three sub-questions and hypotheses in order to explore the differences between monologue and dialogue tasks in terms of the aspects of utterance fluency: speed, breakdown and repair fluency.

Research question 1a**:** *Are there any differences between L2 learners' monologue and dialogue in terms of speed fluency measures?*

Hypothesis 1a: *Dialogic performance will be faster than monologic performance.*

Recently, research studies by Segalowitz (2016) and Peltonen (2020) on L2 speech fluency have recommended studying the concept of fluency in the dialogic context, considering the social aspect of fluency. Based on previous L2 studies (e.g., Riggenbach, 1998; Michel et al, 2007; Michel, 2011; Sato, 2014; Witton-Davies, 2014; Kirk, 2016; Tavakoli, 2016, and Peltonen, 2017), it is hypothesised that dialogic performance will be faster than monologic performance because in dialogue both the speaker and the interlocutor share the responsibility of the talk. They can exchange turns and ask for clarification. In L2 monologic performance, the speech production processing is planned, formulated, and articulated by a single speaker; this could place pressure on attentional resources. By contrast, in dialogue, both speakers can plan, conceptualise, and formulate their speech during the other speaker's turn. This to some extent helps to reduce any unnecessary silences, pauses or hesitations during L2 speech production processing.

Research question 1b: *Are there any differences between L2 learners' monologue and dialogue in terms of breakdown fluency measures?*

Hypothesis1b: *It is hypothesised that dialogic performance will be associated with fewer and shorter mid-clause filled/silent pauses than monologic performance.*

As noted previously (see Hypothesis.1a), regarding L2 fluency in dialogue, both speakers are trying to reach a conceptual understanding of their speech and to maintain the flow and smoothness of the speech by filling the long silences (McCarthy, 2010). This could help to reduce unnecessary pauses to keep the flow of the speech.

Research question 1c*: Are there any differences between L2 learners' monologue and dialogue in terms of repair fluency measures?*

Hypothesis 1c**:** *Based on the interactive nature of dialogue in which both the speaker and their partner are sharing responsibility for the talk, it is hypothesised that L2 learners' oral performance in dialogue will have fewer repairs than in a monologue.*

The literature is divided regarding repair fluency in monologue and dialogue. There are mixed findings as some L2 researchers (e.g., Tavakoli et al., 2020) found that L2 dialogue was associated with fewer repairs than monologue, while others (e.g., Michel, 2011; Witton-Davies, 2014; Huensch and Tracy-Ventura, 2017) showed that there were no significant differences between monologue and dialogue in most repair measures. This area needs further study, considering all types of L1 and L2 language repair fluency measures (Zuniga and Simard, 2019). L2 researchers usually follow one of two approaches to measuring repair fluency. The first approach is measuring all kinds of repairs under one category, *'Repairs'* (e.g., Iwashita et al., 2008; Witton-Davis, 2014). The second approach is evaluating each measure in repair on its own, for example, self-correction, repetitions, false starts, and reformulation. Thus, it can be argued that the way of operationalising repair fluency measures may have an effect on the results. The current study follows both approaches of measuring fluency as a

one measure on its own and all kinds of repairs. This to increase the comparability with other studies.

Research question 2: *To what extent can L2 learners' WMC predict their utterance fluency in monologue and dialogue?*

Hypothesis 2: *Higher WMC learners will be more fluent than lower WMC learners in monologue and dialogue.*

Besides investigating the difference between monologic and dialogic performances in utterance fluency, the relationship between WMC and fluency across the two task modes will be measured. WM and individual difference in WMC have been regarded as strong predictors of complex cognitive activities, such as L2 speaking (Rosen and Engle, 1997; Ahmadian, 2012; Unsworth et al., 2013). This means that there are differences in the oral performances of high- and low-WM individuals. According to Rosen and Engle (1997), Amadian (2012) and Unsworth et al (2013), high WM participants were faster and more fluent than low-WM participants in retrieving verbal items in verbal fluency tasks (e.g., animals' names) from long-term WM. Thus, one can hypothesise that WMC would significantly predict L2 fluency performance because larger WMC allows speakers to store linguistic information and regulate the attentional resources responsible for speech production processing (Daneman, 1991). Thus, the current study has been guided by this prediction. Despite all this, other L2 researchers (e.g., Kormos and Safar, 2008; Michel, 2011; Awwad, 2017) have conducted different studies to examine the influence of WMC on L2 oral processing and production, and these have produced inconsistent results and a lack of agreement on the impact of WMC on L2 production and processing. To this end, this study aims to test this hypothesis in monologue and in dialogue, a first in L2 research.

3.4 Study Design

The design employed in this study is a within-participants repeated-measures design (Cohen et al., 2017; Rogers and Revesz, 2019) because tests, measures and tasks were undertaken with the same participants. In this study, the repeated measures are utterance fluency measures, which were taken for one group of L2 participants in two task conditions: L2 monologue and L2 dialogue. There are two independent variables each with two levels of operationalisation and one dependent variable with three levels of operationalisation. The first independent variable is L2 oral task performance with two levels (task modes): monologue and dialogue. The second independent variable is WM capacity represented by tests: Operation Span Test (OST) and Backward Digit Span (BDST), whereas the dependent variable is L2 utterance fluency measured by speed, breakdown and repair (see Table 10). By including all aspects of utterance fluency measures and two complex WMC tests, this research aims to capture the differences between L2 oral dialogic and monologic performances and to establish whether there are relationships between the L2 oral performances in the two task modes (monologue vs. dialogue) and WMC. Table 10 presents the design of the study and the independent and dependent variables.

Wilcoxon Signed Rank and Friedman tests were used to examine the differences between two task modes in terms of speed, breakdown and repair. Regression analyses and correlation tests were also employed to answer the second research question. Correlation research is a quantitative method used to test the existence of relationships between variables. This test can also make predictions of whether or not two or more variables are interrelated. This method does not measure cause-and-effect relationships (Mackey and Gass, 2005). The correlational analysis of this study looked at the variables of L2 utterance fluency in monologue and dialogue, and WMC

measured by the backward digit span test and the operation span test, and it evaluated their relationships in terms of direction and strength. A strong correlation is called a positive coefficient: '+1'. This happens if it is found that when a participant scores high on one variable, such as the WM test, she is highly likely to score high on another variable of L2 utterance fluency, such as L2 dialogic performance. In another example, a low WMC score and a low fluency score means that there is a strong positive correlation because both variables have gone down. In contrast, a negative coefficient of '-1' implies the opposite relationship between two variables. For example, a high WMC score and a low fluency score means that there is a strong negative correlation, because an increase in one variable means a decrease in the other variable and vice versa. The third coefficient is '0', implying no significant relationship between two variables – this means that they are unrelated (Dornyei, 2007).

Table 2. The Study Design and Variables

Study Design	Independent Variables	Dependent Variables
Repeated measures Within-participants ↓ One group N = 64	**1-Task conditions (2 levels):** a-Monologue b-Dialogue **2-Working Memory Capacity (2 Levels):** a- OST b-BDST	L2 Utterance Fluency in monologue and dialogue (3 levels): 1- Speed 2- Breakdown 3- Repair

3.5 Setting and Participants

The participants were 64 volunteer female students. They were third-year bachelor students at the Department of English Language and Translation at The University of Jeddah, Saudi Arabia. Their ages ranged from 20 to 22 years (M= 20.93 and SD= .78). The participants speak Arabic as their L1 and English as their L2. They had

studied English as L2 for eight years in elementary, secondary and university settings taught by L1 Arabic teachers. All of the participants were living in Saudi Arabia at the time of the study. Although all the L2 participants were from the same educational level (third-year bachelor students), their L2 proficiency level was tested to ensure homogeneity among the L2 participants (see section 3.6.1).

3.6 Tasks and Materials

In SLA research, individual differences, such as L2 proficiency, task types, WMC and gender, can be measured by two approaches. In one, researchers intentionally match and mismatch participants' age (e.g., adults, teenagers, children), WM (low, high), or aptitude (low, high) to the treatment variables in order to compare the scores in both groups after the treatment (Vatz et al., 2013). The second approach is to run a correlation test to find the relationship between the participants' measured individual differences and their performance in the proposed task. This may, for example, entail examining the relationship between individual differences in participants' WMC and their oral production (e.g., Kormos and Trebits, 2011; Wen, 2016). To answer RQ2, The current study adopted the latter approach by testing the relationship between variables, such as WMC and L2 fluency, in monologic and dialogic performances.

To maximise the potential of investigating the relationship between WMC and L2 fluency in the context of monologue and dialogue, besides other extraneous variables (age, nature of the task, L2 proficiency, WMC and gender), the following procedures were considered. First, L2 tasks have frequently been used in SLA research to examine: (1) task variables, such as familiarity, complexity, and structure or (2) task discourse, such as narrative, descriptive or opinion (Tavakoli and Skehan, 2005). L2 task is also used as a research tool to measure participants' performance and how it

varies from one task to another. For example, administrating an L2 task in an interactive or monologic context may affect their L2 complexity, accuracy or fluency (Ellis, 2017). In this study, the participants completed two L2 task types (monologue and dialogue) about popular topics in their country. Further details are provided later in this section.

Additionally, the participants' individual differences in WMC were examined to investigate the predictive power of WMC in explaining the performance of L2 fluency in monologic and dialogic tasks. Finally, regarding the participants' gender, it was not possible to access male participants because of gender segregation in the educational system of Saudi Arabia. Also, as a lecturer teaching at the University of Jeddah, it was easier and more convenient to contact the students and their teachers at the female campus. This selection is called convenience or opportunity sampling. The participants were willing to volunteer as they were students at the same researcher's institution (Dornyei, 2007).

The following materials were collected from 64 participants in the following sequence (see Figure 2):

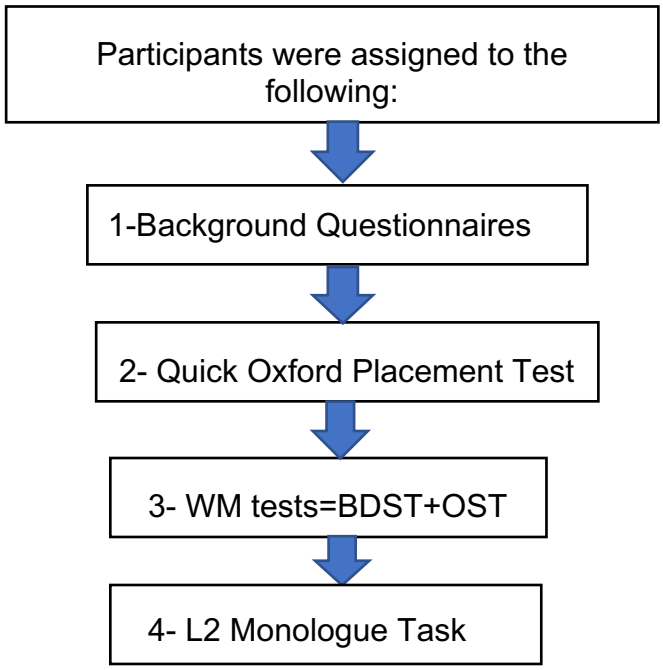


Figure 2. Steps in Collecting Current Study Data

Since there was only one group of participants, they followed the above sequence for data collection. For example, all of the participants took the first two tasks (see Figure 2). Then, with the researcher in a quiet office, each participant took a Backward Digit Span Test, an Operation Span test, and a monologue task. When two participants finished the first four tasks, they were asked to take part in a dialogic task in a pair. To control for the practice effect, a counterbalance order of performances was an option. However, each participant's performance was important for comparing utterance fluency in monologue and dialogue. Also, as an example of preparedness, both task modes contained guideline statements, questions, or keywords. This decision was taken after piloting the study. Thus, to ensure the equality between two task modes in terms of content palnning, preparedness examples were provided in the prompt cards.

3.6.1 Background Questionnaires and Proficiency Test

All participants received L1 background questionnaires (Appendix 1) to prevent there being any effect of L2 learning experiences outside class or abroad. For example, the questionnaires were used to exclude any participant who had experience learning L2 abroad. The questionnaires asked for information such as the participant's name, age, L2 experience, their learning of English outside the classroom and the number of languages the participant could speak. The biographical background questions revealed that all the participants were homogenous in their age and language learning

experience. They had all started learning English in grade 7 in a foreign language classroom.

The participants completed the grammar section of a standardised general proficiency test – the Oxford Placement Test (OPT) – (Allan, 2004) (Appendix 2) to exclude any outliers and to make sure that the participants recruited were from a similar band of proficiency. OPT is a pen-paper test used to assess participants' English levels of knowledge in grammar, reading and vocabulary. The test consists of two parts. In the first part, the participants can score 1 to 40 points and on the second part another 20. Thus, 60 is the maximum score. The OPT should be completed within 30 to 45 minutes, as mentioned on the first page of the test kit.

Based on the results of the Oxford Placement Test (OPT) (Table 3), the participants were homogenous in terms of L2 proficiency level (M= 37.9 and SD= 4.22). Aligning the scores with the Common European Framework of Reference for Languages (CEFR), all of the participants fell into the category of an intermediate level of proficiency (B1). According to the Common European Framework of Reference (CEFR), English users of intermediate level B1 are considered to be independent English language users. In other words, they "can interact with a degree of fluency and spontaneity and can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue" (CEFR, Council of Europe, 2001, p.24).

Table 3. Descriptive Statistics of OPT Scores

OPT Scores	Mean	Standard Deviation	Range	Minimum	Maximum
N=64	36.09	4.22	26	22	48

3.6.2 Measures of WMC

Previously published SLA studies have not yet agreed on reliable, valid measures for WMC; they tend to choose a task that aims to fully investigate the WM components (phonological loop, central executive, visuospatial, and episodic buffer) that tap into storage and processing. Some WM tests are designed to measure storage and processing separately or in combination (Weissheimer and Mota, 2011). Gilabert and Muñoz (2010) suggest using two WM tests to examine whether processing and storage are interrelated or separated systems. Thus, the dual function of WM in storing and processing has led researchers such as Daneman and Carpenter (1980), Turner and Engle (1989), and Kormos and Safar (2008) to use complex span WM tests in which the subjects are either asked to resolve a mathematical equation or to produce a sentence along with recalling a string of words, digits, or visuals. For example, they test the participants' abilities to read several separate sentences simultaneously and then recall the last word of each sentence. This is called the Reading Span Test (RST) and it was introduced by Daneman and Carpenter (1980). RST draws on both the recall and processing resources of WMC, and it can be conducted in a group or individually. Participants can recall words freely or in a specific order – as many words as they can remember (Juffs and Harrington, 2011).

Fortkamp (1999) highlighted that most research on WM has been carried out using two approaches. The first approach utilises neuropsychology, where the researchers focus on examining the four components of Baddeley's (2000) WM model by using dual tasks. For example, participants are asked to do two tasks simultaneously to compare the performance of one task against the other. The second approach is the psychometric correlation that researchers use to examine the relationship between the WMC and the performance of a cognitive task, such as speaking. This means that

WM tests such as the Speaking Span Test (SST), the RST, or the OST can be used to examine learning skills, for example, speaking, reading, writing, and listening. The second approach has been widely used in previous SLA studies (e.g., Mota, 2003; Guara´-Tavares, 2008; Kormos and Trebits, 2011).

As WM is a multi-dimensional construct, the present study used two complex WM tasks, BDS and OST, to measure the participants' performance in comprehension and production. Thus, the present study follows the psychometric correlational approach and focuses on the relationship between the individual differences in WMC (as measured by two complex WM tests, BDST and OST) and L2 fluency performance in monologic and dialogic tasks.

3.6.2.1 Backward digit span test (BDST)

The automated BDST measures the ability to remember and say numbers in reverse order. It has been used previously by SLA researchers (e.g., Morra, 1994; Kormos and Trebits, 2011) as not only a measure of short-term working memory but also a measure of a complex verbal WMC task. The rationale behind choosing BDST over its opposite, forward-digit test is that the latter measures only short-term phonological memory (Gathercole et al., 1999), whereas BDST can tap into the components of WM, namely, phonological loop, visual sketchpad and central executive (i.e., storage and processing) (Kormos and Trebits, 2011).

BDST is a language-independent test that depends on recalling numbers instead of letters or words, meaning that efficiency does not depend on the language of the task (Osaka and Osaka, 1992). The BDST in this study (see Appendix 3) was designed by the researcher in L1 Arabic to reduce the language proficiency effect because it would be difficult to identify whether the results of the task were measuring WMC or L2

speaking ability (Mitchell et al., 2015). Also, L1 abilities were considered to be less variable across the participants than L2 proficiency.

Following Conway et al. (2005), Kormos and Trebits (2011) and Awaad (2017), the test included seven sets. Each set contained random audio-recorded digits, starting with 3 digits for the first set then increasing in ascending order to 4, 5, 6, 7, 8 and 9 digits. There was a 1 second interval between each digit. Each student sat in a quiet room with the researcher and listened to the seven sets one at a time. For each set, each participant was given three attempts (trials) to repeat the digits backwards. For example, "564" should be repeated as "465". For scoring, I followed standard procedures from the previous studies by Wechsler (1997), Conway et al., (2005), Kormos and Sáfár (2008), Kormos and Trebits (2011) that suggested that each subject should be given three trials in each set to recall the digits (item) correctly once, twice or three times. The score is from 0-9. The highest possible score is 9. If the participant correctly recalls the item two out of three times, the test will continue onto the next set, until the last set with nine digits. BDST was terminated when a participant failed to correctly repeat any set two times out of three presentations, and the BDST was calculated based on the last set the participant recalled successfully twice in a reverse order. The whole test was written out on paper for the researcher to check the participants' answers. The participants were not able to take notes. The descriptive statistics, mean, minimum and maximum scores of the participants' BDST are presented in Table 3 below. Mean and Standard Deviation values of BDST (M = 4.10, SD=0.98) were similar to other studies that used BDST to predict L2 utterance fluency such as Kormos and Safar (2008) (M= 5.29, SD =1.29) and Awwad (2017) (M = 5.16, SD = 1.22).

Table 4. Descriptive Statistics of BDTS Scores.

WM BDST	N	Range	Min.	Max.	Mean	SD
	64	3	3	6	4.10	0.98

3.6.2.2 Operation span test

OST is a verbal working memory test that was first introduced by Turner and Engle (1989). The OST aims to measure the central executive of WMC, coupled with two subcomponents – the phonological loop and the visual-spatial sketchpad (Baddeley, 2017). OST is known as a language-independent test (Unsworth et al., 2005). The test in the present study was adapted and designed by the researcher in L1 Arabic to avoid any confounding between L2 proficiency and the WM test (Juffs and Harrington, 2011). OST is a one-on-one task in which each participant reads aloud arithmetic problems that appear in a PowerPoint presentation and says whether each operation is “correct” or “incorrect”. The reason for asking the participants to read each sentence aloud with no time limit is to avoid a practice effect. The test also requires each participant to remember the word that appears on the screen after each arithmetic operation. Figure 3 shows an example of how L1 equations and words are presented in OST. The words in the OST are monosyllabic nouns with 4–6 letters. After piloting the test, it was decided that the presentation of each arithmetic problem should last for 3.60 seconds, whereas 1.15 seconds was enough for the presentation of each word. The test had four sets and within each set there were four levels. The first level started with two words, followed by levels with three, four and five words. The same presentation was repeated three more times, making a total of 56 words.

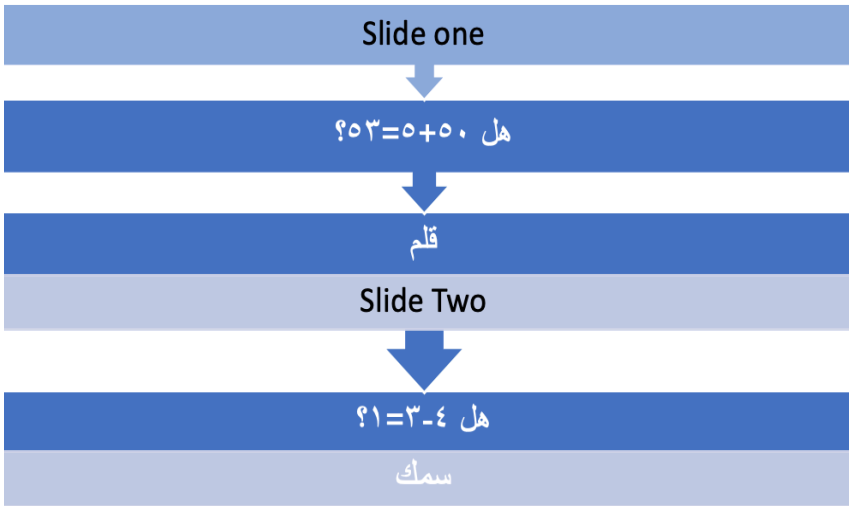


Figure 3. Example of Two-block Question in OST

3.6.2.2.1 Procedure

After each level in a presented set, the participant was provided with a sheet numbered from 1 to 56 (see Appendix 4), and asked to fill it in by writing down the words they can remember, if possible with the correct words in each level. However, the participants were informed that they could not start by writing out the last word from the sentence they had just read. The answer sheet showed the level number, and next to each level was a number of question marks to give the participant hints about how many words they should remember to write. For example, level one had two question marks (??), level two had three question marks (???), and so on. Each participant had ten to fifteen minutes to complete the test. One half of the operations were correct, and the other half were wrong to avoid rehearsal (Duncan et al., 2012). Based on Conway et al.'s (2005) scoring guidelines, each correctly recalled word was worth one point. The highest possible score was 56. In order to be included in the data analysis, the participants' scores should reach no less than 85% accuracy in the arithmetic part of the test. This means that if participants' scores were less than 45 in the arithmetic part, they would be excluded from the data (Conway et al., 2005; Mitchell et al., 2015).

To make sure that all the participants achieved at least 85% accuracy in the processing part of the OST, each participant's score was calculated in Excel software. For example, a score of 53 was divided by the perfect score of 56 and multiplied by 100 (=53/56*100). If the participant's score was less than 45 in the arithmetic part, which was less than 85% accuracy for the component of processing, this participant was excluded from the data. This decision is important to ensure a shared or an equal contribution of both WM components: storage and processing (Conway et al., 2005; Mitchell et al., 2015). The results indicated that all but two participants achieved above 85% accuracy percentage (M = 53.57, SD = 1.84). These two participants, who had 82% accuracy, were discarded from the data.

Table 3 reports the descriptive statistics for the total memory span for 64 participants. The scores are calculated based on PCL (partial credited load) scoring. This is the sum of all elements (words) correctly recalled in all sets regardless of whether the words were recalled in the correct order or not (Conway et al., 2005). The highest possible score for OST was 56. The PCL scoring method shows the variabilities in the participants' WM capacities. This method also allows better comparison with previous studies (Conway et al., 2005).

Table 5. Descriptive Statistics for OST Storage Component.

WM OST	N	Range	Min.	Max.	Mean	SD
	64	28	20	48	30.57	8.15

3.7 L2 Speaking Tasks

The oral performances of utterance fluency in both task modes, monologue, and dialogue, were collected to examine the differences between them in terms of speed,

breakdown, and repair. The questions of both tasks were selected to encourage L2 participants to pay attention to meaning rather than form (Hanzawa, 2021).

3.7.1 Monologue Task

Previous studies have highlighted the importance of using different topics when a participant engages in two modes of speaking (e.g., monologue and dialogue) in order to avoid the practice effect (Tavakoli, 2016). The monologic task in the present study was an argumentative task about a familiar topic to be discussed and argued by the L2 participants. Cucchiarini et al. (2010) argue that answering questions spontaneously can help learners to speak more fluently than in a picture description task. In this study, each participant received a task card (see Appendix 5) with the statement and the keywords. This is similar to the structure of the IELTS test. The participants were asked to give an opinion with justifications about the following statement and how far they would agree with it: '*Social media (e.g., Twitter, Instagram, Snapchat, and WhatsApp) is the current and future of marketing in Saudi Arabia'.* Different kinds of preparedness such as task repetition, focus on form and taking notes have been found to be effective for increasing L2 oral task performance (Skehan, 2014b). In the monologic task, the participants had two minutes to plan and take notes and two minutes to speak. The time was measured with a stopwatch. The participants were not allowed to use notes when speaking as they can prevent smooth speech. Piloting the task showed the need to add the English keywords shown in Table 5 because the participants found it difficult to find the right words for arguing and expressing their opinions. For example, some participants had the Arabic vocabulary, and asked for their English equivalents.

Table 6. English Keywords Given to Participants in Monologue Task

advertisement	famous	followers	celebrities	fashionista
attract	popular	products	influencers	accessible

3.7.2 Dialogue Task

The instructions for the dialogue task required the participants to exchange opinions and discuss the following statements with a partner, deciding whether they agreed or disagreed, with justifications. *'Do you agree/disagree with allowing women to drive in Saudi Arabia?'*

This statement was followed by a number of short statements and guiding questions on a prompt card (Appendix 6) to give the participants more ideas about what to say about this particular topic (e.g., advantages vs disadvantages). This topic was chosen because it allowed the students to speak about a familiar topic in a natural situation, i.e., dialogue. When a speaker shares a personal or true experience with an interlocutor and discusses it further in a conversation, this may result in fluent spontaneous speech with fewer unnecessary pauses (Derwing et al., 2004; Kahng, 2014). Thus, 64 participants were divided into pairs, making 32 pairs in total. Each pair met with the researcher in a quiet room to conduct the dialogic task. As previously done in monologic task, the participants were allowed to take notes. The participants were given two minutes to prepare individually and familiarise themselves with the topic before they started speaking.

3.7.3 Rate the Complexity of the Task

The choice of the dialogue and monologue tasks was guided by the following criteria. First, both tasks were rated for the level of difficulty by ten experienced teachers. The results of this rating showed that task type (monologue and dialogue) and the topics were a medium level of difficulty and suitable for the students' level of proficiency. Second, the topics of both tasks had not previously been discussed in the classroom, thus avoiding the practice effect.

Ten experienced L2 English instructors were asked to rate the monologue and dialogue tasks for their degree of complexity and difficulty on a 5-point scale, from extremely easy to extremely difficult (Tavakoli, 2009). The L2 instructors were asked to first read both tasks (see Appendices 5 and 6) and their instructions. Then the researcher handed them a rating task paper and asked the instructors to underline their answers (see Table 5). The English instructors' rating, to some extent, guided the choice of the tasks. For example, they considered the topics and the questions applicable and suitable for the students' level. The tasks were scored 3 for difficulty on a 5-point scale.

Table 7. The Rate of the Task Complexity.

Extremely easy Extremely difficult

1	2	3	4	5

3.8 Pilot Study

Conducting a pilot study is a very important step that should be taken before the implementation of the main study as it helps the researcher to broadly understand the topic and procedures of the study (Lancaster, 2015). It also provides the researcher

with an initial estimate for calculating the number of participants (Lancaster, 2015). Additionally, a pilot study shows the main challenges and complications that arise during the data collection stages and it proposes possible suggestions and changes for the actual data collection. Conducting a pilot study helps to test the suggested research instruments in order to ensure that these instruments are suitable for answering the research questions. For example, piloting the current study provided a deep insight into how the participants performed the WM tests and this determined the amount of time that the participants needed to understand and complete the tests. Furthermore, piloting the dialogic task helped the researcher to decide whether the question or the topic suggested in the conversation task were suitable for the students' L2 level in order to talk equally. It also helped to avoid any unclear instructions that might cause misunderstanding between the participants in the actual data collection.

3.8.1 Procedures and Instruments

The pilot study was conducted at the English Language and Translation Department at the University of Jeddah, during the period of 13th January 2019 to 16th January 2019. First, agreement was obtained from this department to conduct this study. On the 14th January, I visited the participants' classroom with the permission of their instructor to explain the purpose of the study in Arabic and to answer any questions that they might have before collecting the informed consent sheets (Appendix 6). It was stressed that the participants could stop or withdraw from the study at any time during the data collection procedure. The L2 participants who volunteered to be a part of the pilot study were fourteen level-6 (third-year) undergraduate students. Their ages ranged from 20 to 21. The data collection took three days, and the participants met the researcher three times at the university. The first meeting was held in order to

complete the Quick Oxford Proficiency Test (OPT) and a language background questionnaire. The second meeting was a one-on-one meeting with the researcher to complete the two WM tests and the monologic and dialogic tasks in a quiet space. The five instruments were administered as follows:

3.8.2 Day 1: Oxford Placement Test OPT

Fourteen L2 students took the paper and pencil OPT (see Appendix 3) in a classroom. The participants' responses were scored on a scale of 1 to 60, with each correct answer being worth one credit. The researcher marked each of the fourteen students' tests using the key answers sheet that was provided with the test kit, and then went on to align the overall participants' scores based on the CEFR (Council of Europe, 2001). Based on the results of the OPT (M=36.78 and SD=6.99), the students' level of proficiency was an intermediate level of L2 competence. In the CEFR, a score of 30–39 is equivalent to B1, and English users of B1 level are considered to be independent English language users. In other words, they "can interact with a degree of fluency and spontaneity and can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue" (CEFR, Council of Europe, 2001, p.24).

3.8.3 Background Questionnaire

After completing the OPT, the participants received an L1 (Arabic) background questionnaire (see Appendix 2) including questions about their English language learning history and biographical information. The participants' answers showed that they did not know any languages other than their L1 Arabic and L2 English. Additionally, the L2 participants had not studied or lived abroad. Thus, it can be

assumed that they were homogenous in terms of language learning experience and time spent learning L2.

Days 2 and 3, the following data were collected:

3.8.4 Operation Span Test (OST)

The PowerPoint slides were set and timed to run automatically on a laptop screen (Appendix 4). The laptop was given to the participant, who was asked to begin when she was ready. A model example was introduced so that all the participants could familiarise themselves with the task procedure. Almost all the participants were able to carry out the processing part of the test accurately. Table 6 shows the mean and standard deviation scores (*M* = 52, *SD* = 2.34) for the processing part. The mean score of 52 indicates that the participants' scores were above 45 for accuracy (Conway et al., 2005). As for the word recall scores (WM storage), Table 6 shows the mean and standard deviation (*M* = 22.28, *SD* = 4.90) of the number of words that were correctly recalled by the L2 participants.

Table 8. Participants' WM Scores in OST

OST	N	Min.	Max.	Mean	SD
Processing scores	14	48.00	56.00	52.64	2.34
Storage scores		16.00	36.00	22.28	4.90

3.8.5 Backward Digit Span Test

After taking the OST, each participant took part in the BDST (see Appendix 3). The test instructions were explained in the participant's L1. Also, a trial example was

introduced to ensure that the participant understood the test procedures. Each correct answer was worth one credit. Table 7 shows the M and SD scores of the participants' OST (*M* = 4.86, *SD* = 2.18).

Table 9. Participants' Scores in BDT

Backward Digit Test	N	Mini	Max	M	SD
Scores	14	2.00	8.00	4.86	2.18

3.8.6 Monologue and dialogue tasks

For the monologue, each participant received a monologic task (see Appendix 5) to discuss the following statement: '*Social media (e.g., Twitter, Instagram, Snapchat, and WhatsApp) is the current and future of marketing in Saudi Arabia'.* The participants were asked to plan for two minutes and speak for two minutes as well. For the L2 speaking, the participants utilised the entire two minutes of the planning time in order to prepare their speech. The time taken on the task varied from 45 seconds to one minute (i.e., time the students spoke for). The participants' performances were audio recorded on a digital audio recorder.

For the dialogue, the fourteen participants were paired up randomly, so there were seven interlocutor pairs in total. Each pair met the researcher in a quiet office in order to perform a dialogic task in L2. The participants were asked to sit facing each other. Each pair were provided with a written prompt (see Appendix 6), which included instructions to exchange opinions and discuss the following statement with a partner: "*How far do you agree with allowing women to drive in Saudi Arabia*?" Their performances were recorded on a digital audio recorder. The researcher did not set a

time on this occasion for the participants to speak; therefore, the participants' speaking time varied, and some participants did not talk enough. Thus, the researcher decided to instruct the rest of the participants (four pairs) to plan for two minutes and perform the task for two minutes each participant. The other reason why some participants spoke less was their need to take notes on a piece of paper in order to plan what they said and gather ideas.

3.8.7 Changes and Suggestions for Main Data Collection

After piloting the current study, a few suggestions were added to the main study. In the OST, it was decided that the duration of each arithmetic problem presentation on the PowerPoint should be 3.60 seconds and that 1.15 seconds was enough for the word presentation. This decision was made because when piloting the OST the participants finished solving the arithmetic operations in less than 5 seconds, and then stayed silent until the next slide presentation.

After piloting the monologic and dialogic tasks, both tasks were reconstructed to ensure equality of instructions in both tasks and to enhance the participants' performances in both task modes. The pilot study showed the need to add more vocabulary items to the monologic task, whereas in the dialogue the participants needed more guiding questions to carry on the speech. Additionally, the results of the pilot study suggested the need to add a number of L2 keywords to the monologue task and to fine-tune the dialogue task instructions by adding more L2 statements, in order to reduce unnecessary pauses and encourage speaking for a longer period of time. One source of limitation that could have affected the results was that the two tasks were not identical in the kind of speaking function (genre) they require for task

completion: one was an argumentative task (monologue) and the other a discussion (dialogue).

In the context of monologue and dialogue, the researcher suggested allowing the participants two minutes planning time to take notes and prepare their L2 speech; however, in the main data collection, the participants were informed not to use the notes when speaking, as recommended by Fortkamp (2000) and Ahmadian et al. (2012). The reason behind this was to prevent any slowness or dysfluency in speech processing caused by the participants reading from the cards. Also, Nitta and Nakatsuhara (2014) noted that long pre-task planning may prevent participants from interacting in the task collaboratively and decrease their speech rate. They might focus on trying to remember their previously planned speech, which causes dysfluency.

Finally, the researcher decided to not show the stopwatch to the participants, but instead to interrupt them when the time was up. This was because in the pilot study it was noted that the participants were worried about the time remaining on the task when the timer was counting down the seconds on the phone screen, which affected their attention given to the task, as well as the flow of their speech.

3.9 Ethical Considerations

Ethical approval (see Appendix 7) has been granted by the University of Leeds and the University of Jeddah. In order to obtain consent, the researcher took the following steps:

3.9.1 Obtaining Informed Consent

The administration office and the participants at the University of Jeddah received the information sheet (Appendix 8), which explained the aim and process of this study and

contained the required consent form for the participants (Appendices 9 and10). They needed to sign this form to signal their willingness to be a part of the study before data collection began. All the information on the consent form was translated to L1 Arabic to ensure the participants' understanding.

3.9.2 Risks

The study was conducted in my home country, Saudi Arabia, at the University of Jeddah. I travelled to the participants' university for the whole period of data collection. The data collected from the participants were anonymous and no personal information was taken. The participants were informed orally and in writing on the consent sheet (Appendix 8) that their voices would be audio recorded for the purpose of the study and that they had the right to withdraw from the study at any time in the data collection process. All the recordings were saved on a secure device on the university drive. For the purpose of this study, I am the only one who had access to the participants' recordings. To this end, there was no known risk except that some volunteer participants might not complete some parts of the tests in the data collection process; therefore, the researcher had to rearrange for alternative participants.

3.10 Main Data Collection Procedures

The data were collected in three weeks and during students' regular school days. All the participants were given a brief explanation about the purpose of the study by their lecturer. All measures (tests and tasks) were administrated in the following order: the Oxford Quick Placement Test and background questionnaire; BDST and OST; and the monologue and dialogue tasks. All tasks were completed in a quiet room without disturbances and the participants were able to choose a day and time of their

convenience. To ensure the participants' understanding for all tasks, L1 and L2 instructions were provided for each task, as well as a practice example.

The period of data collection was from 11th March 2019 to 1st April 2019. The researcher checked each student's willingness to participate in the study and explained that the participant had the right to withdraw from the study at any time. Then, the researcher gave each participant time to read and sign the consent form. Each participant took approximately 25-30 minutes to complete the OST, BDST, monologue, and dialogue tasks. Ten participants were excluded from this study because 7 of them had less than 20 in the proficiency test, 2 students did not reach 85% accuracy in the automated OST and 1 student did not complete both working memory tests. The total number of participants included in the analysis was sixty-four.

Each participant completed the OST then the BDST (for tests details see Section 3.6.2.1). They all used the researcher's laptop to complete the WM tests. The L1 OST instructions were explained orally to each participant in individual one-to-one meetings with the researcher. Additionally, the instructions for the OST were provided on the first slide of the PowerPoint presentation. In OST, each participant was asked to complete two tasks: (1) read aloud a sentence that contained a simple arithmetic problem, such as addition, subtraction, or multiplication, and decide whether the operation was correct or incorrect; and (2) remember the last word next to each equation. The time spent on completing this task was 10–15 minutes for each participant. Next, the participants completed the BDST. The participant was required to listen to random digits that increased in ascending order (3–9) and then repeat these digits backwards. There were seven sets and the first set started with three digits. Each participant had three trials to repeat the numbers backwards correctly. A trial

example was provided: “632” should be repeated as “236”. The BDST was printed on paper for the researcher to check each participant’s answers and calculate their score. The average time spent on the BDST was between 4 and 7 minutes long.

For the monologic task, each participant spoke into a microphone that was connected to a Philips voice recorder (DVT8010) to record the L2 speech performances about a popular topic in their country (see Appendix 5). The researcher audio recorded all of the participants’ responses with their knowledge and agreement. After collecting data from two participants individually, the researcher randomly paired them to perform the dialogue task. In the same room, there was a table with two chairs facing each other to allow the two participants to discuss the given topic in a dialogue setting. Each pair was provided with a written prompt (see Appendix 6). The dialogue task was recorded using the same device (Philips voice recorder (DVT8010)) and the recordings were transferred to a computer for further analysis.

3.11 Data coding

3.12 Fluency Measures

SLA researchers (e.g., Skehan, 2003, 2014; Kormos and Denes, 2004; Kormos, 2006; de Jong et al., 2012, 2015; Kahng, 2015) have generally agreed on the most reliable measures for different aspects of utterance fluency under three major subcomponents – speed, breakdown, and repair. These are the composite measures of mean length of syllables (MLS), phonation time and ratio (PTR), and mean length of run (MLR).

Since this study concentrates on a low-order fluency (narrow sense), selected measures of temporal variables are used to measure the smoothness and speed of overall oral proficiency and delivery. The 17 most commonly used measures in L2

fluency research are presented in Table 10 below, together with their definitions and calculation methods. They capture different aspects of fluency in monologue and dialogue and enhance comparability with previous studies. The first four temporal variables measure the speed of oral delivery. To ensure comparability across studies, speech rate, articulation rate, PTR and MLS were included for measuring speed in both task modes. However, some L2 researchers (e.g., de Jong, 2016; Tavakoli, 2016) recommended carefully choosing fluency measures to avoid overlap between the measures and to capture different fluency dimensions. For example, articulation rate is a pure measure that captures the speaking speed without pauses. Some measures of speed fluency have not been examined before, such as length of speech sample. This measure was included in this study with the aim of enhancing the reliability of utterance fluency measures and gaining a broad understanding of the complex construct of utterance fluency in both task modes. Other measures have been used less in the literature of L2 fluency, such as PTR and all repair measures. These measures were also included in this study to examine their reliability among other measures.

The rest of the measures that are presented below were used to calculate disfluency of oral delivery such as pause location, pause type, reformations, repetitions and false starts. Although the participants were given two minutes to talk, the recording length for each participant varied. Due to the variety of speaking lengths, the measures of speech rate, pauses and repairs were counted by the amount of time spent speaking in seconds and then multiplied by 60 to measure one minute of each participant's speaking time (e.g., Kormos and Denes, 2004; Lahmann et. al, 2015).

On the one hand, Most SLA researchers (e.g., Kormos and Denes, 2004; Bosker et al.,2012; de Jong, 2016) usually use frequency measures of disfluencies (e.g., pauses, repairs or repetitions) by counting the number of disfluencies divided by the length of time spent speaking excluding silent pauses (SPs). For example, reformation is counting the number of phrases or clauses that are repeated with modifications to the word order, syntax, or morphology, and dividing by the total spoken time excluding SP (de Jong, 2016). On the other hand, some researchers (e.g., Ginther et al., 2010) corrected fluency measures for the total time of speaking including SPs. Correcting fluency measures means using common calculation methods for measuring temporal variables of fluency, including or excluding silent pause time.

de Jong (2016; 2018) recommended using frequency measures corrected for speaking time excluding SP time because using the total time may result in measures that confound with SPs. Thus, using measures that are unrelated theoretically and correcting them for the length of speaking time excluding SPs is a reasonable decision when comparing speakers within a language, for example, filled/unfilled pauses, repairs, mean duration of SPs, and mean length of syllable. Following Bosker (2013) and de Jong (2016), the measures in this study were calculated per spoken time excluding SPs.

Table 10. Dependent Variables of 3 Categories and 17 Measures of L2 Fluency in Monologue and Dialogue Tasks

Utterance fluency		Label of the measures	Operational definitions
1.	Speed Fluency	Speech rate	(SR) Number of syllables per minute. It was calculated by dividing the total number of syllables by total response time including all utterances and pauses, multiplied by 60 (Nitta and Nakatsuhara, 2014).
2.		Mean length of syllable	(MLS) was counted by dividing phonation time (time taken to perform the task excluding SPs) by the total number of syllables (Bosker et al., 2013). It is a measure of MLR. MLR is defined as mean number of syllables that are produced between SPs of 0.25 or more (Riggenbach, 1991; Towell et al.,1996).
3.		Articulation rate	(AR) Total number of syllables divided by spoken time (phonation time) excluding silent and turn pauses and multiplied by 60 (Kormos and Denes, 2004; Ginther et al., 2010; Witton Davies, 2014).
4.		Phonation time ratio	(PTR) The percentage of time taken to perform the task excluding SPs (phonation time)/total response time, multiplied by 100 (Kormos, 2006).
5.		Number of silent pauses/60	(NSPs) more than 250 ms per minute spoken time. SPs were calculated by dividing the total number of SPs by total amount of time spent speaking in seconds (exclude SPs), multiplied by 60 (Kormos and Denes, 2004, Ginther et al., 2010; de Jong, 2016; 2018).

6.	Breakdown Fluency	Number of Mid-clause silent pauses/60	(MCSPs) of 250 ms or more divided by the time spent speaking (excluding SPs) multiplied by 60 (Bui and Huang 2016; Kahng, 2017).
7.		Number of End-clause silent pauses/60	(ECSPs) of 250 ms or more divided by time spent speaking (excluding SPs) multiplied by 60 (Bui and Huang 2016; Kahng, 2017).
8.		Mean length of mid-clause SPs	(MCSPs)Total length of MCSPs/total number of MCSPs that last 250 ms or longer.
9.		Mean length of end-clause SPs	(ECSPs)Total length of ECSPs /total number of ECSPs that last 250 ms or longer.
10.		Number of filled pauses/60	Number of lexical/non-lexical filled pauses (FPs) per minute spoken time (e.g., *eh, um, uh, yea, you know*) (de Jong, 2013). Pauses that lasted 250 ms or longer were calculated by dividing the total number of FPs by the total amount of time spent speaking (excluding SPs), and then multiplying by 60 (Kormos and Denes, 2004; de Jong, 2016, 2018).
11.		Number of Mid-clause filled pauses/60	(MCFPs) Number of mid-clause FPs of 250 ms or more divided by the time spent speaking (excluding SPs) multiplied by 60.
12.		Mean length of mid-clause FPs	Total length of MCFPs/ total number of MCFPs that last 250 ms or longer.
13.	Repair Fluency	Reformulations	(RF) Number of any phrases or clauses that are repeated with modification to word order, syntax, or morphology, divided by total spoken time (excluding SPs) and multiplied by 60 (Kormos and Denes, 2004; de Jong, 2016).
14.		Repetitions	(RP) Any repeated word, phrase, or clause without any modification. The number of partial or complete

			repetitions per second spoken time was calculated by dividing the total number of repetitions by the total time spent speaking, excluding pauses and multiplying by 60.
15.		False Starts	(FS) Number of utterances or ideas that are discarded or eliminated without being completed. It was calculated by dividing the total number of false starts by the total time spent speaking, excluding pauses and multiplying by 60.
16.		Replacements	(RPL) Number of words or phrases that are immediately substituted for another (Ellis et al., 2005). It was calculated by dividing the total number of replacements by the total time spent speaking, excluding SPs and multiplying by 60 (Bosker et al., 2013).
17.		Total number of repairs	Total number of all kinds of repairs divided by time spent speaking and multiplied by 60.

3.13 Monologue Analysis

L2 participants (N = 64) performed a monologue task (see Appendix 5) from which 17 fluency measures were derived (see Table 10). Various decisions were made about the analysis of the monologues. The first decision was related to the transcriptions of L2 speech samples. L2 participants' speaking performances were first transcribed manually by a free online tool called Transcriber, following Witton Davies (2014) and Kormos (2004) (see Figure 4). The researcher was able to type the audio content while listening to each participant's audio file that appeared at the top of the page. An expert PhD researcher in language education rated 10% of the monologue and dialogue data to check interrater reliability of the transcriptions. A high level of interrater reliability 90% was achieved among raters.

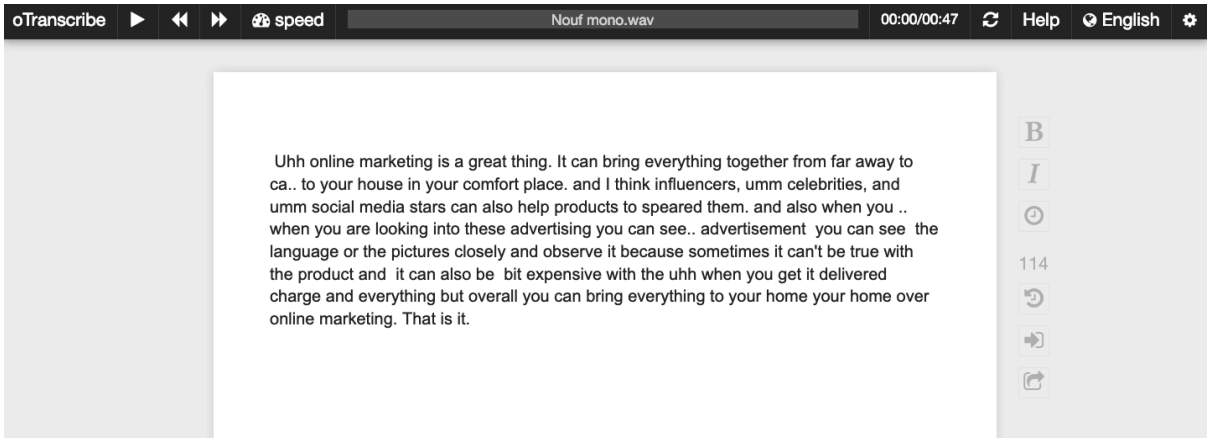


Figure 4. Transcriber.com, Free Online Tool

Second, in psycholinguistic studies, a syllable is a basic unit used to calculate temporal fluency measures (e.g., speech rate). Syllables per minute is the unit of reference used in the current study instead of number of words per minute. This is because using syllables/60 is a reliable measure that has been widely adopted in the literature of language teaching and SLA (e.g., Lennon, 1990; Kormos and Dénes, 2004; Tavakoli and Skehan, 2005; Derwing et al., 2009; Hilton, 2009; Ishikawa, 2015; Peltonen, 2018). Additionally, number of words/60 differs by genre; for example, dialogue speech samples were longer and had more words than the monologues and this might result in a higher speech rate (Tauroza and Allison, 1990). Therefore, using syllables/60 was recommended for calculating speed measures.

Following Thai and Boers (2016) Syllable Count program (www.syllablecounter.com) was used for the English transcripts to calculate the number of syllables per minute speaking time (see Figure 5). Syllable Count is a free online tool that can be used for calculating the total number of syllables in a sentence or a paragraph. This method helped to reduce the effort of counting the number of syllables manually, which saves time, especially for large data. It also helped to avoid errors in calculating syllables

manually. Intra-rater reliability was considered for a 10% sample of the transcription to ensure that the software syllable counting was reliabile. An expert PhD researcher in language education rated 10% of the monologue and dialogue data to check the interrater reliability of the syllable counting. A high level of interrater reliability 95% was achieved by the raters.

Figure 5. Syllable Counter Free Online Tool

The third decision for monologue analysis was related to the AS-unit (Foster et al., 2000) that was chosen as a main unit of speech segmentation. Pause location was identified as being either between or within AS-units. An AS-unit is defined by Foster et al. (2000) as "a single speaker's utterance consisting of an independent clause, or a sub-clausal unit, together with any subordinate clause(s) associated with either" (p.365). These terms are defined in turn. Starting with the definition of *an independent clause,* it is a clause that contains at least a finite verb and can stand alone, for example, |I agree with that|. *An independent sub-clausal unit* contains one or more phrases that "can be elaborated to a full clause by means of recovery of ellipted

elements from the context of the discourse or situation" (Foster et al., 2000, p.366). For example, |A*: how long have you lived here?* B*: three years*|. Additionally, the independent sub-clausal unit can be an irregular sentence, for example, |*Oh poor lady*| or a non-sentence, for example, |*yes*|. Finally, *the subordinate clause* consists of at least a finite or a non-finite verb plus subject, object, adverb or complement. For example, |and you would be surprised :: how he can work| (2 clauses, 1 AS-unit) (Foster et al., 2000, p.366).

Foster et al. (2000) showed that an AS-unit is an adequately reliable unit that can facilitate the analysis of spoken data. Furthermore, Foster et al. (2000) considered AS-units as genuine units of planning because many pauses in native speakers' speech occur at clause boundaries. Therefore, a syntactic unit is a valid unit of planning. AS-units are suitable for spontaneous speech transcripts that have a fragmentary nature. Furthermore, AS-units are also designed to deal with the characteristics of L2 spoken language, such as false starts, hesitations and reformations. For example, repetition is a common dysfluency feature in L2 speech and AS-units can be used to deal with repetition in a clause, for example:

| Uhhh sometimes [we found we found] some product |

| [It's it's uhh it's] It will be umm :: work first for me |

|Vertical lines| are used to mark the beginning and the end of AS-units (speech units).

[Square brackets] include repetition examples.

:: A double colon is used to mark a new clause within an AS-unit.

Also AS-units enable researchers to analyse speech that contains more than one clause (multi-clause units), for example: | It is easier to me :: to buy from social media | (2 clause, 1 AS-unit).

The final decision for monologue analysis was about silent pause length. Silent pauses are known as pauses that last longer than 250 milliseconds. Reviewing the literature, it could be seen that previously published studies on fluency differed in how they set the minimum length of a pause. However, the majority of L2 researchers, such as Towell et al. (1996), Bosker et al. (2012), de Jong et al. (2012), Kahng (2014) and Tavakoli (2016) adopted 250 ms as a minimum pause length. Moreover, Huensch and Tracy-Ventura (2016) used 250ms as a threshold pause for both L1 and L2 speaking data. According to Baker-Smemoe et al. (2014), any pause shorter than 250ms "may include onsets of stop consonants and any pause longer than this may miss important shorter pauses" (p.714). This study followed the previously mentioned studies and the silent pause exclusion criterion was set at 250 milliseconds or longer, regardless of their length.

Thus, in the current study, the length of silent and filled pauses was measured in milliseconds using PRAAT (Boersma and Weenink, 2012). PRAAT is free computer software used to analyse speech in terms of pauses, syllables, and speed. This was done by listening to each speech excerpt and using the function of zooming in and zooming out to examine the waveform and spectrogram for the length and type of pauses. Then the duration of all pauses was typed in the transcript. For example (see Figure 6), FPs number and duration are shown between parentheses:

|*But it's but (uhh)* (0.29) SP (0.52) *most of them:: it expensive* |
The duration of filled pause is 0.29ms.

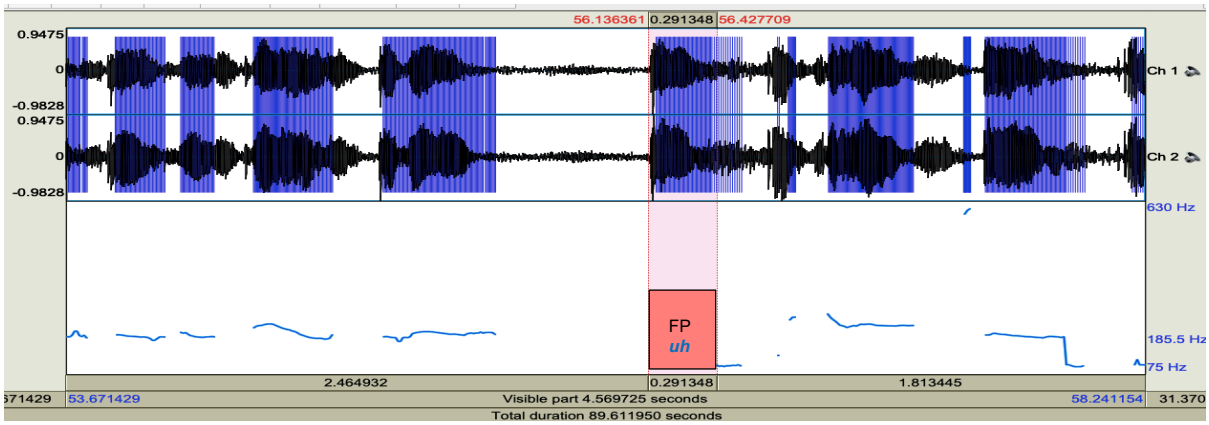


Figure 6. Speech in the Pause Condition

An example of silent pause duration is shown between parentheses in Figure 7:

|*I can't use anything abaaa about it (1.02)*|.

Duration of silent pause is 1.02 ms.

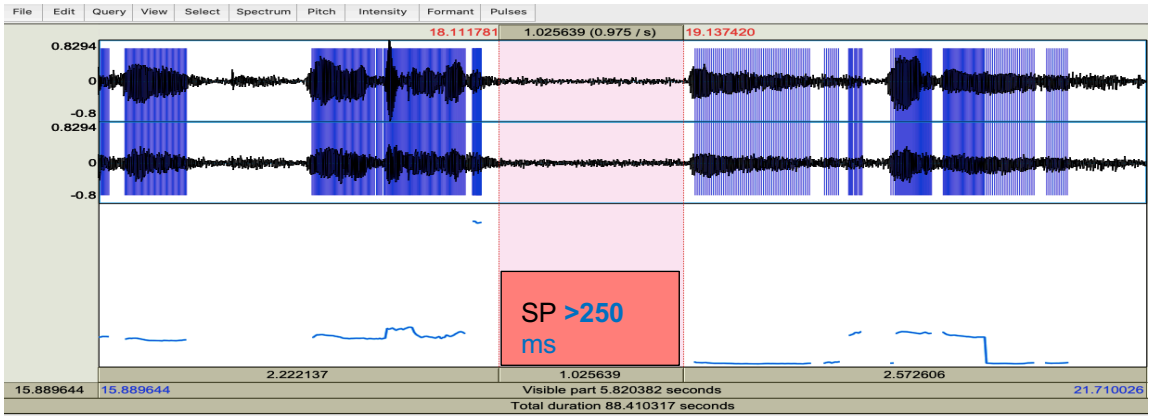


Figure 7. Speech in the Silent Pause Condition

3.14 Procedures

Total speaking time, and the number, location, and duration of silent and filled pauses were identified manually by PRAAT (Boersma and Weenink, 2012). Clause boundaries were marked and timed manually in the transcripts. For example, the silent/filled pauses were marked per AS-unit in the transcripts as being either at the end of an AS-unit (i.e., end-clause silent/filled pause) or within AS-unit (i.e., mid-clause

silent/filled) (e.g., Kormos and Dénes, 2004; Iwashita et al., 2008; Ginther et al., 2010; Michel, 2011). Filled pauses were marked by brackets, for example, (*umm), (uhh), and (yea).* Furthermore, the transcripts were annotated with symbols (see Appendix 11) for repair measures, for example, [/-] = false starts, [/] = repetitions, [&] = replacements, and [///] = reformations. For samples of the coded transcribed monologue data (see Appendix 12). For further details on the measures of intra-rater reliability for coding AS-Units and clause boundaries see section 3.18.

3.15 Dialogue Analysis

Dialogue analysis was similar to monologue analysis but was different in one procedure, which was dividing the conversation into two speaking tiers. Starting with the transcription process, the transcripts of 32 pairs (64 participants in total) were analysed using the same methods that had been used previously in the monologue. For example, the Transcriber, Syllables Counter and AS-units were considered in speech coding and segmentation. Additionally, similar to the monologue analysis, PRAAT software was used to identify filled/silent pauses' location, duration, and number. In dialogue, separate tiers for each speaker's speech were considered for the detailed analysis of each speaker (see Figure 8). For example, the dialogue for each pair was separated into tiers. Tier 1 belongs to speaker 1 and tier 2 belongs to speaker 2. Other dialogue measures, e.g., pause turn, between turn pauses, interruptions, were not included because they were beyond the scope of this study. The same measures of fluency (see Table 10) are used and operationalized equally in monologue and dialogue as this study aims to make a comparison between two modes in terms of speed of speech and smoothness of oral delivery (i.e., narrow fluency). Thus, these measures are used to cover all dimensions of L2 utterance fluency.

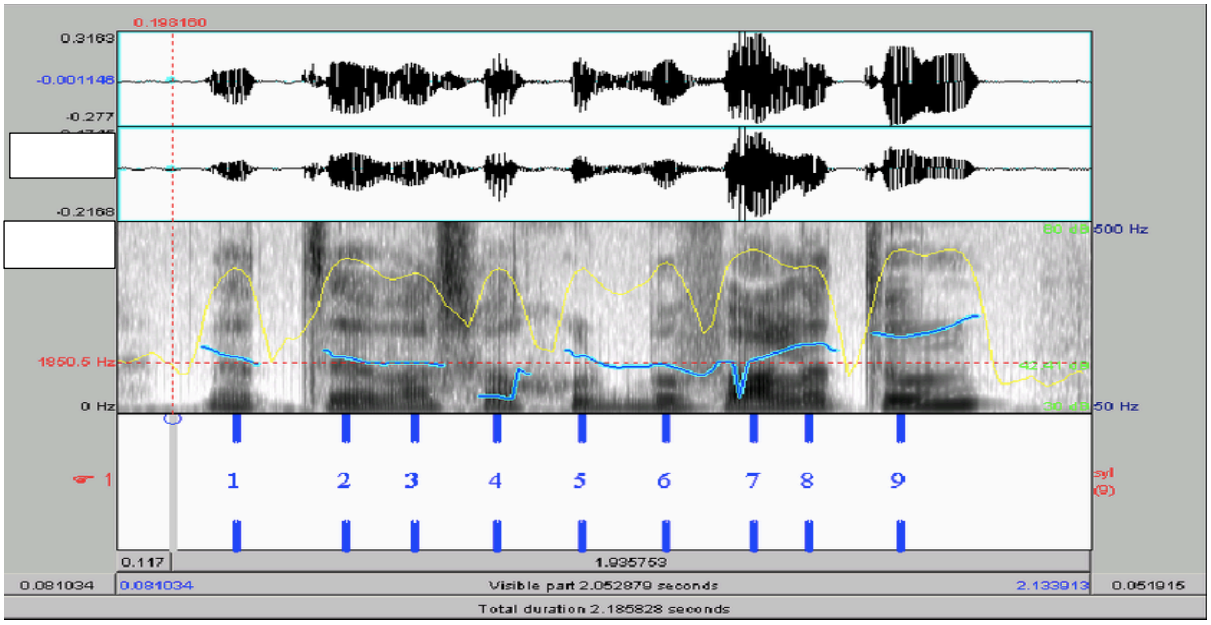


Figure 8 Dialogue Speech Analysis in PRAAT.

3.16 Summary of Dialogue Speech Analysis in Stages:

1. Recorded conversations between speaker A and B were separated into tiers for further analysis.
2. Pause types and locations were marked and timed for each speaker.
3. Repair fluency identified and marked for both speakers (e.g., reformations, false starts, etc.).
4. As in monologue, the number of syllables for each speaker was counted by an online free tool called Syllablecounter.
5. The transcription was checked twice by the researcher. The first time, the researcher listened to each track and read the transcript at the same time to check the accuracy of the typing of the words and sounds of the speakers' speech. The second time, the researcher added all kinds of pauses to the transcriptions and measured the time of speech for each participant individually

by playing the audio recording in PRAAT and zooming in and out to identify the time and pauses in the wave forms.

3.17 Data Analysis

Seventeen measures were used to operationalise the three aspects (speed, breakdown, and repair) of utterance fluency. Each participant contributed scores in all measures of L2 fluency in monologue and dialogue. For the data analysis of the research, IBM SPSS 25 was used to run descriptive and inferential analyses for each utterance fluency measure: speed (speech rate, articulation rate, MLS, PTR); breakdown (SPs, FPs, mean length of silent/filled pauses, end-clause SPs, and mid-clause silent/filled pauses); and repair (reformations, repetitions, false starts). Descriptive analyses of the mean and standard deviation were introduced to obtain the initial results about the significant differences in L2 oral fluency between monologic and dialogic performances (see Section 4.3). Non-parametric tests (Friedman's ANOVA and Wilcoxon signed Ranks) were used due to assumption violation of the normality tests. Normality was assessed by estimating skewness and kurtosis values as well as by applying the Kolmogorov-Smirnov and Shapiro-Wilk's tests (Pallant, 2007).

For the main analyses corresponding to the first research question, Friedman's ANOVA was conducted to determine whether there were any statistically significant differences between the mean scores of dialogic and monologic performances in terms of utterance fluency. Friedman's ANOVA is an alternative for a repeated measures ANOVA, while Wilcoxon's signed-rank test is considered as an alternative for a paired/dependent sample t-test (Field, 2013). Cohen's (1988) definitions of effect size were also included. For all analyses, we considered a p-value less than 0.05 as

statistically significant. Regression analysis was carried out to examine the power of WMC in predicting the L2 utterance fluency in monologue and dialogue (see Section 4.4).

3.18 Inter-rater Reliability

The oral performances of monologue and dialogue were transcribed, segmented, coded, and scored according to the measures selected to assess utterance fluency (see Table 8 for all fluency measures). To ensure the reliability of the segmented, coded and scored fluency measures and WMC tests, 10% (7 participants) of the data from monologue and dialogue were re-coded by a second bilingual (Arabic-English) independent PhD researcher. A third rater was consulted to resolve some disagreement in recoding the repair measures of fluency. Cohen's kappa statistics were applied to check the consistency of the placement of the clause boundaries and AS Unit, the PRAAT measurements of fluency in speed, breakdown repair, and in the calculation of the scores of the WM tests. The kappa is one of the most commonly used statistics to establish inter-rater reliability, which shows the degree of agreement among raters. The kappa coefficient was introduced by Cohen (1960), and it ranges between −1 and +1, with a higher number being desirable. According to Cohen (1960) criterion, the kappa (κ) values can be interpreted as follows: κ values ≤ 0 signify no agreement and 0.01–0.20 indicate slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 almost perfect agreement among the raters.

In light of the above criterion, a series of Cohen's κ was carried out to determine if there was any agreement between the first rater (the researcher) and the second rater (the PhD researcher in language education) on the coding of different measures of

utterance fluency and WMC across monologues and dialogues. The results showed that inter-rater reliability ranged between 0.68 and 0.78 for speed, breakdown and repair measures in monologue and dialogue. Also, the inter-rater reliability for WMC tests ranged between 0.65 and 0.76. Based on Cohen (1960) criterion described above, it can be concluded that the inter-rater reliability for our variables was substantial. The percentage of 78% is in line with previous studies; for example, an inter-coder agreement of 73% was reached by Levelt (1983), 76% by Blackmer and Mitton (1991), 70% by Oomen and Postma (2001) and 75% by Kormos and Declerck (2012). Thus, this substantial inter-rater reliability allowed the researcher to complete the data analysis with confidence because Cohen's kappa score confirmed the robustness of the procedure and the consistency of the transcription, coding, and testing.

3.19 Conclusion

This chapter provided a detailed explanation of the purpose of the current study, the research questions and hypotheses, the design of the study, the participants and settings, and the research methods used to collect the data. Additionally, a pilot study and ethical approval procedures were introduced. Then the following were discussed with justifications: the data collection, transcriptions, coding, and analysis. This study used a wide range of L2 fluency measures that clearly captured each aspect of speed, breakdown, and repair. The results of the current study will be presented through the descriptive and inferential statistical analyses and answers to each research question will be introduced in the next chapter.

Chapter 4: Results

4.1 Introduction

This chapter aims to answer the following research questions:

1- Are there any differences between monologue and dialogue in terms of utterance fluency measures?

   1a. Are there any differences between monologue and dialogue in terms of speed fluency measures?

   1b. Are there any differences between monologue and dialogue in terms of breakdown fluency measures?

   1c. Are there any differences between monologue and dialogue in terms of repair fluency measures?

2- To what extent can WMC predict the performance of L2 participants' utterance fluency in monologue and dialogue?

This chapter starts with introducing the results of the current study by employing a within-subject repeated measures design to examine the differences between the L2 utterance fluency performance in monologue and dialogue. After analysing the monologue and dialogue tasks of each L2 participant in terms of the suggested measures (see Table 10, p.110), I had a score for each participant in each measure: L2 utterance fluency, Backward Digit Span test (BDST), and Operation Span Test (OST). Descriptive statistics for each participant's score were calculated using IBM SPSS 26 then the researcher ran the following tests: normality test, Friedman's ANOVA test, Wilcoxon's signed ranks test, regression analysis, and correlation test. Effect sizes – Cohen's *d* (1988) – were also included to evaluate the degree of importance of the results.

As explained before, task modes (monologue and dialogue) and WMC (OST and BDST) were independent variables, whereas the dependent variables were 17

measures (see Table 10, p.100) used to operationalize the three aspects of L2 utterance fluency in monologue and dialogue (i.e., speed, breakdown, and repair). According to CEFR the 64 participants were intermediate level of language proficiency (B1) (M= 37.9 and SD= 4.22).

4.2 Testing Assumptions of Normality

According to Curran et al. (1996), skewness and kurtosis values should be within the criteria determined across a variety of statistical analyses. Based on the criteria set out for the assessment of univariate distribution, values ranging between ±1 refer to a "very good" distribution, while values of +/-2 refer to an "acceptable" distribution. Skewness values greater than 2 and a kurtosis value greater than 7 refer to "concern" over the distribution of the data. In light of these criteria, each of the fluency variables in the monologue and dialogue was evaluated to determine if they were normally distributed (see Table 11). Some of the main variables were not normally distributed because their skewness and kurtosis values were within the range of "concern". Only 13 variables were approximately normally distributed as their skewness and kurtosis values were within the +1 and -1 range. The following variables met the assumptions of the parametric tests: in monologue measures – sample length, speech rate, articulation rate, end-clause silent pauses, total repairs, repetitions; and in dialogue measures – speech rate, articulation rate, silent pauses, mid-clause silent pauses, filled pauses, mid-clause filled pauses, and total repairs.

In addition, the results of the Kolmogorov-Smirnov and Shapiro-Wilks tests (see Appendix 13) confirmed that the majority of the variables were not normally distributed because the *p*-values associated with these two tests were < 0.05. Only 11 variables had a *p*-value greater than 0.05, referring to a roughly normal distribution of the data.

Table 11. Skewness and Kurtosis Values for Utterance Fluency in Monologue and Dialogue

	Skewness		Kurtosis	
Variable	Statistic	Std. Error	Statistic	Std. Error
Monologue speed				
Length of sample	0.69	0.29	0.12	0.59
Number of syllables	1.24	0.29	1.46	0.59
Speech rates	-0.22	0.29	0.98	0.59
Articulation rates	0.19	0.29	1.43	0.59
PTR	-1.41	0.29	2.66	0.59
MLS	1.77	0.29	4.29	0.59
Dialogue speed				
Length of sample	1.41	0.29	2.80	0.59
Number of syllables	1.97	0.29	5.43	0.59
Speech rates	-0.22	0.29	-0.09	0.59
Articulation rates	-0.24	0.29	0.55	0.59
PTR	-1.26	0.29	1.90	0.59
MLS	2.21	0.29	5.83	0.59
Monologue breakdown				
SPs/60	2.20	0.29	10.07	0.59
Mid-CSPs/60	2.85	0.29	14.29	0.59
End-CSPs/60	0.14	0.29	0.37	0.59
M length Mid-CSPs	4.17	0.29	20.89	0.59
M length End- CFPs	3.49	0.29	13.91	0.59
FPs/60	3.25	0.29	18.82	0.59
Mid-CFPs/60	2.38	0.29	11.67	0.59
M length Mid-CFPs	6.84	0.29	49.55	0.59
Dialogue breakdown				
Silent Pauses/ 60	0.42	0.29	0.23	0.59
Mid-CSPs/60	0.98	0.29	2.30	0.59
End-CSPs /60	1.83	0.29	6.02	0.59
M length Mid-CSPs	4.54	0.29	26.51	0.59
M length End-CSPs	4.19	0.29	21.37	0.59
FPs/60	0.39	0.29	0.55	0.59

Mid-CFPs/60	0.29	0.29	0.25	0.59
M length mid-CFPs	7.91	0.29	63.07	0.59
Monologue repair				
Repairs/60	0.75	0.29	0.66	0.59
False Starts/60	1.55	0.29	1.89	0.59
Reformations/60	1.02	0.29	0.31	0.59
Repetitions/60	0.72	0.29	1.03	0.59
Replacements/60	1.04	0.29	0.21	0.59
Dialogue repair				
Repairs/60	0.68	0.29	0.61	0.59
False Starts/60	1.18	0.29	1.37	0.59
Reformations/60	1.06	0.29	1.43	0.59
Repetitions/60	1.09	0.29	1.54	0.59

4.3 Differences between Monologue and Dialogue in Terms of Utterance Fluency

Friedman's ANOVA and Wilcoxon signed ranks tests were used to answer the following main research question: *Are there any differences between monologue and dialogue in terms of utterance fluency measures?* This question was divided into three sub-questions and hypotheses in order to explore the differences between monologue and dialogue tasks in terms of (1) speed, (2) breakdown, and (3) repair fluency measures.

The Friedman test was first run to reveal whether there were any statistically significant differences between the oral performances (e.g., speed, breakdown, and repair) in the two modes. Friedman Test is the non-parametric equivalent to the One-Way repeated measures ANOVA Test. Friedman Test used to test the same participants more than once (Larson-Hall, 2015) and to compare mean scores of three or more dependent variables. Since the results of Friedman's ANOVA showed statistically significant differences, the post-hoc test, the Wilcoxon signed ranks test, was then run in order

to determine where the significant differences were based between dependent variables. The Wilcoxon signed ranks test is the non-parametric alternative to a paired-samples t-test or a dependent samples t-test. It was used to compare repeated measurements on a single sample to detect whether the sample's mean ranks were statistically varied (Larson-Hall, 2015).

4.3.1 Differences between Monologue and Dialogue in Speed Fluency Measures

In a dialogue the interlocutor and the speaker share the responsibility to talk and exchange turns (Tavakoli, 2016). During the other partner's turns, both (speakers and interlocutor) have an opportunity to prepare for what to say next (Michel, 2011). Thus, it was predicted that the speed of L2 learners' oral performances would be faster in dialogues than in monologues. Speed fluency was operationalised through six measures, namely *length of speech, number of syllables, speech rate*, *articulation rate, mean length of syllables (MLS), and phonation time and ratio (PTR)*. The results of this hypothesis are presented as follows: descriptive statistics, Friedman ANOVA, and Wilcoxon signed ranks test.

4.3.1.1 Descriptive statistics

The descriptive analyses median (MED), mean (M), standard deviation (SD), minimum (Min) and maximum (Max) scores were run for the L2 measures of speed fluency in monologue and dialogue and they are presented in Table 12. Although the differences between monologues and dialogue have not been tested yet, the overall results of descriptive statistics indicated that dialogues were more fluent than monologues. The number of syllables per minute was greater in dialogue (M = 252.67, SD =151.55) than

in monologue (*M* = 172.15, *SD* = 82.07). Speech rates were higher in dialogue (*M* = 156.57, *SD* = 41.50) compared to in monologue (*M* = 142.47, *SD* = 36.80).

Table 12. Descriptive Statistics for the Dependent Variables of Speed Fluency

Speed fluency	Median		Mean		SD		Min.		Max.	
	Mono.	**Dial.**	**Mono**	**Dial**	**Mono**	**Dial.**	**Mono**	**Dial.**	**Mono**	**Dial.**
Length of sample	66.50	85.84	74.32	97.18	30.80	47.58	26.00	32.11	168	281
Number of syllables	147	211	172	252	82.07	151	66.00	68.00	451	925
Speech rates	146	157	142	156	36.80	41.50	51.95	67.92	251	256
Articulation rates	183	200	181	192	37.99	46.13	83.66	75.94	305	306
Phonation time and ratio	80.52	83.52	78.12	81.38	9.63	10.26	42.78	48.20	92.09	95.71
Mean length of syllable	0.33	0.30	0.34	0.33	0.09	0.10	0.20	0.20	0.72	0.79

(N = 64). Mono= monologue. Dial= Dialogue.

To answer the following sub-question – *Are there any differences between monologue and dialogue in terms of speed fluency measures?* – the Friedman test was run to test whether there were any statistical differences in the speed fluency measures in monologues and dialogues. The results of the statistical test in Table 13 shows that there were statistically significant differences between speed fluency measures in the two modes, χ^2 (11) = 58.23, *p* = 0.00, meaning that the *p* value was <0.05. In addition, mean ranks scores show that these differences were statistically significant between the two modes as the participants were consistently more fluent in dialogue than in monologue. The exception to this was for MLS, which was slightly higher in monologue (*M* = 1.66) than in dialogue (*M* = 1.34).

Table 13. Mean Ranks for Dependent Variables of Speed Fluency

Speed Fluency Measures	Mean Rank		Chi-Square	df	Asymp. Sig.
	Monologue	**Dialogue**			
Length of sample	4.34	5.48	58.23	11	0.00
Syllables	8.88	10.69			
Speech rate	7.47	8.28			
Articulation rate	9.98	10.42			
Phonation time and ratio	4.44	5.02			
Mean length of syllables	1.66	1.34			

The significant results obtained from the Friedman test did not provide enough information in terms of where the differences between the ranges laid. Thus, it was necessary to run a number of Wilcoxon signed rank tests (see Table 14) on all the dependent measures of utterance fluency to show where exactly the differences laid between variables (Pallant, 2007; Field, 2013).

The use of the Wilcoxon signed ranks test could cause a Type 1 error rate inflation, meaning that the null hypothesis could be rejected even though it was probably true. For example, it might be assumed that there were differences between variables, when in fact there were no differences between them. Therefore, the probability value (alpha < 0.05) needed to be adjusted using the Bonferroni correction in order to minimise Type 1 errors (Pallant, 2007; Field, 2013). This was achieved by dividing the significance level of 0.05 by the number of variables used in the study. This estimation gave a new alpha (α), which was used to compare the *p* value for each of the Wilcoxon signed ranks tests.

Additionally, effect size (r) was calculated in order to estimate the importance of the findings of the current study. Effect sizes illustrate the proportion of the differences between the means (Fidell et al., 2013). Following Pallant (2011, p. 230), the size effects were calculated by dividing the Z value by the square root of N ($Z\sqrt{n}$). Z is the value from the statistics of the Wilcoxon signed ranks test and N is the total number of the cases or observations in pairs, not the number of subjects (Pallant, 2011). Cohen's (1988) guidelines for indicating the strength of the effect size (*r*) were adopted as: .10 ≤ *r* < .30 representing a small effect size; .30 ≤ *r* < .50 representing a medium effect size; and *r* ≥ .50 representing a large effect size.

The results of the Wilcoxon test analysis are presented in Table 14. The results show statistically significant differences between the two task modes (monologue and dialogue) with different effect sizes between the participants' performances in the monologue and dialogue tasks. The Bonferroni adjustment was applied as follows: 0.05\6 = 0.008 where 0.05 refers to the alpha value and 6 refers to the number of target variables compared in the analysis. This computation produced a new alpha value of α = 0.008, meaning that the p-values of the speed fluency measures should be < 0.008.

Table 14 Test Statistics for Differences in Speed Fluency between Dialogue and Monologue

Speed fluency	Pairs	Z	Asymp.Sig.2-tailed	Effect size
Length of samples	Monologue vs dialogue	-3.89[b]	0.00	-0.34
Syllables	Monologue vs dialogue	-5.25[b]	0.00	-0.43
Speech rates	Monologue vs dialogue	-3.10[b]	0.00	-0.27
Articulation rates	Monologue vs dialogue	-2.40[b]	0.02	-0.21
PTR	Monologue vs dialogue	-3.09[b]	0.00	-0.27
MLS	Monologue vs dialogue	-2.23[c]	0.03	-0.20

Note. *b* is based on negative ranks; *c* is based on positive ranks.

Table 14 presents the p-values of the speed fluency measures. In the one hand, there were statistically significant differences between the two modes in length of samples (Z = -3.89, *p* = .00) with a medium effect size *r* = 0.34; number of syllables (Z = -5.25 *p* = 0.00) with a medium effect size (*r* = 0.43); speech rate (Z = -3.10, *p* = 0.00) with a small effect size (*r* = 0.27); and PTR (Z = -3.09, *p* = 0.00) with a small effect size (*r* = 0.27). On the other hand, the participants' MLS (Z = -2.23, *p* = 0.03, *r* = 0.20) and articulation rates (Z = -2.40, *p* = 0.02, *r* = 0.21) were not statistically significant in monologues or dialogues and they had small effect sizes.

Thus, the hypothesis 1a is confirmed. The participants produced significantly longer speech samples (see Table 10) (*MED* = 85.84), a greater number of syllables (*MED* = 211), faster speech rates (*MED* = 16.05), and higher PTR (*MED* = 83.52) in dialogues compared to monologues. The median values in monologue speed fluency measures were: *MED* = 66.50 for speech samples, *MED* = 15.50 for speech rates, and *MED* = 80.52 for PTRs. No significant differences were found between the MLS in monologues (*MED* = 0.33) and dialogues (*MED* = 0.30), or between the articulation rate in dialogues (*MED* = 200.29) and monologues (*MED* = 18.21).

4.3.2 Differences between Monologue and Dialogue in Breakdown Fluency Measures

It was hypothesised that the breakdown fluency of L2 learners' oral performances would be less in dialogues than in monologues. Breakdown fluency was operationalised through eight measures, namely, silent pauses/60 (SPs), mid-clause silent pauses/60 (mid-CSPs), end-clause silent pauses/60, mean length mid-CSPs, mean length of end-CSPs, filled pauses/60, mid-filled pauses/60, and mean length of mid-clause filled pauses.

4.3.2.1 Descriptive analysis

The MED, M, SD, Min. and Max. scores for the monologue and dialogue breakdown fluency variables are reported in Table 15 below. In the one hand, generally descriptive statistics scores suggested that monologue speech samples had more silent pauses/60 (M = 24.16, SD = 9.36), end-clause silent pauses/60 (M = 14.09, SD = 4.21), and filled pauses/60 (M = 12.73, SD = 6.53*)* than the dialogue. On the other hand, the participants had higher scores on mid-CSPs/60 (M = 10.40, SD = 5.24), mean length of mid-CSPs (M = 0.90, SD = 0.75), mean length of end-CSPs (M = 84, SD = 0.71), mid-clause filled pauses/60 (M = 12.25, SD = 5.91) and mean length mid-clause filled pauses (M = 1.07, SD = 4.06) in the dialogue compared to the monologue.

Table 15. Descriptive Statistics of the Dependent Variable Breakdown Fluency

Breakdown Fluency	Median		Mean		SD		Min.		Max.	
	Mono.	**Dial.**	**Mono.**	**Dial**	**Mono.**	**Dial.**	**Mono.**	**Dial**	**Mono.**	**Dial**
Silent Pauses/60	22.91	17.23	24.16	17.28	9.36	7.00	9.54	3.52	72.46	37.90
Mid-clause SPs/60	9.23	10.67	10.18.	10.40	6.57	5.24	1.25	2.20	46.75	30.53
End-clause SPs/60	14.06	6.58	14.09	6.89	4.21	4.11	4.77	0.00	25.71	25.69
Mean length Mid-CSPs	0.58	0.70	0.71	0.90	0.46	0.75	0.34	0.31	3.44	5.75
Mean length End-CSPs	0.59	0.65	0.76	0.84	0.50	0.71	0.42	0.00	3.37	5.04
Filled Pauses/60	12.35	12.40	12.73	12.51	6.53	6.13	2.87	0.00	51.42	31.82
Mid-CFPs/60	10.09	11.90	10.26	12.25	5.53	5.91	2.46	0.00	39.74	29.89
Mean length mid-CFPs	0.55	0.52	0.73	1.07	1.06	4.06	0.36	0.00	8.58	33.00

To answer the following sub-question: *Are there any differences between monologue and dialogue in terms of breakdown fluency measures?* The Friedman test was run to compare the scores for the breakdown fluency measures in monologues and dialogues, and to investigate whether the participants' scores for breakdown fluency were statistically different across monologues and dialogues. The results of the statistical tests presented in Table 15 show that there were statistically significant differences between the two modes in terms of breakdown fluency measures**,** $\chi^2(15) = 765.96$, $p = 0.00$. The mean ranks and Friedman test results are presented in Table 16 below. According to the results, there were statistically significant differences between breakdown fluency measures across monologues and dialogues tasks. L2 participants scored significantly higher in monologue compared to dialogue in the following measures: silent pauses/60 (mono. *M* = 15.35 vs. dial. *M* = 13.62), end-CSPs (mono. *M* = 12.34 vs. dial *M* = 8.38), filled pause/60 (mono. *M* = 11.98 vs. dial. *M* = 11.51), and mean length mid-clause filled pause (mono. *M* = 3.04 vs. dial. *M* = 2.80). In contrary, mean length mid-clause silent pauses (mono. *M* = 3.43 vs. dial. *M* = 4.37), mid-filled pauses/60 (mono. *M* = 9.98 vs. dial. *M* = 11.18), and mean length end-clause silent pauses (mono. *M* = 3.76 vs. dial. *M* = 4.12) were greater in the dialogue than in the monologue.

Table 16 Mean Ranks for Breakdown Fluency

Breakdown Fluency	Mean Rank		Chi-Square	df	Asymp. Sig.
	Monologue	**Dialogue**			
Silent Pauses/60	15.35	13.62			
Mid-clause SPs/60	10.05	10.09			
End-clause SPs/60	12.34	8.38	765.96	15	.00
Mean length Mid-CSPs	3.43	4.37			
Mean length End-CSPs	3.76	4.12			
Filled Pauses/60	11.98	11.51			

Mid-CFPs/60	9.98	11.18
Mean length Mid-CFPs	3.04	2.80

As a post-hoc test, the Wilcoxon signed ranks test was carried out on each dependent variable of breakdown fluency to compare the mean score of each participant in monologue and dialogue. The differences between the two modes are presented in Table 17. As previously, the Bonferroni adjustment was applied as follows: 0.05\8 = 0.006 where 0.05 reflects the alpha value and 8 reflects the number of variables to be compared. For this analysis, the new α was set to 0.006 and any significant results for breakdown fluency measures should therefore be < 0.006.

Table 17. Test Statistics for Monologue and Dialogue Differences in Breakdown Fluency.

Breakdown fluency	Pairs	Z	Asymp. Sig.(2-tailed)	Effect size
SPs/60	Dialogue vs monologue	-5.190[b]	0.000	-0.46
Mid-CSPs/60	Dialogue vs monologue	-0.796[c]	0.426	-0.07
End-CSPs/60	Dialogue vs monologue	-6.380[b]	0.000	-0.56
M length Mid-CSPs	Dialogue vs monologue	-2.960[c]	0.003	-0.26
M length End-CSPs	Dialogue vs monologue	-1.164[c]	0.245	-0.10
FPs/60	Dialogue vs monologue	-0.298[b]	0.766	-0.03
Mid-CFPs/60	Dialogue vs monologue	-2.759[c]	0.006	-0.24
M length Mid-CFPs	Dialogue vs monologue	-1.265[b]	0.206	-0.11

Note. *b* is based on positive ranks; *c* is based on positive ranks.

According to Table 17, there were statically significant differences between the two modes in silent pauses/60 (Z = -5.190, p = .000) with a medium effect size (r = 0.46), end-clause silent pauses (Z = -6.380 p = .000) with a large effect size (r = 0.56), mean length mid-clause silent pauses (Z = -2.960, p = .003) with a small effect size (r = 0.26), and mid-clause filled pause (Z = -2.759, p = .006) with a small effect size (r = 0.24).

No significant statistical differences were found between the monologues and dialogues in the mid-clause silent pauses (Z = -.796, p = .426), which had a very small effect size (r = 0.07), mean length of end-clause silent pauses (Z = -1.164, p = .245) with a very small effect size (r = 0.04), filled pauses/60 (Z = -.298, p = .766) with a very small effect size (r = 0.03) and the mean length of mid-clause filled pauses (Z = -1.265, p = .206) with a small effect size (r = 0.11).

Table 15 presents the median values for each measure of breakdown fluency. L2 participants produced more silent pauses/60 (*MED* = 22.91) and end-clause silent pauses/60 (*MED* = 14.06) in the monologues than in the dialogues, whereas silent pauses/60 (*MED* = 17.23) and end-clause silent pauses/60 (*MED* = 6.58) were fewer in dialogues. Moreover, the participants' mean length of mid-clause silent pauses (*MED* = 0.70) and the number of mid-clause filled pauses/60 (*MED* = 11.90) were longer in dialogues than in monologues, whereas mean length of mid-clause silent pauses (*MED* = 0.58) and number of mid-clause filled pauses/60 (*MED* = 10.09) were shorter in monologues. Thus, the hypothesis is partially confirmed that dialogue included less breakdown measures (e.g., silent pauses) than monologue. The data suggest that the mean length of mid-clause silent pauses is a core fluency feature. In dialogues, the participants paused more frequently, using filled pauses in the middle of clauses. Additionally, the duration of silent pauses within clauses was longer in dialogues than in monologues. These findings demonstrate that L2 speakers used their listening time to plan what to say next or to monitor their speech production processing. Thus, listening time is considered to be a communication strategy that benefits speakers in terms of keeping the flow of speech in dialogue.

4.3.3 Differences between Monologue and Dialogue in Repair Fluency Measures

It was hypothesised that the repair fluency in L2 learners' oral performances would be less in dialogues than monologues. Repair fluency was operationalised through five measures, namely, total repairs/60, repetitions/60, reformations/60, replacements/60, and false starts/60.

4.3.3.1 Descriptive statistics

As previously done with the speed and breakdown measures, MED, M, SD, Min., and Max. scores are reported (see Table 18) to obtain a general understanding about the characteristics of each repair fluency measure across monologues and dialogues. Generally, based on the L2 participants' mean sores presented in Table 16 below, the participants' oral performances in replacements, reformations, total repairs, and repetitions were similar in monologue and dialogue except for false starts. False starts were higher in dialogue (*M* = 0.43) than in monologue (*M* = 0.00).

Table 18. Descriptive Statistics of the Dependent Variable Repair Fluency.

Repair Fluency	Median		Mean		SD		Minimum		Maximum	
	Mono.	**Dial.**	**Mono.**	**Dial.**	**Mono.**	**Dial.**	**Mono.**	**Dial.**	**Mono.**	**Dial.**
Total Repairs/60	5.65	5.47	5.76	5.90	3.41	3.09	0.00	0.81	16.52	15.10
False Starts/60	0.00	0.43	0.69	0.62	1.00	0.74	0.00	0.00	3.88	3.29
Reformations/60	0.75	0.99	1.03	1.17	1.20	1.07	0.00	0.00	4.67	4.93
Repetitions/60	3.10	3.26	3.26	3.60	2.32	2.27	0.00	0.00	11.44	10.79
Replacements/60	0.00	0.00	0.76	0.49	0.96	0.81	0.00	0.00	3.77	2.97

(N=64). Mono= monologue. Dial.= dialogue.

To answer the following sub-question: *Are there any differences between monologue and dialogue in terms of repair fluency measures?* The Friedman test was run in order to compare the scores of repair fluency in monologues and dialogues, and to investigate whether the participants' scores were significantly statistically different on the repair fluency measures across the two task modes. The statistical tests presented in Table 19 show that there were significant differences between the repair fluency measures in the two modes**,** $\chi^2(9) = 371.03$, $p = 0.00$. In the one hand, according to the mean ranks and Friedman test results, the mean scores of reformations (RF) and repetitions (RP) were slightly higher in dialogue (RF, *M* = 4.80), (RP, *M* = 7.36) than in monologue (RF, *M* = 4.04), (RP, *M* = 6.95). On the other hand, replacements were higher in monologue (*M* = 3.58) than in dialogue (*M* = 3.16). No significant differences were found between monologues and dialogues in total repairs/60 (mono. *M* = 8.95 vs. dial. *M* = 9.07) and false starts/60 (mono. *M* = 3.56 vs. dial. *M* = 3.53). However, these statistics are not sufficient to find out where the differences lay, therefore there is a need to run further tests like the Wilcoxon test (see Table 20), as reported below.

Table 19. Mean Ranks for the Dependent Variables of Repair Fluency Measures

Repair Fluency Measures	Mean Ranks		Chi-Square	df	Asymp. Sig.
	Monologue	**Dialogue**			
Total repairs/60	8.95	9.07			
False starts/60	3.56	3.53			
Reformations/60	4.04	4.80	371.03	9	0.00
Repetitions/60	6.95	7.36			
Replacements/60	3.58	3.16			

The test statistics in Table 20 below illustrate where the differences laid between the measures of repair in two modes (monologue vs dialogue). Again, the Bonferroni adjustment was applied to avoid Type error as follows: 0.05\5= 0.01 where 0.05 presents probability value and 5 presents the number of variables compared in the analysis. The new α = 0.01 suggests that for the comparison the *p* value of repair fluency measures should be set to $p < 0.01$.

Table 20. Test Statistics for Monologue and Dialogue Differences in Repair Fluency

Repair fluency	Pairs	Z	Asymp. Sig.(2-tailed)	Effect size
Repairs/60	Dialogue vs monologue	-0.264[b]	0.792	0.02
False Starts/60	Dialogue vs monologue	-0.317[c]	0.751	0.03
Reformations/60	Dialogue vs monologue	-0.832[b]	0.405	0.07
Repetitions/60	Dialogue vs monologue	-0.806[b]	0.420	0.07
Replacements/60	Dialogue vs monologue	-1.738[c]	0.082	0.15

Note. *b* is based on positive ranks; *c* is based on positive ranks.

Table 20 indicates that there were no statistically significant differences between the two modes in all repair measures: repairs/60 (Z= -.264, p=.792) with a small effect size (r = 0.02), false starts/60 (Z= -.317, p=.751) with a small effect size (r = 0.03), reformations/60 (Z= -832, p=.405) with a small effect size (r = 0.07), repetitions/60 (Z = -806, p=.420) with a small effect size (r = 0.07), and replacements/60 (Z= -1.73, p=.082) with a small effect size (r = 0.15).

Thus, these results suggest that the participants did not statistically differ in their repair fluency across the monologue and dialogue tasks, meaning that the participants performed similarly in terms of repairs in both modes. Additionally, the median values (SeeTable18) between the two modes were very similar in repair fluency. For example, in total repairs/60 *MED* = 5.65 in monologues and *MED* = 5.47 in dialogues, in

repetitions/60 *MED* = 3.10 in monologues and *MED* = 3.26 in dialogues, and in replacements/60 *MED* = 0.00 in monologues and *MED* = 0.00 in dialogues.

4.4 WMC and L2 Utterance Fluency in Monologue and Dialogue

4.4.1 Multiple Regressions for WM Predicting Oral Performance

Following the analyses of monologue and dialogue Means in terms of L2 utterance fluency measures, multiple regression analyses were conducted to answer the second question: *To what extent can variations in WMC scores predict utterance fluency in monologue and dialogue?* The hypothesis predicted that L2 learners who had a higher WMC would be more fluent than low WMC leaners in monologic and dialogic performances. Regression analysis was used to calculate the statistical significance of the participants' results in the individual independent variables and in the model itself. Regression analysis is a statistical test that is used to know the proportion of variance in the dependent variables that can be explained by the independent variables (Pallant, 2013). In this study, WMC was an independent variable and utterance fluency measures in monologue and dialogue were employed as continuous dependent variables. Multiple regression was used to examine how much variance in the L2 fluency measures could be explained by WMC in monologue and dialogue. WMC was measured by OST and BDST, which gave different estimates of WMC. The OST scores ranged between 22 and 54 (*M* = 40.84, *SD* = 6.71) while the BDS scores ranged between 3 and 6 (*M* = 4.10, *SD* = 1.02). Further details were discussed in the discussion chapter.

Regarding the dependent variables, the composite measures of speed fluency, breakdown fluency and repair fluency in monologue and dialogue were employed. The

composite measure of speed fluency comprised the length of speech samples, number of syllables, speech rates, articulation rates, phonation time ratio, and mean length of syllables. The composite measure of breakdown fluency contained silent pauses/60, mid-clause silent pauses/60 (mid-CSPs), end-CSPs/60, mean length of mid-CSPs, mean length mid-cause SPs pauses (mid-CSPs), filled pauses/60, mid-filled pauses/60, and mean length mid-CFPs. Finally, the repair fluency incorporated measures of total repairs/60, false starts/60, reformations/60, and repetitions/60.

Prior to testing the main hypothesis, multiple regression assumptions were examined, i.e., normality of distribution, linearity, and collinearity (Field, 2013; Pallant, 2013). The first test examined whether the outcome variables (i.e., composite measures of speed fluency, breakdown fluency and repair fluency in monologue and dialogue) were normally distributed. Skewness and kurtosis values are reported in Table 21. There was no violation in terms of the distribution of the dependent variables (utterance fluency) as the skewness and kurtosis values fell within a “very good” symmetry of a normal distribution, between +1 and -1 (West et al.,1995). Only the values of monologue breakdown were slightly outside of the recommended threshold values of skewness and kurtosis – 3.94 and 23.92, respectively.

Table 21. Descriptive Statistics and Tests of Normality

					Skewness		Kurtosis	
Variables	**Min**	**Max**	**Mean**	**SD**	**Statistic**	**SE**	**Statistic**	**SE**
1. Monologue speed	35.43	10.72	65.48	14.78	0.64	0.30	0.20	0.59
2. Dialogue speed	40.00	13.66	77.36	21.00	0.70	0.30	0.11	0.59
3. Monologue breakdown	36.70	24.03	73.66	26.45	3.94	0.30	23.92	0.59
4. Dialogue breakdown	22.09	11.72	62.17	21.34	0.35	0.30	-0.30	0.59
5. Monologue repairs	0.00	33.04	11.52	6.82	0.76	0.30	0.66	0.59
6. Dialogue repairs	1.62	30.21	11.80	6.20	0.68	0.30	0.61	0.59
7. Operation span test	22.00	54.00	40.84	6.71	-0.64	0.30	0.42	0.59

8. Backward digit span	3.00	6.00	4.11	1.03	0.41	0.30	-1.03	0.59

In addition, the residuals were analysed in terms of normality; this is critical for confirming the results emerging from regression analysis. Residuals should be independent of one another to validate the results of regression analysis. The term 'residual' refers to the difference between the observed values of the outcome variables and the values of the predictor variables. It is generally determined by a scatter plot produced from the statistical analysis of the regression test. The expectation is that there should be no values falling outside the range of ±3 (Tabachnick and Fidell, 2013). According to this rule, there were no issues related to the residuals because all the values were between ±3 (see Figure 9).

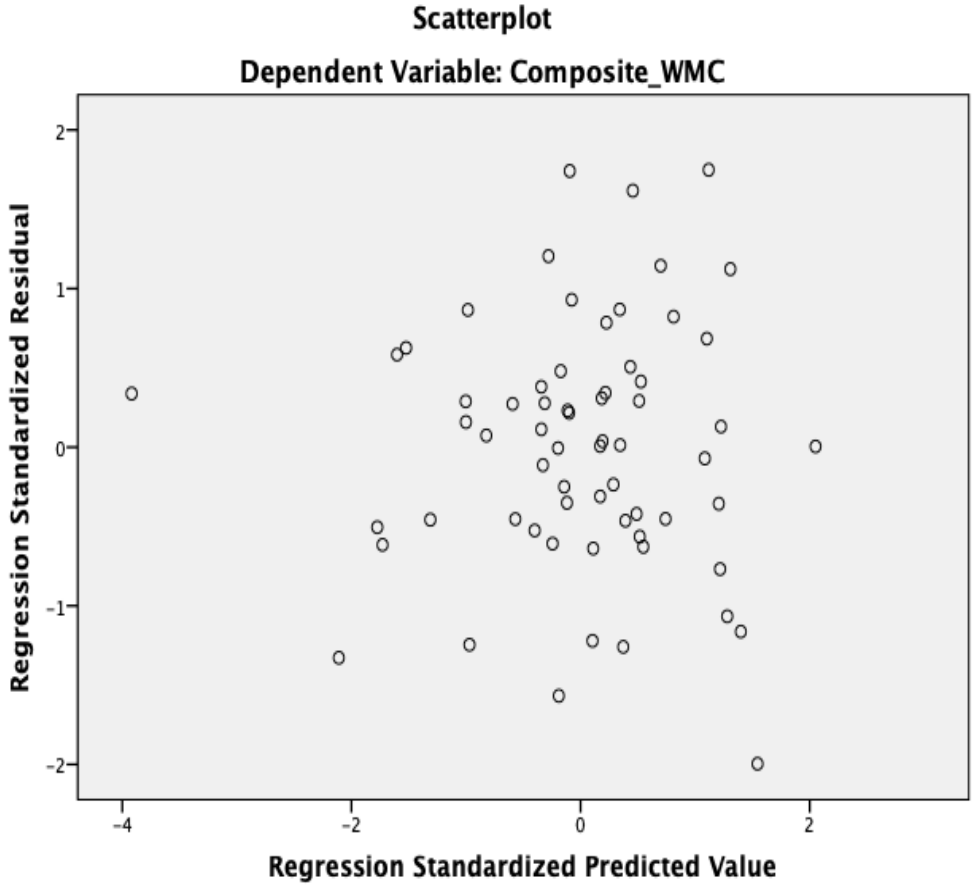


Figure 9. The Residuals' Values Fall between ±3

The second assumption was a correlational analysis that was carried out to explore the linear relationship between the WMC and utterance fluency measures. The linearity between the independent and dependent variables is an important assumption of regression analysis. Thus, Pearson correlation coefficients between composite utterance fluency measures in monologues and dialogues and WMC were computed. The results of this test are presented in Table 22. The results demonstrate that OST and BDST were not statistically significantly correlated with the composite measures of utterance fluency in monologue and dialogue. As can be seen from Table 20, the correlation coefficients ranged from $r = 0.00$ to $r = 0.18$.

Table 22. Correlations between Utterance Fluency Measures in Monologue and Dialogue and WMC

Variables	1.	2.	3.	4.	5.	6.	7.	8.
1. Monologue speed	1							
2. Dialogue speed	.64**	1						
3. Monologue breakdown	-.36**	-.21	1					
4. Dialogue breakdown	-.26*	-.23	.31*	1				
5. Monologue repairs	.04	.03	.25*	-.05	1			
6. Dialogue repairs	-.02	.10	.06	.32*	.19	1		
7. Operation span test	-.02	.00	-.03	-.05	.12	.03	1	
8. Backward digit span test	.15	.17	-.21	-.06	-.09	.09	.18	1

**. $p < 0.01$; *. $p < 0.05$

The final assumption of regression analysis is multicollinearity. There should be no issues of multicollinearity among the independent variables (Field, 2013; Pallant, 2013). It is important to ensure that the independent variables are not highly correlated

with each other to discriminate the relationships between each independent variable (WMC tests) and dependent variable (utterance fluency measures) (Pallant, 2013). The rule of thumb is that correlations between the independent variables that are ≥ .80 raise issues of multicollinearity (Pallant, 2013; Tabachnick and Fidell, 2013). Thus, multicollinearity was examined using Variance Inflation Factor (VIF) and tolerance tests. VIF shows if there is a collinear relationship between independent variables. According to Kutner et al. (2005), threshold values for VIF of at least 5 and tolerance statistics of less than 0.2 signify issues of collinearity. Thus, in this study no issues of multicollinearity were detected in this analysis as VIF (1.03) and tolerance (0.97) did not conflict with the above-mentioned threshold values.

4.5 WMC and L2 Fluency in Dialogic and Monologic Performances

After testing all assumptions of the regression analysis, WMC tests were set as predictors while monologue and dialogue utterance fluency served as outcomes. Six separate multiple regression analyses were run (see Table 23) to examine the predicted roles of OST and BDST in explaining the variance in the oral performances of L2 utterance fluency across monologue and dialogue. To assess the effect size (R^2) eta square, the criterion recommended by Cohen (1988) was used, where r^2 = 0.14 reflects a small effect size, r^2 = .39 reflects a medium effect size, and r^2 = .59 and above reflects a large effect size.

Table 23. Summary of Multiple Regression Analysis Predicting Utterance Fluency in Monologue and Dialogue from the WMC

Outcome	Predictor	B	SE	β	t	Sig.	*R*	R^2
		F (2.63) = 0.80, *p* = 0.45						
Monologue speed	OST	-1.12	2.77	-0.05	-0.40	0.69	0.16	0.02
	BDST	22.72	18.13	0.16	1.25	0.21		
Monologue breakdown		*F* (2.63) = 1.42, *p* = 0.25						

	OST	0.01	0.50	0.00	0.03	0.98	0.21	0.05
	BDST	-5.46	3.28	-0.21	-1.66	0.10		
		F (2,63) = 0.79, *p* = 0.46						
Monologue repairs	OST	0.14	0.13	0.14	1.05	0.30	0.16	0.03
	BDST	-0.74	0.86	-0.11	-0.87	0.39		
		F (2.63) = 0.92, *p* = 0.40						
Dialogue speed	OST	-0.84	3.95	-0.03	-0.21	0.83	0.17	0.03
	BDST	35.09	25.87	0.17	1.36	0.18		
		F (2.63) = 0.16, *p* = 0.85						
Dialogue breakdown	OST	-0.13	0.41	-0.04	-0.31	0.75	0.07	0.01
	BDST	-1.09	2.70	-0.05	-0.40	0.69		
		F (2.63) = 0.27, *p* = 0.76						
Dialogue repairs	OST	0.01	0.12	0.02	0.12	0.91	0.09	0.01
	BDST	0.55	0.78	0.09	0.70	0.49		

The results presented in the above Table show that all the models were not statistically significant regarding speed, breakdown, and repair fluency. With regard to the individual contribution of the independent variables (OST and BDST), they failed to contribute significantly to all models and to explaining the variation in the participants' oral performances in both modes. The p-values of all models were greater than 0.05. For example, for monologue speech rates *F* (2.63) = 0.80, *p* = 0.45, therefore the *p*-value was larger than 0.05.

Additionally, the amount of variance explained in the models is as follows: in model 1, the outcome variable is *monologue speed fluency* and the WMC tests (OST and BDST) explain 2% of the variation in the participants' speed fluency (r^2 = 0.02, small effect size). For model 2, *monologue breakdown fluency* is considered as the outcome variable, and the WMC tests explain 5% of the variation in the participants' breakdown fluency (r^2 = 0.05, small effect size). For model 3, *monologue repair fluency* is the

outcome variable, and the WMC tests explain 3% of the variation in the participants' repair fluency (r^2 = 0.03, small effect size). For model 4, the *dialogue speed fluency* is the outcome variable, and the WMC tests explain 3% of the variation in the participants' speed fluency (r^2 = 0.03, small effect size). For model 5, *dialogue breakdown fluency* is the outcome variable, and the WMC tests account for 1% of the variation in the participants' breakdown fluency (r^2 =0.01, small effect size). Lastly, in model 6, where *the dialogue repair fluency* is the outcome variable, the WMC tests account for 1% of the variation in the participants' repair fluency (r^2 = 0.01, small effect size). Thus, it can be concluded that the amount of variance explained in all models was not large enough to reach significant levels in the regression analysis.

4.6 Conclusion

To conclude this section, the non-parametric Wilcoxon signed-rank test was used to find the significant differences between dialogic and monologic performances in terms of speech, breakdown and repair fluency. Additionally, multiple linear regressions were computed to examine the predictive power of WMC, as measured by OST and BDST, on the L2 utterance fluency measures of speed, breakdown and repair in monologic and dialogic tasks. The findings of the current study suggested that dialogue was fluent than monologue in most of the speed and breakdown fluency measures. Additionally, no significant differences were found in the performance of monologue and dialogue in terms of repair fluency measures.

In line with the previous studies (e.g., Fortkamp, 2000; Mizera, 2006; Kormos and Trebits, 2011; Awaad, 2017; Georgiadou and Roehr-Brackin, 2017) that studied WMC in relation to L2 processing and production, there were no significant correlations between WMC and oral fluency in monologue and dialogue. The data suggested that

individual differences in working memory capacity could not predict the variations in the L2 oral performance of monologue and dialogue. Finally, future studies are recommended on WMC and L2 oral fluency in two modes to uncover the power of WMC using different kinds of WMC measures. It can be suggested that including L1 fluency measures is necessary to find the reason of disfluency, whether it is a matter of personal speaking styles or individual differences in WMC.

Chapter 5: Discussion

5.1 Introduction

This chapter discusses in greater detail the findings that are presented in the results chapter. Then, the practical and theoretical contributions of this study to L2 fluency research will be discussed in detail. The first research question sets out to examine the differences between monologic and dialogic performances in terms of speed, breakdown, and repair fluency. In addition, the aim of the second research question is to test the power of the participants' working memory (WM) in predicting L2 oral fluency performance in dialogue and monologue.

5.2 Summary of Key Findings

This study was mainly designed to examine L2 utterance fluency in monologue and dialogue. It sought in an original way to examine the relationship between L2 oral fluency performance in dialogue and monologue and working memory capacity (WMC). Additionally, the current study aimed to address and evaluate the differences between the operationalization of L2 utterance fluency measures in two modes,

dialogue and monologue. Thus, this study hypothesized that L2 participants would be more fluent in terms of speed, breakdown, and repair in L2 dialogue tasks than L2 monologue tasks because dialogue is an interactive task where two speakers share the responsibility of maintaining the flow of speech. Moreover, the current study predicted that WMC can explain the variance in the oral fluency performances of L2 participants in the two modes because WM directs and distributes attentional resources to link form and meaning. To summarize the results presented in the previous chapter, two research questions and sub-questions and their related main findings are summarized below.

RQ1: Are there any differences between monologue and dialogue in terms of utterance fluency?

Speed, breakdown, and repair are the three sub-dimensions of L2 fluency (Skehan, 2003) that are used in the current study to capture different aspects of L2 utterance fluency in monologue and dialogue. The results are clearly in line with the existing findings of, for example, Riggenbach (1998), Michel et al. (2007), Michel (2011), Sato (2014), Witton-Davies (2014), Kirk (2016), Tavakoli (2016), and Peltonen (2017) that L2 participants are significantly more fluent in discussion dialogues than in opinion monologues for most fluency measures except for repair fluency (e.g., Huensch and Tracy–Ventura, 2017).

RQ1a: Are there any differences between monologue and dialogue in terms of speed fluency measures?

There are significant differences between monologic and dialogic performances in terms of speech rates, the length of speech samples, the number of syllables, and

phonation-time ratios. The L2 participants spoke faster in dialogue as demonstrated by higher speech rates, longer speech samples, a greater number of syllables, and higher phonation-time ratios. As for the data obtained from the monologues, the L2 participants had lower speech rates (SRs), shorter speech samples, fewer syllables, and lower phonation-time ratios (PTRs). In addition, there were no significant differences between monologues and dialogues in terms of articulation rates (ARs) and mean length of syllables.

To answer the above research questions in greater length, it is necessary to discuss each measure of each aspect of utterance fluency.

5.3 Speed Fluency

Speed fluency, including SRs, ARs, PTRs, and mean length of runs (MLRs), is an essential aspect of fluency because it is a clear and less ambiguous fluency measure than breakdown and repair measures (Kahng, 2014) because they can be easily calculated and identified from the transcripts. For example, SR is calculated by dividing total number of syllables by the total speaking time including pausing time. Speed fluency represents the articulation phase of speech production processing where the speech sound is ready to be produced (Lambert, 2017). The analysis of the dialogue and monologue speech samples of the current study indicates that dialogic task performance was fluent than monologic tasks in terms of SR and PTR measures. Having different speech rate values for the same speaker in two different tasks can be caused by the variation in the number of pauses (de Jong et al., 2012; Witton-Davies, 2014). In other words, the more a speaker pauses, the slower their SR will be (de Jong, 2016).

Furthermore, speech sample length and the number of syllables were higher in the dialogic task, whereas the MLS was not significant but still slightly higher in monologue (*M* = 1.66) than in dialogue (*M* = 1.34). The effect size associated with this difference was small (*r* = 0.21), so we fail to reject the null hypothesis that there are no differences between monologic and dialogic performances in terms of MLS. The higher MLS in monologue can be explained by the fact that a monologue is usually one long uninterrupted turn, while a dialogue consists of short turns (de Jong et al., 2012). Thus, the MLR is influenced by duration of turn (Witton-Davies, 2014). Additionally, although the articulation rate (AR) was not statistically significant between the two modes, it indicates that the more syllables per minute of speaking time, the faster the speaking rate. Thus, AR as a pure measure of speed confirmed that the participants are faster in dialogic task performance. The non-statistically significant results of MLS and AR across the two modes are in line with Witton-Davies's (2014) study, which showed that MLS was not significant across modes and did not vary by genre, indicating that the presence of an interlocutor may not influence the performance of both variables in narrative tasks. This might be due to the small sample size (64 participants), which could influence the non-significant results.

It can be argued that in the monologic task there are fewer opportunities for language learning to be improved compared to in the dialogic task because of the reduced input of new feedback and information which are included in the monologic task. Additionally, based on the literature, in a monologue the speech production processing is planned and conceptualised by a single speaker and thus is only possible with self-monitoring (Kormos, 2006; Michel, 2011). This may place a great deal of pressure on the attentional resources of the WM, which is responsible for regulating and organizing the language in the long-term memory. By contrast, in dialogue the speakers can plan

and conceptualize the speech during the partner's turn. This helps to reduce any unnecessary pauses or hesitations during the L2 speech production (Tavakoli and Foster, 2008, Michel, 2011). For example, in this study it could be possible that the L2 speakers divided their attention between planning, formulating, articulating, and monitoring their speech in the monologic task while they are speaking, whereas in the dialogic task L2 speakers allocated their attention to planning their own turns during their partners' turns and to using communicative strategies (e.g., pauses, repetitions) to sustain the flow of dialogue. Therefore, the dialogic performance is faster than the monologic one.

Speaking L2 is to a very large extent not an automatic process (de Jong and Perfetti, 2011) and L2 speakers therefore depend on their attentional resources to retrieve lexical grammatical knowledge. Depending on declarative knowledge alone requires extensive semantic phonological knowledge, with more access fluidity and attentional control. This is possible for native speakers or highly advanced language learners who are able produce language automatically. Based on the literature, less fluent intermediate L2 speakers in monologue depend more on their declarative knowledge than dialogue where they have less access fluidity to lexical information (Segalowitz, 2010). Whereas, in dialogue they produce more fluent speech due to the collaborative nature of the dialogic task.

Furthermore, consistent with the literature, this research found that collaborative speech and exchange roles in dialogue facilitate speech production processing to allow the speakers to speak more fluently than in monologue (Garrod and Pickering, 2004). The goal of dialogue is that both the speaker and the interlocutor reach a common conception or understanding of what they are talking about, or the

conversation will fail (Garrod and Pickering, 2004). Achieving a common perception in dialogue requires speakers to accommodate and relate their choices of syntactic and phonological structures to the ongoing dialogue (Kootstra et al., 2009). This leads speakers to use each other's vocabulary and linguistic structures and it limits the interlocutor's selections of language production (e.g., changing the topic) (Kootstra et al., 2009). The example below was taken from the dialogic samples of this study showing that speaker 2 hesitated or stopped talking during the conversation (see lines 7 and 8) with the speaker 1 (partner). This might be because the speaker 2 did not know the right vocabulary, or she was still planning her speech while the other partner (speaker 1) completed speaker 2's turn (in lines 9, 10, 11) in order to keep up the flow of speech. Thus, they worked together to solve their language problems and keep the talk flowing. Line 7 and 8 suggested that speaker 2 used speaker's 1 words and idea(in lines 5 and 6) to express her opinion about women driving. This could be an example of alignment during the pair interaction task.

Speaker 1:

1 *"Because this the decision would help a lot of their uh in daily life ummm may be a*

2 mother earlier would drive her children to school instead of a foreign driver to

3 pick her umm children and umm and go to the the school. That is not safe"

Speaker 2 :

4 *yeah. I agree I know many children got harassed by driers.*

Speaker 1:

5 *Of course uh for young children it is safer to be picked by tier moms, what do think*

6 about women driving?"

Speaker 2 :

7 women or mothers have a lot of benefit of driving.It is safer for their children to be

8 with their mother when they go to school by the car. Also women uhhh can do….

Speaker 1:

9 *Yea a mother uhhh She can do it herself and she can umm and she can go to the*

10 mall shop uhhh may be do the smallest thing like pick up a cup of coffee it was

11 hard thing to do in the in the past now it is more convenient."

Additionally, the current study findings are in line with empirical studies that compared L2 fluency in monologue and dialogue (e.g., Michel et al., 2007; Gilabert et al., 2011; Michel, 2011; Witton-Davies, 2014; Sato, 2014; Tavakoli, 2016; Peltonen, 2017). Some of these have studied utterance fluency in two modes monologue and dialogue in terms of task complexity, native speakers' judgement, and cognitive hypothesis. For instance, Michel's (2011) study, supported the current study findings of higher speech rates and fewer filled pauses in dialogue than in monologues, when comparing two modes in terms of their complexity, accuracy, and fluency. Similarly, Witton-Davies (2014), Tavakoli (2016, 2018), and Peltonen (2017) found that dialogic tasks were more fluent than monologic tasks in terms of speech rate and AR.

Also, the finding here regarding speed fluency broadly supports the work of other studies in this area (e.g., Tauroza and Allison,1990; Gilabert et al., 2011) demonstrated that L2 participants were faster in dialogue than monologue in terms of pruned and unpruned speech rates. In accordance with the present study results,

Gilabert et al. (2011) also showed that native speaker raters demonstrated that L2 participants were faster in pair-interaction tasks than in monologue tasks in terms of speech rates.

Measures of speech and AR are well known in the field of L2 fluency, showing the differences across learners' levels of fluency (e.g., Riggenbach, 1991; Kormos and Dénes, 2004), or levels of L2 proficiency (e.g., Peltonen, 2017; Tavakoli et al., 2020). Recently, Os et al. (2020) suggested that speech rate is an essential measure to determine individual fluency and fluency perception in native and non-native speech. It also seems to be the best choice for L2 research fluency to differentiate between fast and slow speakers' rates (Witton-Davies, 2014). The influential study of Kormos and Dénes (2004) demonstrated that high-proficiency learners showed higher levels of L2 speech fluency than lower-intermediate university students in SP, MLR, PTR, and mean length of pause. More recently, Tavakoli et al. (2020) examined four assessed proficiency levels of learners (A2, B1, B2 and C1 CEFR) using temporal fluency measures of speed, breakdown, and repair. The results showed that these fluency measures were usually able to differentiate learners across proficiency levels. For example, speed and composite measures were able to differentiate the three proficiency levels of A2, B1, and B2, while breakdown measures, such as the length of silent pauses, were able to distinguish single low-proficiency level participants (A2) from higher proficiency level participants. As for the frequency of silent pauses, this was able to distinguish two low-proficiency levels (A2 and B1). Finally, for repair fluency, filled pauses and repairs were used by higher proficiency levels to build and punctuate their speech fluency.

RQ1b: Are there any differences between monologue and dialogue in terms of breakdown fluency measures?

On the one hand, the dialogues were associated with fewer silent pauses/60 with a medium effect size (r = 0.46) and with end-clause silent pauses/60 with a large effect size (r = 0.56). The medium to large effect sizes lead us to reject the null hypothesis that there is no difference between monologue and dialogue in breakdown fluency. Also, dialogues had a longer mean length of mid-clause silent pauses with a small effect size (r = 0.24) and more mid-clause filled pauses/60 with a small effect size (r = 0.24). On the other hand, monologues were associated with higher numbers of silent pauses/60 and end-clause silent pauses/60. Additionally, monologues had a shorter mean length of mid-clause silent pauses/60, and fewer mid-clause filled pauses/60. Additionally, participants in dialogues had more frequent mid-clause silent pauses/60 and filled pauses/60, yet a shorter mean length of mid-clause filled pauses/60 compared to those in monologues. However, there were no significant differences found between monologic and dialogic performances in the following: mid-clause silent pauses/60 (r =0.07), mean length of end-clause silent pauses/60 (=0.04), filled pauses/60 (r =0.03), and mean length of mid-clause filled pauses/60 (r = 0.11). These non-significant results are due to very small effect sizes that fail to reject the null hypothesis. Thus, overall, dialogue was more fluent than monologue, in particular in L2 SRs/60, length of samples, number of syllables/60, mean length of mid-clause SPs/60, and mid-clause FPs/60.

5.4 Breakdown Fluency

5.4.1 Filled and Silent Pauses

In this study, breakdown fluency sets out to analyse both silent and filled pauses in terms of location, frequency, and length. It is important to first generally explain some reasons for pauses in L2 oral performance. Pauses are produced for: (1) speech planning, to think about the subsequent performance (Tavakoli and Wright, 2020); (2) lexical retrieval; and (3) the monitoring stage of speech processing (Witton-Davies, 2014). Thus, pauses are necessary in speech, whether they are filled or silent. Some L2 individuals tend to use silent pauses or filled pauses without differentiating between them; for example, they use filled pauses as an alternative to silent pauses and this could be related to a speaker's speaking style (Towell et al., 1996; Witton-Davies, 2014; de Jong et al., 2015). Interesting findings from Leonard's (2015) longitudinal study showed no significant gains regarding the measurements of first and second language filled/unfilled pauses and hesitations during a study abroad period. This means that the pausing phenomenon is not related to the L2 proficiency level, but to individuals' speaking styles. In addition, a recent study by Shea and Leonard (2019) found a weak relationship between both filled and silent pauses, and participants' L2 proficiency level.

Filled pauses can be a communication strategy that users may employ to help them recall and deliver their intended speech (Witton-Davies, 2014). Carroll (2004) found that less fluent speakers use repetitions and filled pauses in conversation to reflect 'interactive achievements' that can maintain the flow of speech. Traditionally, pauses have been considered as examples of disfluency, but recently filled pauses have been shown to have a communicative role, for example, in showing difficulty retrieving

words, utterances, or ideas or showing thinking and hesitation during speech processing (Tavakoli and Wright, 2020). Interestingly, filled pauses can assist fluency rather than dysfluency. For example, FPs may help speakers to deal with the pressure of processing time by avoiding long silent pauses and introducing lexical or non-lexical filled pauses (Kahng, 2014).

In this study, it can be argued that filled pauses play an important role in an interactive task, meaning that the interactive and collaborative nature of a dialogue encourages L2 speakers to perform more fluent speech and sustain the smooth flow of the L2 speech during the interaction. They do this by producing filled pauses, such as *umm, you know, uhh, like* (Tavakoli et al., 2020).

Similarly, Sato (2014) argues that pauses in dialogue might be used to buy some time to think of suitable phrases, whereas in monologue pauses are considered as hesitation in the speaker's speech. It has been argued in SLA research (e.g., Skehan, 2014b) that pauses are necessary in L1/L2 speech production processing and that these pauses can be explained as a monitoring process (Levelt, 1989) that less advanced L2 speakers use in their oral performance. This is in accord with Tavakoli's (2011; 2018) study indicating that frequent pauses are used as self-monitoring and formulation processes to pay attention to the speech form with regard to accuracy. Therefore, in the current study, the higher frequency of filled pauses and their location in the dialogue can be suggested to help the dialogic task to be more fluent than the monologic task.

It is also possible to argue that the repeated use of filled pauses in dialogue is an example of taking a partner's needs into consideration (Tavakoli, 2016). For example, sentence five below shows the repeated use of *you know*, which can work as a

confirmation check of the listener's understanding. Sometimes a speaker intentionally uses disfluencies to assess the listener's understanding, to ask for attention, or to show empathy (Witton-Davies, 2014).

There are, however, other possible explanations by Crible and Pascual (2020), showing that in the fluent speech of English, Spanish, and French, filled pauses and repetitions usually co-occur with silent pauses at the beginning of a speech turn in a dialogue sample (see examples 1 and 4 below). With this in mind, additional qualitative analysis is recommended to reveal the additional functions of lexical filled pauses, non-lexical filled pauses, and repair fluency in conversation.

1 *Uhh (0.89) (0.28) yeah (1.46). And also all of my friends (0.55) uhh (0.39) are driving*
2 and one of my sisters. And uhh (0.58) (1.08) yeah for me the advantage is uhh (0.29)
3 to be (0.27) um (0.67) I can uhh (0.54) (0.5) do my own works to myself (0.44)
4 Uhh(0.31) (0.5) not asking anyone. Um (0.55) (0.69) the disadvantages, I think.

Sentences from 1 to 4 are from current L2 dialogue samples where the duration of filled pauses is enclosed in blue parentheses, whereas orange parentheses represent the duration of silent pauses. Line 1, which is the beginning of the turn, shows a filled pause *(uhh)* followed by a silent pause (0.28). This could indicate that some learners use pauses when they face problems with formulating the intended message during speech (Kormos, 2006; Michel, 2011; de Jong et al. 2013; Kahng, 2017).

In terms of the number of SPs, dialogue was associated with fewer SPs. This result also supports the work of (Tavakoli, 2010), who found that silent pauses decrease the speech flow. It is possible to tentatively speculate from these findings that the interactive nature of dialogue encourages participants to produce fewer silent pauses because the interlocutor uses the partner's turn to plan and think about what to say next. Instead of using silent pauses, L2 speakers use more filled pauses in the dialogic

tasks to tell the interlocutor that they are still holding the turn and that they are about to speak.

Thus, qualitative questionnaires, e.g., stimulated recall, are recommended to understand the reasons behind the more fluent speech in L2 dialogic tasks compared to those in L2 monologic tasks. Moreover, stimulated recall can help the researcher to know more about the use of silent and filled pauses during speech. For example, pausing might have been used to keep the flow of speech, hold the floor or think of what to say next. Pausing might be used because the speaker does not have anything to say. de Jong et al. (2013) argues that pausing can be related to other factors such as individual differences. It can therefore be assumed that some L2 fluency measures, such as filled pauses, silent pauses, and repairs, are related to L1 fluency (Derwing et al., 2009; Segalowitz, 2010; de Jong, 2016). Including L1 fluency measures from the same participants is also recommended for future study to examine whether the L1 pausing phenomena affect pauses in L2. Taking L1 fluency measures into account could reveal whether disfluency is caused by personal speaking styles or L2 proficiency.

One source of limitation in this study that could have affected the results of FPs/SPs was that the instructions for the monologic and dialogic tasks were slightly different. For example, vocabulary items were provided in monologic task but statements and guiding questions were added to the dialogic task. Additionally, the two task modes were not identical in terms of the kind of speaking function (argument vs discussion).

5.4.2 Pause Location

Pause location is an essential element to consider when measuring pauses because it refers to a particular stage in L1 (Levelt, 1989) and L2 (de Bot, 1992) speech

production models. Internal pauses are mid-clause silent/filled pauses, whereas external pauses are end-clause silent/filled pauses. Previous L2 fluency researchers have indicated that pause locations are one of several temporal variables that affect fluency performance. For example, mid-clause silent/filled pauses were found to be most the important variable when measuring fluency (Witton-Davis, 2014; Peltonen, 2020). Pauses location have also been shown to influence native speakers' judgements on speech fluency (e.g., Skehan et al., 2016).

Furthermore, Witton-Davis argues that mid-clause pauses affect fluency more than end-clause pauses. This could be explained by the fact that native speakers have more external pauses than internal pauses, as internal pauses are found more in non-native speech, to allow the speaker to recall linguistic knowledge (Foster and Tavakoli, 2009). Since L2 is not an automatic process, L2 speakers pause to think of suitable words or to revise their clauses before completing the oral production processing. This seems to be consistent with de Jong's (2016) findings, which showed that L2 participants paused longer and more frequently within clauses than L1 speakers. In reference to Skehan et al.'s (2016) study, the comparison between native and non-native speakers indicates that pauses at the end of analysis of speech unit (AS-unit) represent the stage of conceptualization, while pauses that take place within ASU represent lexical retrieval that is related to formulator and articulator processing.

5.4.3 Mid- and End-Clause Silent Pauses

The majority of language learning studies (e.g., Tavakoli, 2011; Skehan et al., 2016; de Jong, 2018; Lambert et al., 2017; Peltonen, 2020) assessed the performance in each of the three stages of L2 speech production model by L2 fluency measures (breakdown, speed, and repair fluency). They have demonstrated that pause locations

reflect difficulties in speech production processing and are indicative of additional efforts put into planning for utterances. For example, for L1, mid-clause pauses occur before uncommon words and are associated with retrieval of lexical information, while L1 end-clause pauses are related to planning for the next clause. In contrast, in L2 speech, mid-clause pauses are associated with content planning and lexical encoding. Pauses at clause boundaries indicate long-time planning for lexical retrieval and grammatical organization of L2 utterances, so they are related to the formulator and/or articulator stages (Skehan et al., 2016).

Furthermore, mid-clause pauses have been found to be correlated with perceived fluency and L2 proficiency. For example, native speakers' judgments on participants L2 utterance fluency have been found to be strongly correlated with mid-clause pauses (Kahng, 2014). Also, there is a positive correlation between mid-clause pauses and low L2 proficiency participants (Shea and Leonard, 2019; Tavakoli et al., 2020).

de Jong (2018) argues that mid-clause FPs and the length of mid-clause SPs indicate lexical retrieval linked to the formulation stage. Moreover, dialogues allow the creation of new ideas and sometimes new topics, which require need more time for the planning for the utterances (Michel, 2011).This could be an indication that the interlocutors spent more time during their partners' turns in discussion dialogue on planning their utterances. Thus, the collaborative nature of the dialogues made the participants' oral performances more fluent than in monologues because in dialogue there is pressure on the speaker from the interlocutor to keep the conversation going and to avoid long interrupted pauses.

Additionally, in the current study, there were significant differences in the mean length of mid-clause silent pauses in both modes; mid-clause SPs were longer in dialogue

than monologue (dialogue *M* = 4.37 vs. monologue *M* = 3.43) with a small effect size (*r* = 0.26). However, there were no significant differences in the number of mid-clause silent pauses. This result corroborated Peltonen's (2017) findings of a non-significant difference in mid-clause SPs between monologue and dialogue across two groups of L2 proficiency levels (G1 and G2). It was slightly unexpected that the difference in mid-clause SPs in both modes did not reach statistical significance. This finding could be explained by the trade-offs between the different elements of language proficiency. For instance, in Peltonen's (2017) study, high proficiency participants (G2) possibly attempted to produce more complex language, resulting in a higher processing load and more mid-clause pauses. She argues that in order to confirm this hypothesis, an analysis of complexity and accuracy is needed. Another possible reason for non-significant mid-clause SPs is that both groups G1 and G2 had very similar scores for mid-clause silent pauses. In the current study, the mean score of the dialogic performance was *M* = 10.09, whereas the mean score of the monologic performance was *M* = 10.05 with a very small effect size (*r* = 0.07). This led to the rejection of the null hypothesis that there is a difference between monologue and dialogue in mid-clause silent pauses.

Regarding the significant findings of the mean length of mid-clause silent pauses, these are consistent with Peltonen's (2017) study, who found longer mid-clause SPs in the L2 dialogic task. A possible explanation for longer mid-clause silent pauses in dialogue is that less proficient L2 speakers tend to depend mostly on longer silent pauses to develop and produce utterances, which are linked to the processing of formulation and articulation. In contrary, L2 proficient speakers tend to use repair and filled pauses to build and punctuate utterances. As for the dialogic performance in this study, the L2 participants produced longer L2 mid-clause silent pauses. This is

possibly, as mentioned before, because L2 is not an automatic process and dialogue allows more ideas to be created during the speaking process (Tavakoli et al., 2020). Therefore, participants spend more time looking for suitable lexical information and linking this to the most appropriate phonological sound. It can be assumed that L2 participants treat silent pauses as filled pauses to sustain the speaker's turn during L2 speech processing. More broadly, qualitative research is also needed to determine the reasons behind the speakers' pauses in dialogue.

The sentences below (1-5) are from L2 participants' data in dialogue to show silent pauses (duration > 0.25 seconds) that are inserted in the middle of clauses or phrases instead of at the end of clauses. These are in orange parentheses. The possible explanation for these silent pauses is that they are used to buy time to recall the lexical and linguistic knowledge required to formulate and produce speech (Götz, 2013). They are related to the formulator and articulator stages of speech processing (Skehan et al., 2016).

1 *Nowadays many ca (0.27) companies and business owners (0.46) rely on social*
2 *media Influencer (1.46). And this is (0.51) increase fraud and all of social media*
3 *influencer (0.55). It will be (0.28) um (0.91) work first for me and uh (0.42) um (1.06)*
4 *you know uh easy for me and so on (0.44). Um (1.04) uh (1.05) and it's an easy*
5 *way to (0.39) to uh to be a famous.*

Here, for example, in line five the mid-clause filled pauses seem to indicate that the speaker is holding her turn to retrieve lexical information. Also, the speaker uses filled pauses (um and uh) in combination with mid-clause silent pauses enclosed in orange parentheses, possibly to allow more time for the formulator and the articulator to produce utterances. When speaker hesitates or stops talking, it may she not know the right vocabulary, or still planning for the speech (Peltonen, 2017). In lines 2 and 4, the silent pauses enclosed in black parentheses are examples of end-clause pauses or

what are known as grammatical juncture pauses (Goldman-Eisler, 1968), which are connected to the conceptualiser. Thus, this could be indicative of planning for the next sentence because the idea of the sentence or the intended speech is completed and followed by end-clause silent pause.

5.4.3.1 Mid-clause filled pauses

There were significant differences in the number of mid-clause filled pauses in dialogue and monologue; dialogues had more mid-clause filled pauses than monologues. A comparison of the current study's findings with those of other studies, for instance Witton- Davies (2014), showed that dialogue was often associated with more mid-clause filled pauses than monologue. Furthermore, these results are in agreement with those obtained by Peltonen (2017), who indicated that high-proficiency participants, who scored higher in fluency, produced more mid-clause filled pauses in dialogue than monologue. Additionally, filled pauses are considered as a stalling mechanism to sustain the flow of the speech and reduce the use of silent pauses (Peltonen, 2017). An interesting similar finding by Tavakoli et al. (2020) demonstrated that advanced speakers produced filled pauses and mid-clause filled pauses, whereas less advanced speakers produced mid-clause silent pauses. A possible explanation for the significant findings in the current study might be that the participants benefitted from the use of mid-clause filled as a communicative strategy to hold the speaker's turn or to keep the flow of the speech in the dialogic performance (Peltonen, 2017).

The following sentences (6–16) between speakers A and B are examples of L2 mid-clause filled pauses of > 0.25. They are surrounded with blue parentheses and inserted in mid-clause/phrase by L2 speakers in dialogue to contribute to the flow of

the speech. Filled pauses can be lexical (e.g., *you know, I see, yeah*) as in lines 12, 15, and 16, or non-lexical (e.g., *uhh, umm or aha*) as in lines 6–16.

Speaker A

6 *I think there is a lot of advantage for this thing umm (0.57) like uhh (0.59) the women*
7 *who need to uhh (0.45) for example, umm (0.59) (0.48) the widows who doesn't*
8 *have uhh (0.68) his husband her husband. May be uhh (0.33) (0.97) she had a lot*
9 *of children so she can umm (0.26) (0.83) help herself and her kids with uhh (0.49)*
10 *life (0.65).*

Speaker B

11 *Uh yes that's correct. (Yeah) for me I think uhh (0.29) the advantages are umm we*
12 *can't (you know uhh) list the advantages because they because they are a lot*
13*(0.51). uhh (0.33) one advantages is that the woman is umm (0.27) independent*
14*(0.52). Also the price of the gas umm (0.41)*

Speaker A

15 *Yes as we know from the 2018 (0.41) uhh (0.45). You know uhh the (0.47)*
16 you know the gas and (0.29)petrol uhh (0.33) has the price is (0.46) uh raised up.

Moving to the mean length of mid-clause filled pauses measure, none of the comparisons indicated a significant difference across modes. This result reflects that of Tavakoli et al. (2020), who also found that the length of mid-clause filled pause was not significant across proficiency levels. This further supports Segalowitz et al.'s (2017) study, which demonstrated that the length of filled pauses was not an appropriate enough measure of fluency to be classed as one of the core of L2 utterance fluency measures, such as mean length of silent pause, MLR, and PTR. Furthermore, this result is in line with Witton-Davies's (2014) study, which showed that the length of filled pauses was not significant across monologic and dialogic tasks. Measuring the mean length of mid-clause filled pauses is not a suitable measure to

demonstrate differences between monologue and dialogue performances. Therefore, we could not reject the null hypothesis.

Finally, the differences between the two modes in silent pause frequency and end-clause silent pauses were significant, indicating that monologues had more silent pauses with a medium effect size (r = 0.46), and more end-clause silent pauses with a large effect size (r = 0.56). These findings are in line with Tavakoli's (2016) study, which showed that monologues had more external end-clause pausing than dialogues. More recently, Tavakoli et al. (2020) indicated that the number of end-clause silent pauses did not differentiate the different levels of proficiency, as all L2 participants, regardless of their level, had similar numbers of end-clause silent pauses. This suggests that end-clause silent pausing is a characteristic of L1 speakers. An implication of this finding (end-clause SPs in monologue) is the possibility that because monologue is an extended uninterrupted talk by one speaker, whereas dialogue is short, interrupted turns with shared pauses between turns that could belong to both speakers. This combination of findings provides some support for the argument that the L2 participants' pausing behaviour in the current study could also be explained in the light of speech production processing (e.g., Levelt, 1989; de Bot, 1992; Kormos, 2006). In this case, the end-clause SPs (external pausing) in monologue would help speakers during the conceptualisation phase, while clause internal (mid-clause SPs/FPs) as in dialogue would help speakers during their formulation and articulation phases of speech production processing.

Despite this, the above results are clearly in line with existing findings (e.g., Michel, 2011; Witton-Davies, 2014; Kirk, 2016; Peltonen, 2020), which indicate that the performance of speakers in dialogue outperforms their performance in monologue.

This suggests that the presence of an interlocutor in dialogue helps the speaker to produce faster speech. It can be argued that the interactive nature of dialogue conversation between a speaker and an interlocutor could actually motivate speakers to exchange speech interactively, and they thus show their willingness to communicate by speaking faster and reducing hesitations, repetitions, and silent pauses (Tavakoli, 2016). Therefore, in the current study, pause frequency, duration, and location (e.g., mid-clause filled pauses, duration of mid-clause silent pauses) are important findings contributing to the breakdown of fluency in the dialogic performance.

5.5 Repair Fluency

RQ1c: Are there any differences between monologue and dialogue in terms of repair fluency measures?

Repair means the change of speakers' language in terms of repetitions, false starts, reformations, or replacements. Fluency in L2 can be affected at any stage of L2 speech production processing because L2 speech requires conscious processing (Witton-Davies, 2014). Repair, for example, occurs in the monitoring stage of speech processing and it takes the form of replacements, reformations, repetitions, hesitations, or false starts; these could reduce the speed and flow of the speech performance and cause dysfluencies. However, slow speech is not necessarily wrong because it may be caused by self-monitoring, which could affect the speech flow (de Bot, 1992; Kormos, 2006). Kormos (2006) argues that L2 speakers benefit from monitoring their speech because it allows them to pay attention to the structures and different aspects of the language. According to Witton-Davies (2014), more proficient speakers tend to be concerned about the accuracy of their oral production, and as a

result they produce fewer repairs. However, there are no differences in terms of number of repairs between L2 speakers of low and intermediate levels of proficiency.

Lambert (2017) argues that utterance fluency dimensions, including repair, are related to Levelt's (1989) blueprint stages of oral speech processing. For example, the speech rate is related to the final stage of speech production processing, the articulation stage. Breakdown measures, as mentioned before in this thesis, are related to the formulation stage. Finally, repairs, such as self-repair, are related to the monitoring phase, where the participants use their attentional resources for self-correction (Kormos, 1999). In addition, Peltonen (2020) and Witton-Davies (2014) suggest that repair measures, including reformations, repetitions, false starts, and replacements, seem to be related to Levelt's blueprint of the speech production model and the role of each measure is basically different. For example, reformations reflect monitoring at one of the three stages of speech processing: when choosing a preverbal message, when choosing grammatical and phonetic sounds, or in the articulation stage. Repetitions have a similar function to filled and silent pauses in that they allow more online planning time to maintain the speed of speech during conceptualisation or formulation. During this time, they think about what to say next and how to say it. Replacements and false starts are L2 words or utterances that passed the conceptualisation and formulation stages but were disregarded incomplete when they were partially articulated because the monitor detected a problem, either in the conceptualization or formulation stages.

While the differences between monologue and dialogue were statistically significant for some measures of speed and breakdown, the non-parametric Wilcoxon signed-rank test showed no statistical differences between the two modes in repair fluency. Although repair is a good measure, the results showed that the participants did not

differ in their use of repair fluency across the monologue and dialogue tasks for all repair measures, meaning that the participants performed similarly in terms of repairs in both modes. For example, the median values for false starts/60 were very similar across modes (e.g., monologue *MED* = 3.56 vs. dialogue *MED* = 3.53) with a small effect size ($r = 0.03$). Similarly, the median values of the Wilcoxon test in the measure of total repairs/60 in monologues (*MED* = 8.95) and dialogues (*MED* = 9.07) were close to each other with a small effect size ($r = 0.02$). Close median values were also found in repetitions/60 (*MED* = 6.95) in monologues and in dialogues (*MED* = 7.36) with a very small effect size ($r = 0.07$).

The non-significant results of all the repair measures are in line with previous studies (e.g., Michel, 2011; Witton-Davies, 2014). In Michel's (2011) study, the results of unpruned speed and repair fluency did not reach significant differences in monologue and dialogue. Although repair fluency was not significant in Witton-Davies's (2014) study, in dialogue the participants repeated themselves more than in monologue. However, when he combined filled pauses and repetitions in a single measure called repetition, the results were found to be significant and there were fewer repetitions in dialogue. Thus, it can be argued that decisions made when choosing the fluency measures can affect the results. For example, repetitions are commonly known for buying more time to plan for utterances. They are similar to pausing phenomena that help to maintain the flow of speech. Therefore, repetitions and filled pauses seem to be related to each other than repetitions and reformations or silent pauses and filled pauses (Witton-Davis, 2014). Another example of combining repair measures is in Iwashita et al.'s (2008) study, where instead of measuring reformations, repetitions, and false starts alone, different types of repairs were grouped under the word 'Repairs'.

Thus, here in the dialogic task, although the results were not significantly different, the L2 participants used slightly more repetitions during dialogic speaking (monologue *MED* = 6.95 vs. dialogue *MED* = 7.36). A possible explanation for this might be L2 participants use more repetitions to avoid silent pauses while planning and to sustain the floor. This finding was also reported by Peltonen (2017), where advanced speakers used more repetitions to fill the silence while planning for monologue and to hold the floor in dialogue.

The results from SLA literature regarding the relationship between repair fluency and L2 proficiency are complex and mixed. The relationship between the number of repairs and dysfluency is not clear in previous studies (Kormos and Dénes, 2004). A recent study by Tavakoli et al. (2020) showed that the number of repairs was reduced in more advanced participants' speech. In contrast, Huensch and Tracy-Ventura's (2017) study showed that participants studying abroad developed in the dimensions of speed and breakdown but not in repair. The possible explanation for this is that the proficiency level did not influence the development of repair.

Therefore, it can be assumed that repair is a matter of individual differences or personal traits (Kormos and Dénes, 2004; Kahng, 2014). Similarly, Zuniga and Simard (2019) demonstrated that repair fluency is an underlying trait, and it is not related to proficiency level. This means that L1 self-repair is linked to L2 self-repair. Even if the learner's L2 proficiency increased, L2 self-repair would not change because it is a stable personality trait.

Finally, while several measures of speed and breakdown fluency have been shown to be reliable signs of L2 fluency, further studies are needed on repair fluency (de Jong et al., 2013) to uncover the reasons behind using repair during L2 speech production.

For example, the question of whether repairs are a matter of personal speaking style or related to L1 behaviour in L2 monologic and dialogic performances is still unanswered. The above results also imply that more qualitative research and in-depth analysis are required to uncover the aspects of language that were repeated, replaced and corrected and to consider the functions of different types of repairs. Dimensions of repair fluency are to some extent neglected and the focus in the literature is on the measures of speed and breakdown fluency. Therefore, the following conclusion can be drawn from the present study: repair fluency measures in monologue and dialogue could be related to individuals' speaking styles or to participants' L1 repair behaviour. To prove this claim, the study should be repeated in the future using L1 repair fluency measures to ensure that the results obtained are not affected by L1 fluency behaviour.

5.6 WM Predictability of L2 Oral Performance in Monologue and Dialogue

The second question was: To what extent can WMC predict utterance fluency in monologue and dialogue? This section is dedicated to discussing the main findings in relation to both the research question and the hypothesis. This study hypothesised that individual differences (IDs) in WMC can predict the variations in the dialogic and monologic performances of 64 L2 participants. The second question adopted the statistical approach of multiple regression analysis to examine the relationship between WMC and utterance fluency in monologue and dialogue, and to examine the role of WM in explaining the performance of 64 L2 participants in utterance fluency in monologue and dialogue. Additionally, the findings will be linked to the previous studies on WMC and utterance fluency. WM was captured through Backward Digit test (BDST) (Kormos and Safar, 2008, Kormos and Trebits, 2011) and Operation Span

Test (OST) (Turner and Engle, 1989) tests, whereas, L2 fluency was measured by speed, breakdown, and repair.

Based on previous studies, WM is responsible for dividing and guiding the attention of L2 speakers during online processing (Ahmadian, 2012, Baddeley, 2012). The WM attention could be expandable to support and revise L2 lexical, syntactic, and semantic aspects (Robinson, 2011). Additionally, Wright (2013) argues that the variation in WMC may cause variability in retrieving and producing the structure of the target language because the limitation of WMC and the lack of access fluidity to linguistics and syntactic knowledge results in dysfluent speech (Segalowitz, 2016).

The limited access to the mental lexicon makes the processing of L2 speech production challenging because it is not entirely automatic (Kormos, 2006). In L2, WM is assumed to be fully loaded by the need to allocate attentional resources to the required stage in speech production processing. This is because L2 speech production is a sequential process (Skehan, 2015). Furthermore, Skehan (2014a) argues that WM is not an expansible system, meaning that difficult and challenging tasks are assumed to limit the capacity of WM by reducing the memory capacity and attentional resources. As a result, the amount of oral data that can be processed is reduced and this may cause dysfluencies, such as filled pauses, silent pauses, hesitations or repetitions (Kormos, 2006).

Following Mota (2003) and Gilabert and Muñoz (2010), the current study hypothesised that participants with higher WMC would speak more fluently than lower WMC in terms of speed, breakdown, and repair. This is because a large WMC improves the processing of retrieving lexical information, as a result improving the quality of the L2 oral production. It was expected that WMC and L2 utterance fluency measures would

be significantly correlated. It was also expected that the variations in WM would explain the variations in L2 participants' oral fluency performance across monologue and dialogue in terms of speed, breakdown, and repair fluency measures. Examining the relationship between WMC and L2 speech production is a traditional approach that has been adopted by L2 correlational studies on WMC and cognitive behaviour (e.g., Daneman and Carpenter, 1980; Daneman and Green, 1986; Turner and Engle, 1989; Mota, 2003).

Thus, the present study used a correlational approach to define the degree of association between WMC and measures of L2 utterance fluency in monologue and dialogue and to determine whether WMC reliably predicts L2 oral fluency performance. Pearson's correlation coefficient (*r*) was used to measure the relationship between both BDST and OST with fluency measures of speed breakdown and repair in monologue and dialogue, and multiple regression analysis was used to examine the predictive power of WMC.

The results demonstrated that the OST and BDST were not statistically significantly correlated with the composite measures of utterance fluency in monologue and dialogue (speed, breakdown, and repairs). Also, the multiple regression analysis was not significant in all models F (2.63) = 0.80, p = 0.45. This conclusion of non-significant findings is in line with Awwad's (2017) study, which investigated the predictive power of language proficiency and WM in explaining the oral performance of L2 participants in lexical complexity, syntactic complexity, fluency, and accuracy. Awwad's findings showed that the variations in the L2 participants' WM, as measured by BDST, and their language proficiency failed statistically to explain the variations in syntactic complexity, accuracy, speed, and pausing fluency. However, WM was found to be

significantly correlated with accuracy and syntactic complexity, but not with L2 fluency. Furthermore, his study showed that the individual contribution of each dependent variable to the model was statically significant for the independent variable of language proficiency test, but not for the other independent variable of BDST.

Furthermore, a comparison of the current study findings with those of other studies (e.g., Fortkamp, 2000; Mizera, 2006; Kormos and Trebits, 2011; Awaad, 2017; Georgiadou and Roehr-Brackin, 2017) confirms that there are no associations between WMC and L2 utterance fluency measures. For example, Fortkamp (2000) argues that WM is task specific. In her study, there was no significant correlation between WMC, as measured by an operation word speaking test and a Speaking Span Test (SST), and L2 oral fluency, as measured by speed, breakdown, and repair.

A similar finding was also reported by Mizera (2006), who investigated the relationship between WMC as measured by SST and L2 fluency as measured by speed, breakdown, and repair. The overall findings revealed weak correlations between the SST, repair and speed fluency measures, and no significant correlations were found between SST and other L2 fluency measures (e.g., pausing) ($r = .33$ to $.36$, $p < .05$). Similarly, Weissheimer and Mota (2009) found that higher scores of WMC as measured by the SST were not a good predictor of the L2 participants' performance in pruned and unpruned speech rates. Similar results were also reported by Kormos and Trebits's (2011) study, where WMC had a small effect on L2 participants' oral performance in terms of accuracy in more complex tasks. Further results indicated that there was a significant correlation between BDST and complexity, except for accuracy and fluency. Finally, Kormos and Trebits's (2011) findings seem to be consistent with other research by Georgiadou and Roehr-Brackin (2017), who found

that the number of pauses was reduced in L2 participants with higher scores of BDST, whereas there was no significant relationship between both WMC tests and the fluency measures of self-repair, repetition, and speech rate.

The present study results contradict Mota's (2003) study, in which high WMC participants, as measured by SST, were found to be correlated positively with L2 complexity and accuracy, but there was a small to moderate correlation with speed fluency as measured by MLR and pruned and unpruned SRs. However, using SST measures not only executive WMC but also L2 proficiency because the participants need to produce correct grammatical sentences in order to acquire high scores in the test.

Additionally, the current study findings opposed Gilabert and Muñoz's (2010) study, in which low positive correlations were found between the fluency measures of unpruned speech rate and the number of syllables/60 ($r = .23$, $p < .05$), and the Reading Span Test (RST) and complexity ($r = .24$, $p < .05$), but not with accuracy. Gilabert and Muñoz's (2010) study set out to examine the power of WMC, as measured by RST, in predicting the oral performance of 59 Spanish participants in terms of fluency, accuracy, and complexity. Thus, these conflicting results of speed fluency between Mota (2003) and Gilabert and Muñoz (2010) could be related to the way they measured and operationalised WM with L2 fluency. In Mota's (2003) study, a small to medium correlation was found between the SST and speed fluency, but not breakdown fluency. No significant correlation was found between the RST and speed fluency. This latter finding broadly supported the work of other studies in this area linking speed fluency with WMC, and finding that variations in WM could not be

explained by the variations in L2 speed fluency (e.g., Fortkamp, 2000; Mizera, 2006; Kormos and Trebits, 2011; Awwad, 2017; Georgiadou and Roehr-Brackin, 2017)

To conclude this section, the majority of the findings on L2 fluency and WM mentioned above showed no significant results in the correlation tests. It can therefore be assumed that WMC cannot predict the oral performance of L2 fluency in monologue and dialogue. This is not surprising as a similar result had been reported in previous studies (e.g., Fortkamp, 2000; Mizera, 2006; Kormos and Trebits, 2011; Awaad, 2017; Georgiadou and Roehr-Brackin, 2017) that examined the association between WMC and L2 oral performance in L2 utterance fluency. There is, therefore, a definite need to update or use more WM tests that can assess L2 production and processing. It is also clear that future research is recommended on WM and utterance fluency in monologue and dialogue, but the operationalisation and methods used in this study could be taken as a framework to be used again for further thorough and systematic investigation. It is possible that L2 proficiency is a reason for L2 fluency as in Awwad's (2017) study, where language proficiency appeared to be a reliable variable to explain variations in L2 participants' speed and pausing fluency. Juffs and Harrington (2011) argue that the impact of WM could be varied by task condition, language proficiency, and learners' age. Usually, research on WM is considered correlational in nature and participants' numbers could be another reason for a non-significant correlation because correlational studies require a large number of participants to achieve statistical power (Juffs and Harrington, 2011).

Furthermore, it can be argued that the idea that BDST is able to examine only the central executive is not theoretically supported (Albarqi, 2020). This is because the updated model of Baddeley includes episodic buffer as a new WM component that

takes over the storage job that was previously assigned to the central executive component. Finally, the absence of a relationship between WMC and L2 oral fluency in monologue and dialogue might be related to reasons other than WMC. For example, L2 speakers sometimes feel shy about correcting themselves many times (Lennon, 1990) because this may cause breakdown in the oral communication and slow down the speech (Kormos, 2006).

5.7 Conclusion

This chapter discussed the results of the current study which were obtained to answer the two research questions and the three sub-questions. The research findings were introduced and explained in greater details with regards to the suggested hypotheses, previous research on L2 fluency and WMC in monologue and dialogue, speech production models, and utterance fluency measures.

Chapter 6 Conclusion

6.1 Introduction

This chapter presents the conclusion from the current study findings. It further discusses the implication for literature, teaching, learning and testing. This chapter concludes with a discussion of the limitations of the current study as well as its implications for designing a follow-up second study, and wider suggestions for future research beyond the scope of this project.

6.2 Summary of KeyFindings

L2 Fluency has been studied less in dialogic performance and there is a limitation in examining the operationalizations of fluency measures across oral monologic and dialogic performances. This study has used variety of measures to fill this gap in the literature. The aim of the present study was to examine the differences between L2 oral fluency as measured by speed, breakdown, and repair in monologue and dialogue task. The second aim of this study was to test the predictive power of WMC, as measured by BDST and OST, in explaining the variations in L2 dialogic and monologic performances. The findings make several contributions to the current literature. First, the hypothesis was partially confirmed that dialogic performance was more fluent than

monologic performance, as indicated by the statistically significant differences between two modes in eight temporal fluency measures: speech rates, speech samples' length, number of syllables, PTRs, silent pauses, end-clause silent pauses, duration of mid-clause SPs, and duration of mid-clause FPs. Furthermore, the findings also confirmed that there were no significant differences in the use of repairs in dialogue and monologue due to similar values in both modes. Speech rates were faster in dialogue. Silent pauses and end-clause silent pauses were fewer in dialogue than monologue. Finally, longer mid-clause SPs and more mid-clause FPs were associated with dialogic performance.

Taken together, the current study findings are in accordance with those of other fluency studies (e.g., Witton-Davies, 2014; Tavakoli et al., 2016; Peltonen, 2020) that found speech rates, speech samples' length, number of syllables, phonation-time ratios, silent pauses, end-clause silent pauses, duration of mid-clause SPs, mid-clause FPs are the most common and reliable measures that represent fluency in monologue and dialogue except for repairs. Repair fluency could be related to L1 individuals' speaking style. The findings also suggested that it would be reasonable to use monologue fluency measures to measure dialogue fluency. However, some modifications should be applied to the length of the speech samples by standardizing measures of fluency per minute of speaking time (de Jong, 2016). This helps to minimize the differences caused by pauses, especially in dialogic data that have both individual SPs and turn pauses. Traditionally mid-clause FPs and duration of SPs were treated in the literature as examples of dysfluency; however, some researchers (Kormos, 2006; Kahng, 2014) look at these pauses as ordinary way to process the speech and show the need for more time to formulate the utterances (Peltonen, 2017).

It is important to discuss the fact that, even if a measure of L2 fluency does not show any significant difference between monologue and dialogue tasks, it does not mean that this measure is not accurate or informative (Pallotti, 2009). The possible reason for this lack of difference could be that both groups are similar in this particular measure. Additionally, Pallotti (2009) demonstrates that researchers should pay attention to not only differences and variations among groups in terms of measures, but also to similarities and contrasts. For example, if two groups of participants do not show any significant differences, this is considered an interesting finding, even if this measure does not show any difference after a period of time. It seems possible that this result is due to a personal trait that does not change or vary. Thus, it does not mean that this measure is poor or invalid, but it must show its basic construct adequately (Pallotti, 2009).

Furthermore, the present study draws on existing research on both WMC and L2 speech production in monologue and dialogue, to follow the hypothesis that working memory capacity is related to controlled processing in L2 speech performance where attentional resources depend on individuals' WMC. Participants' WMC as measured by OST and BDST was not a strong predictor to explain the variations in L2 oral performances in monologue and dialogue tasks. The lack of significant correlation between WMC and L2 fluency in both modes suggests further qualitative and quantitative studies in this area because to our knowledge there has been no studies on the relationship between WMC and L2 dialogic performance.

6.3 Implications for Literature, Teaching and Testing

Based on the discussed results of the current study, the implications for literature/research, language learning, teaching and testing are presented in this

section. The findings of this study have significant implications for the understanding of how language learning could be improved in interactive tasks. This study could contribute to gaining a deeper understanding of speech production processes in relation to WMC and L2 fluency in an interactional context. This study could also contribute to increasing the reliability and generalisability of L2 utterance fluency measures in dialogic and monologic performances by comparing the results from the fluency measures that were used in this study to those used in the previous literature on L2 fluency.

According to Michel (2011), the intake of new information and feedback is greater in dialogue than in monologue. In monologue, the speakers depend mostly on their own attentional resources. As a result, the high pressure on attentional resources could cause a lack of attention to linking form and meaning. For example, in monologue, the stages of conceptualisation, formulation, and monitoring, are performed by a single speaker, while in dialogue the speaker has a chance to plan, conceptualise, and formulate the speech during the interlocutor's turn (Tavakoli and Foster, 2008).Taken together, the results of the current study suggest that dialogic tasks are more fluent than monologic tasks in terms of speech rates, PTR, mid/end-clause silent pauses and mid-clause filled pauses. Additionally, the dialogue samples are faster than monologue samples, as the number and duration of SPs/FPs while speaking help to maintain the flow of speech in dialogic tasks.

Examining the relationship between L2 fluency in dialogic and monologic performances is an essential requirement for L2 acquisition and language proficiency assessment because language users rarely use language in monologic contexts. The findings of this study have significant implications for language testing in terms of the

levels of fluency expected in monologic and dialoguc tests. In fact, the inclusion of dialogues in oral testing is more valid and beneficial than including monologue because dialogue involves face-to-face conversation between the examiner and the candidate (Witton-Davies, 2014). This allows more time, during the partner's turn, for the candidate to think about their subsequent utterances. For example, speaking tests such as IELTS or OPI are more valid than the monologue tasks in TOEFL. Furthermore, the interactive dialogue test may reduce the length of the candidate's speech turns because the speech alternates between the examiner and the candidate (Witton-Davies, 2014). For instance, in IELTS or OPI, when the candidate produces short answers, the examiner moves on to another question. This process is considered to be a natural way of communication in dialogue, but it is not possible in the TOEFL test, which depends on elicited monologue. According to Witton-Davis (2014), it seems that judging oral fluency in dialogic format tests (e.g., IELTS) is more appropriate than using monologic format tests (e.g., TOEFL) for assessing learners' fluency because dialogic tasks are more authentic and closer to everyday situations.

Another significant point of the present research is related to language teaching and learning. It has produced findings that will be beneficial for L2 English instructors in the English Department at the University of Jeddah in many ways, for example by showing that students' speaking abilities in interactional contexts can be improved to some extent by using more dialogic activities in the classroom. This study also provides reasons for students' dysfluencies and suggestions for L2 speaking tasks that other language instructors may consider adopting to promote and facilitate L2 production. For example, L2 learners should be encouraged to use some of the proposed strategies to help maintain the flow of their speech in real-life communication. When L2 students encounter problems with their speaking, they could

use filled pauses rather than remaining silent or repeating the same speech without modifications (Peltonen, 2017).

Another implicationof this study relates to language teaching and learning in the classroom. Language instructors/teachers can encourage learners to understand and evaluate their speech fluency in terms of pauses, hesitations, and self-corrections. For example, learners can record their oral performances using videos or audio recording tools that are easily accessible on their smart phones in order to analyse their speech in terms of fluency. Language teachers may also encourage L2 learners to listen to and transcribe their conversations and to try to identify words or sentences that affect the flow of their speech. Teachers can show learners a model example and explain that fluency means smooth delivery of speech performance, and that dialogue is co-constructed speech where two speakers share the responsibility of maintaining the flow of the speech. Learners could also record their L1 and L2 speech samples and compare their L2 with their L1 and evaluate their weaknesses and strengths. They may try to identify the common disfluencies in their speech and try to avoid them in the future in order to make their L2 as fluent as their L1. This would help to raise learners' awareness about their own speaking styles.

6.4 Limitations and Future Suggestions

One source of weakness in this study that could have affected the measurements of L2 utterance fluency is sample size. It may suppress the differences between dialogic and monologic performances. This study should be repeated using a larger sample size in order to produce more findings that are statistically reliable and focused. In the current study it was difficult to recruit more participants as this would have been time consuming and beyond the scope of this thesis. Future research on L2 fluency should

consider different factors that could affect the oral fluency of the L2 learners. For example, online planning, social context, task types, amount of instructions and topics are among the factors that can affect L2 fluency performance. Comparing dialogue and monologue is a bit difficult. Findings from the current research cannot be easily attributed to the features of monologue and dialogue. Sometimes measures of L2 utterance fluency may vary according to the nature of the task, context, interlocutor, and topic (Witton-Davies, 2014). For example, the results of this study could be affected by the different instructions given to the students in the monologic and dialogic tasks as the former provides only a few vocabulary items whereas the latter provides a range of guiding questions/statements for the students to borrow from. Another limitation that future studies should consider is a counter-balanced task in order to avoid the practice effect.

In this study, the learners studied L2 English from written texts, such as English literature, drama, grammar and translation. The opportunities to develop and practice speaking skills in the university are limited. In addition, the main language is Arabic and even when teaching English, the teachers sometimes use Arabic in translation and to explain grammar and vocabulary. Additionally, some EFL learners are shy and do not initiate speech. They need to be triggered to participate and engage in oral activity. Furthermore, the teacher-centred method of teaching can reduce students' fluency development because they are receivers in the classroom.

The generalisability of the current study's results is subject to certain limitations. For instance, studying L2 fluency measures alone is not enough (Segalowitz, 2010); L1 measures should also be included to help understand whether the participants' performance is L2 specific or L1 specific. For example, Towell et al. (1996) argued

that the length and number of pauses were not predictors of fluency, but rather characteristics of individuals' speaking style.

The majority of the fluency studies reviewed in this study compared the speech samples of L1 fluency and L2 fluency from different participants with different L1s. For example, de Jong (2016) used L1 Turkish and English and L2 Dutch participants. It is rare to see studies where the researchers collect both L1 and L2 speech samples from the same participants, except for Derwing et al. (2009) and Peltonen (2018), Duran-Karaoz and Tavakoli (2020). Thus, further research is recommended to compare individuals' fluency measures across L1 and L2 speech samples of the same participants in order to create fluency profiles in oral language production (Lintunen, 2019). Including L1 measures would contribute to this line of research and help to gain a deeper insight into the influence of L1 temporal measures on L2 fluency.

The findings from the previous studies (e.g., Derwing et al., 2009; Duran-Karaoz and Tavakoli, 2020) indicated that some of L2 breakdown and repair fluency measures such as filled pauses, hesitations, and replacements were to some degree influenced by personal speaking styles or L1 fluency behaviour. There is little data published in this area in terms of used languages, individual differences, and task types. More studies are needed to investigate the role of L1 oral fluency in predicting the performance of L2 fluency in terms of speed, breakdown, and repair.

Thus, related to this study, including L1 fluency measures from the same speakers in monologue and dialogue could help to explain whether there is any relationship between L1 Arabic and L2 English in terms of speed, breakdown, and repair. On the one hand, if the L1 speech production measures correlate with L2 speech production measures, this study would be in line with previous empirical SLA studies, such as

Bradlow et al. (2017), which found that fluent L1 speakers may also be fluent L2 speakers. This could lead to the tentative conclusion that fluency is a trait. On the other hand, if there is no meaningful relationship between L1 and L2 speech production in terms of fluency, L1 speech fluency is considered as an automated oral production processing, as explained in Levelt's model, and L2 fluency is therefore a language-specific trait.

Duran-Karaoz and Tavakoli (2020) suggest that there is a strong relationship between L1 and L2 fluency in speakers with high L2 proficiency level because L2 speech production processing is not entirely parallel, especially for less advanced level of proficiency speakers. The controlled and serial processing of L2 is to great extent results in working with insufficient linguistic knowledge that makes the speech less fluent with slower speech rate, unnecessary hesitations and pauses (Kormos, 2006). Duran-Karaoz and Tavakoli (2020) argue that when lower-level proficiency speakers improve their L2 with regular practice, explicit instruction, or authentic language exposure, the L2 production speech becomes smoother and more automatic. Furthermore, the linguistics knowledge increases allowing for parallel processing and thus the processing of L2 speech becomes similar to L1 processing.

Finally, a mixed-method design could be used to obtain broad results for L2 fluency in dialogic and monologic performance and WMC. For example, qualitative research could be conducted, including stimulated recall questionnaires. Dialogic performance could be analysed in terms of the interlocutor's contribution. For example, the analysis can include important aspects of dialogic performance, e.g., turn taking, fillers, interruptions, collaborative completions… etc.

Further research might usefully adopt the interactive alignment model (Garrod and Pickering, 2004) as a framework for L2 speech processing in dialogue. This could provide a clear view and support for L2 fluency in interactive tasks, for example, by examining speakers' gestures (e.g., hands), movements (e.g., head, eyes), and behaviour (e.g., speech tone) during the conversation.

6.5 References

Ahmadian, M.J. 2012. The Relationship between working memory capacity and L2 oral performance under task-based careful online planning condition. *TESOL Quarterly*. **46**(1), pp.165–175.

Albarqi, G.A. 2020. *Self-monitoring behaviour of L2 learners: Proficiency level, dual task paradigm and working memory capacity*. Unpublished Ph.D. thesis, University of Reading.

Albert, Á. and Kormos, J. 2011. Creativity and narrative task performance: An exploratory study. *Language Learning*. **61**(1), pp.73-99.

Allan, D. 2004. *Oxford Placement Test 2: Marking kit with user's guide and diagnostic key*. Oxford University Press.

Awwad, A.A. 2017. *The effects of task complexity manipulated by intentional reasoning demands on second language learners' speech performance: interaction with language proficiency and working memory*. Unpublished Ph.D. thesis University of Reading, UK.

Baddeley, A. 2012. Working memory: Theories, models, and controversies. *Annual Review Psychology*. **63**, pp.1-29.

Baddeley, A. 2015. Working memory in second language learning. In working memory in second language acquisition and processing. *Multilingual Matters.* pp.17-28.

Boersma, P. and Weenink, D. 2012. Praat: Doing phonetics by computer (Version 5.3.82) [Computer software]. *Amsterdam: Institute of Phonetic Sciences*.

Boomer, D.S. 1970. Psycholinguistics: *Experiments in spontaneous speech*. In Goldman-Eisler, F. Academic Press, London and New York. *Lingua*. **25**, pp.152-164.

Bosker, H., Quené, H., Sanders, T. and de Jong, N. 2014. Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*. ***31***(3), pp.349-365.

Bosker, H. 2011. *When is speech fluent? The relationship between acoustic speech properties and subjective fluency ratings.* Talk presented at The Language Acquisition Group, Max Planck Institute for Psycholinguistics. Nijmegen, The Nethrtlands.

Botwinick, J. and Storandt, M. 1974. *Memory related functions and age*. Oxford: Charles C Thomas.

Bradlow, A.R., Kim, M. and Blasingame, M. 2017. Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America.* **141***(2),* pp.886-899

Bui, G. and Huang, Z. 2018. 'L2 fluency as influenced by content familiarity and planning: Performance, measurement, and pedagogy', *Language Teaching*

Research. **22**(1), pp. 94-114.

Carroll, D. 2001. Restarts in novice turn beginnings: *Disfluencies or interactional achievements?* In Gardner, R. and Wagner, J. eds. Second language conversations. pp. 201-220. Continuum: London.

Case, R., Kurland, D.M. and Goldberg, J. 1982. Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*. **33**(3), pp.386-404.

CEFR. 2001. *Common European framework of reference for languages: Learning, teaching, assessment.* Council of Europe*.* Cambridge, UK: Cambridge University Press.

Cohen, J. 1960. A Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. **20**(1), pp.37-46.

Cohen, L., Manion, L. and Morrison, K. 2017. *Research Methods in Education*. Routledge.

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

Cowan, N. 2010. The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science.***19**,(1), pp.51-57.

Crible, L. and Pascual, E. 2020. Combinations of discourse markers with repairs and repetitions in English, French and Spanish. *Journal of Pragmatics*.**156**, pp.54-67.

Curran, P.J., West, S.G. and Finch, J.F. 1996. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*. **1**(1), p.16.

Daneman, M. 1991. Individual differences in reading skills. In Barr, R., Kamil, M. L., Mosenthal, P. B. and Pearson, P.D. eds. *Handbook of reading research,* **2**, pp. 512-538. Lawrence Erlbaum Associates, Inc.

Daneman, M. and Carpenter, P.A. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*. **19**(4), pp.450-466.

Daneman, M. and Green, I. 1986. Individual differences in comprehending and producing words in context. *Journal of Memory and Language*. **25**(1), pp.1-18.

Davis, l. 2009. The influence of interlocutor proficiency in a paired oral assessment. *Language Testing***. 26**(3), 367-396.

De Bot, K. 1992. A Bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*. **13**, pp.1-24.

de Jong, N. 2018. Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*. **15**, pp.1-18.

de Jong, N., Groenhout, R., Schoonen, R. and Hulstijn, J. 2013. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*.

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

de Jong, N., Steinel, M., Florijn, A., Schoonen, R. and Hulstijn, J. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition*. **34**, pp.5-34.

de Jong, N. 2016. Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*. **54**(2), p.113.

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

de Jong, N., Steinel, M.P., Florijn, A., Schoonen, R. and Hulstijn, J.H. 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*. **34**(5), pp.893-916.

Derwing, T., Munro, M., Thomson, R. and Rossiter, M. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*. **31**, pp.533-557.

Derwing, T., Rossiter, M., Munro, M. and Thomson, R. 2004. Second language fluency: Judgments on different tasks. *Language Learning*. **54**, pp.655-679.

Derwing, T. M., and Munro, M. J. 2015. Pronunciation fundamentals*: Evidence-based perspectives for L2 teaching and research.* Amsterdam, the Netherlands: John Benjamins.

Dimova, S. 2012 and Segalowitz, N. Cognitive bases of second language fluency. *Language Testing*. **29**(1), pp.153-156.

Dornyei, Z. 2007. *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford: Oxford University Press.

Duncan, J., Schramm, M., Thompson, R. and Dumontheil, I. 2012. Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin and Review*. **19**(5), pp.864–870.

Duran-Karaoz, Z. and Tavakoli, P. 2020. Prediciting L2 Fluency from L1 Fluency behaviour: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition.* **42,** 671-695.

Ellis, R. 2010. Second language acquisition and language-teaching materials'. In Harwood, N. ed. *English Language Teaching Materials; Theory and Practice.* Cambridge: Cambridge University Press, pp.33-57.

Ellis, R. 2017. Task-based language teaching. *The Routledge handbook of instructed second language acquisition*. **108**, pp.125.

Ellis, R. 2015. *Understanding second language acquisition 2nd .ed.Oxford applied linguistics*. Oxford university press.

Ellis, R., Barkhuizen, G. and Kirkgoz, Y. 2005. *Analysing learner language*. Oxford: Oxford University Press.

Elwood, R.W. 1991. The wechsler memory scale-revised: Psychometric

characteristics and clinical application. *Neuropsychology Review*. **2**(2), pp.179-201.

Felker E.R., Klockmann, H.E. and de Jong, N. 2019. How conceptualizing influences fluency in first and second language speech production, *Applied Psycholinguistics.* **40**(1), pp.111-136.

Fidell, S., Tabachnick, B., Mestre, V. and Fidell, L. 2013. Aircraft noise-induced awakenings are more reasonably predicted from relative than from absolute sound exposure levels. *The Journal of the Acoustical Society of America*. **134**(5), pp.3645-3653.

Field, A. 2013. *Discovering statistics using IBM SPSS statistics*. Sage.

Fillmore, C.1979. On Fluency. In Fillmore, C., Kempler, D. and Wang, W. eds. *Individual differences in language behavior.* pp.85-101. New York: Academic Press

Fortkamp, M.B.M. 1999. Working memory capacity and aspects of L2 speech production. *Communication and Cognition*. **32**(3-4), pp.259-295.

Fortkamp, M.B.M. 2000. *Working memory capacity and L2 speech production: An exploration study.* Unpublished Ph.D. thesis, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

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

Foster, P., Tonkyn, A. and Wigglesworth, G. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics*. **21**, pp.354-375.

Freed, B. E, Segalowitz, N., and Dewey, D. 2004. Contexts of learning and second language fluency in French: Comparing regular classrooms, study abroad, and intensive domestic programs. *Studies in Second Language Acquisition.* **26**(2), pp. 275-301.

Garrod, S. and Pickering, M.J. 2004. Why is conversation so easy? *Trends in Cognitive Sciences*. **8**(1), pp.8-11.

Gathercole, S.E., Service, E., Hitch, G.J., Adams, A.-M. and Martin, A.J. 1999. Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*. **13**(1), pp.65-77.

Georgiadou, E. and Roehr-Brackin, K. 2017. Investigating executive working memory and phonological short-term memory in relation to fluency and self-repair behavior in L2 speech. *Journal of Psycholinguistic Research*. **46**(4), pp.877-895.

Gilabert, R., Barón, J. and Levkina, M. 2011. Manipulating task complexity across task types and modes. Second language task complexity: *Researching the cognition hypothesis of language learning and performance*. pp.105-140.

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

Ginther, A., Dimova, S. and Yang, R. 2010. Conceptual and empirical relationships

between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*. **27**(3), pp.379-399.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in spontaneous speech.* New York: Academic Press.

Götz, S., 2013. *Fluency in native and nonnative English speech*. John Benjamins Publishing, Amsterdam.

Guz, E. 2015. Establishing the fluency gap between native and non-native speech. *Research in Language*. **13**(3), pp.230-247.

Hasselgreen, A. 2005. Assessing the language of young learners. *Language Testing*. **22**(3), pp.337-354.

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

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

Kahng, J. 2020. Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*. **41**(2), pp457-480.

Kahng, J. 2017. The effect of pause location on perceived fluency. *Applied Psycholinguistics*. **39**(3), pp.569-591.

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

Kirk, S.J. 2016. *Second language spoken fluency in monologue and dialogue*. Published Ph.D. Thesis. University of Nottingham.

Kootstra, G.J., van Hell, J.G. and Dijkstra, A.F.J. 2009. Two speakers, one dialogue: An interactive alignment perspective on codeswitching in bilingual speakers.

Kormos, J. 2015. Individual differences in second language speech production. In Schwieter, J.W. ed. *The Cambridge handbook of bilingual processing,* pp. 369.Cambridge: Cambridge University Press.

Kormos, J. 2006. *Speech production and second language acquisition*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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

Kormos, J. and Safar, A. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*. **11**, pp.261-271.

Kormos, J. and Trebits, A. 2011. Working memory capacity and narrative task performance. *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. **2**, pp.267-285.

Lancaster, G.A. 2015. Pilot and feasibility studies come of age. *Pilot and Feasibility Studies*. **1**(1), pp.1-4.

Lambert, C. 2017. Tasks, affect and second language performance. *Language Teaching Research*. **21**, pp.657-664.

Lambert, C., Aubrey, S. and Leeming, P. 2020. Task preparation and second language speech production. *TESOL Quarterly*.

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

Labov, W. 1972. *Language in the inner city; studies in the Black English vernacular.* Philadelphia: University of Pennsylvania Press.

Larson-Hall, J., 2008. Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*. ***24***(1), pp. 35-63.

Larson-Hall, J., 2015. *A guide to doing statistics in second language research using SPSS and R*. Routledge.

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

Leonard, K.R. 2015. *Speaking fluency and study abroad: what factors are related to fluency development?* Published Ph.D. thesis University of Iowa.

Levelt, W.J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA, US: The MIT Press.

Lintunen, P., Mutta, M. and Peltonen, P. 2019. *Fluency in L2 learning and use*. Multilingual Matters.

Lo, J.J., 2020. Between Äh (m) and Euh (m): The distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals. *Language and speech*. ***63***(4), pp.746-768.

Mackey, A. and Gass, S.M. 2005. *Second language research: Methodology and design*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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

Michel, M. 2011. Effects of task complexity and interaction on L2 performance. *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*. ***2***, pp.141-174.

Michel, M., Kuiken, F. and Vedder, I. 2007. The influence of complexity in monologic versus dialogic tasks in Dutch L2. *Journal of Dental Research - J DENT RES*. **45**, pp.241-259.

Mitchell, A.E., Jarvis, S., O'Malley, M. and Konstantinova, I. 2015. Working memory measures and L2 Proficiency. *Working memory in second language acquisition and processing*. **87**, pp.270.

Mizera, G.J. 2006. *Working memory and L2 oral fluency*. Unpublished Ph.D. thesis. University of Pittsburgh, PA.

Mojavezi, A. and Ahmadian, M.J. 2014. Working memory capacity and self-repair behavior in first and second language oral production. *Journal of Psycholinguistic Research*. **43**(3), pp.289-297.

Mota, M. 2003. Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos: Revista De Língua e Literatura Estrangeiras*. **24**, pp.69-104.

Morra, S. 1994. Issues in working memory measurement: Testing for working memory capacity. *International Journal of Behavioral Development.* ***17***(1), pp.143-159.

Nitta, R. and Nakatsuhara, F.2014. A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing*. **31**(2), pp. 147-175.

Oomen, C.C.E., and Postma, A. 2001. Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research,* ***30***, pp. 163-184.

Os, M., de Jong, N. and Bosker, H. 2020. Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and nonnative speech. *Language Learning*. ***70***(4), pp. 1183-1217.

Osaka, M. and Osaka, N. 1992. Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*. **30**(4), pp.287-289.

Pallant, J. 2007. SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows Version 15.

Pallotti, G. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*. **30**(4), pp.590-601.

Payne, J.S. and Whitney, P.J. 2002. Developing L2 Oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*. **20**(1), pp.7-32.

Payne, S. and Ross, B., 2005. Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning and Technology*. ***9***(3), pp.35-54.

Peltonen, P. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-method approach*. Published Ph.D. thesis, University of Turku.

Peltonen, P. 2018. Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal*. **102**(4), pp.676-692.

Peltonen, P. 2017. Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System*. **70**, pp.1-13.

Riggenbach, H. 1998. Evaluating learner interactional skills: Conversation at the micro level. *Talking and testing: Discourse approaches to the assessment of oral proficiency*. pp.53-67.

Riggenbach, H. 2000. *Perspectives on fluency*. Ann Arbor: University of Michigan Press/ESL.

Riggenbach, H. 1991. Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse processes*, ***14***(4), 423-441.

Robinson, P. 2011. *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins Publishing.

Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*. **22**(1), pp.27-57.

Rogers, J. and Revesz, A. 2019. Experimental and quasi-experimental designs. In McKinley, J. and Rose H. eds. *The routledge handbook of research methods in applied linguistics*. New York: Routledge, pp.1-27.

Rosen, V.M. and Engle, R.W. 1997. The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*. **126**(3), pp.211–227.

Sacks, H., Schegloff, E.A. and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*. **50**(4), pp.696-735.

Sato, M. 2014. Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System*. **45**, pp.79-91.

Segalowitz, N. 2010.*Cognitive bases of second language fluency*. New York: Routledge.

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

Segalowitz, N., French, L. and Guay, J.D. 2017. What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion? *Canadian Journal of Applied Linguistics / Revue canadienne de linguistique appliquée*. **20**(2), pp.90-116.

Shea, C. and Leonard, K. 2019. Evaluating measures of pausing for second language fluency research. *Canadian Modern Language Review*. **75**(3), pp.216-235.

Skehan, P. 2015. Limited attention capacity and cognition In: M. Bygate.ed. *Domains and directions in the development of TBLT*. Amsterdam: John Benjamins Publishing. **8**, pp.123-156.

Skehan, P. 2014a. Limited attentional capacity, second language performance, and task-based pedagogy. *Processing perspectives on task performance*. **211**, pp.260.

Skehan, P. 2003. Task-based instruction. *Language Teaching*. **36**, pp.1-14.

Skehan, P. 2014b. The context for researching a processing perspective on task performance. *Processing Perspectives on Task Performance*. **5**, pp.1-26.

Skehan, P., Foster, P., and Shum, S. 2016. Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*. **54**(2), pp.97-111.

Smith, L. and Abouammoh, A. 2013. Challenges and opportunities for higher education in Saudi Arabia: An exploratory focus group. In *Higher Education in Saudi Arabia*. Springer, Dordrecht, pp.167-179.

Tabachnick, B.G. and Fidell, L.S. 2013. *Using multivariate statistics*. 6th ed. International edition. Boston: Pearson Education.

Tauroza, S. and Allison, D. 1990. Speech rates in British English. *Applied Linguistics*. **11**(1), pp.90-105.

Tavakoli, P. 2016. Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*. **54**(2), pp.133-150.

Tavakoli, P. 2009. Investigating task difficulty: learners' and teachers' perceptions. *International Journal of Applied Linguistics*.**19**(1), pp.1-25.

Tavakoli, P., and Skehan, P. 2005. Strategic planning, task structure, and performance testing. In R. Ellis ed. *Planning and task performance in a second language.* Amsterdam: John Benjamins. pp. 239-273.

Tavakoli, P. 2018. L2 development in an intensive study abroad EAP context. *System*. **72**, pp.62-74.

Tavakoli, P. 2020. Making fluency research accessible to second language teachers: The impact of a training intervention. *Language Teaching Research.* pp1-26

Tavakoli, P. 2010. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*. **65**(1), pp.71-79.

Tavakoli, P. and Foster, P. 2008. Task design and second language performance: The effect of narrative type on learner output. *Language Learning*. **58**(2), pp.439-473.

Tavakoli, P. and Hunter, A.M. 2018. Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*. **22**(3), pp.330-349.

Tavakoli, P., Nakatsuhara, F. and Hunter, A. 2020. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*. **104**(1), pp.169-191.

Tavakoli, P. and Wright, C. 2020. *Second language speech fluency: from research to practice*. Cambridge University Press. pp180.

Tavares, M.D.G.G., 2008. *Pre-task planning, working memory capacity, and L2 speech performance*. Unpublished Ph.D. thesis, University Fedral de Santa Catarina, Brazil.

Thai, C. and Boers, F. 2016. Repeating a Monologue Under Increasing Time Pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*. **50**(2), pp.369-393.

Towell, R., Hawkins, R., and Bazergui, N. 1996. The development of fluency in advanced learners of French. *Applied Linguistics*. **17**(1), 84-119

Trofimovich, P. 2013. Interactive alignment: Implications for the teaching and learning of second language pronunciation. In: Levis, J. and LeVell, K. eds. *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, September 2012, Ames, IA: Iowa State University, pp.1-9.[Accessed 18 October 2020]. Available from http://jlevis.public.iastate.edu/pslltconference/4th%20Proceedings/Trofimovich%202012%20

Turner, M.L. and Engle, R.W. 1989. Is working memory capacity task dependent?

Journal of Memory and Language. **28**(2), pp.127-54.

Unsworth, N., Brewer, G.A. and Spillers, G.J. 2013. Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory and Cognition*. **41**(2), pp.242-254.

Unsworth, N., Heitz, R.P., and Schrock, J.C. 2005. An automated version of the span task. *Behavior Research Methods.* **37** (3), 498-505.

Vatz, K., Tare, M., Jackson, S. and Doughty, C. 2013. Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. *Sensitive periods, language aptitude, and ultimate L2 attainment*. pp.273-292.

Wechsler, D. 1997. Wechsler memory scale .3rd ed. *Administration and scoring manual. Psychological Corporation*, San Antonio, TX.

Weissheimer, J. and Mota, M.B. 2009. Individual differences in working memory capacity and the development of L2 speech production. *Issues of Applied Linguistics*. **17**(2).

Weissheimer, J. and Mota, M.B. 2011. Working memory capacity and lexical density in L2 speech production. *Organon*. **26**(51).

Witton-Davies, G. 2014. The study of fluency and its development in monologue and dialogue. Unpublished Ph.D. thesis, Lancaster University, Lancaster, UK.

Wright, C. 2015. Working memory and L2 development across the lifespan: A commentary. In: Wen, E.Z., Mota, M.B. and McNeil, A. eds. *Working memory in second language acquisition and processing*. Multilingual Matters, Bristol, pp.285-298.

Wright, C. 2013. An investigation of working memory effects on oral grammatical accuracy and fluency in producing questions in English. *TESOL Quarterly*. **47**(2), pp.352-374.

Wright, C. and Tavakoli, P. 2016. New directions and developments in defining, analyzing and measuring L2 speech fluency. *International Review of Applied Linguistics in Language Teaching*. **54**(2), pp.73.

Wilson, T. and Don, Z. 1986. The structure of silence between rums in two- party conversation. *Discourse Processes*. **9**(4), pp.375-390.

West, S.G., Finch, J.F. and Curran, P.J., 1995. Structural equation models with nonnormal variables: Problems and remedies. In Hoyle, R. H. ed. *Structural Equation Modeling: Concepts, Issues, and Applications*. Sage, pp.56-75.

Yan, X., Kim, H.R. and Kim, J.Y. 2020. Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency (CAF) features of speaking performances on the Aptis test. *Language Testing.* pp.1-26.

Zalbidea, J. 2017. 'One Task Fits All'? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*. **101**(2), pp.335-352.

Zuniga, M. and Simard, D. 2019. Factors Influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguist Research*. **48**(1), pp.43-59.

Appendices

Appendix 1 Language Background Questionnaire

Please answer the following statements and give explanation if possible.

Name: _____

Age: _____

Year (level): _____

First language: _____

Number of languages you speak: _____

Number of days/months/years lived in an English-speaking country: _____

Number of days/months/years learned English in an English-speaking country _____

Years of learning English as a foreign language: _____

Do you use English outside of school? If so, please describe where and how?

_____

Appendix 2 Quick Placement Test

UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

OXFORD
UNIVERSITY PRESS

Name: .....

Date: .....

quick placement test

Version 2

The test is divided into two parts:

Part 1 (Questions 1 – 40) – All students.

Part 2 (Questions 41 – 60) – Do not start this part unless told to do so by your test supervisor.

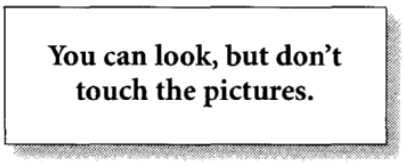
Time: 30 minutes

Part 1

Questions 1 – 5

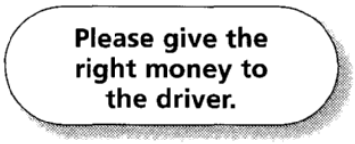
- Where can you see these notices?
- For questions **1** to **5**, mark **one** letter **A**, **B** or **C** on your Answer Sheet.

1



A in an office
B in a cinema
C in a museum

2

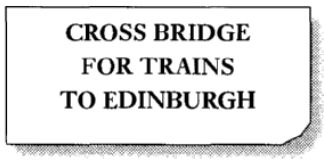


A in a bank
B on a bus
C in a cinema

3

A in a street
B on a book
C on a table

4



A in a bank
B in a garage
C in a station

5

A on clothes
B on furniture
C on food

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** on your Answer Sheet.

THE STARS

There are millions of stars in the sky. If you look **(6)** ..... the sky on a clear night, it is possible to see about 3000 stars. They look small, but they are really **(7)** ..... big hot balls of burning gas. Some of them are huge, but others are much smaller, like our planet Earth. The biggest stars are very bright, but they only live for a short time. Every day new stars **(8)** ..... born and old stars die. All the stars are very far away. The light from the nearest star takes more **(9)** ..... four years to reach Earth. Hundreds of years ago, people **(10)** ..... stars, like the North Star, to know which direction to travel in. Today you can still see that star.

6	**A** at	**B** up	**C** on
7	**A** very	**B** too	**C** much
8	**A** is	**B** be	**C** are
9	**A** that	**B** of	**C** than
10	**A** use	**B** used	**C** using

Questions 11 – 20

- In this section you must choose the word which best fits each space in the texts.
- For questions **11** to **20**, mark **one** letter **A**, **B**, **C** or **D** on your Answer Sheet.

Good smiles ahead for young teeth

Older Britons are the worst in Europe when it comes to keeping their teeth. But British youngsters **(11)** ..... more to smile about because **(12)** ..... teeth are among the best. Almost 80% of Britons over 65 have lost all or some **(13)** ..... their teeth according to a World Health Organisation survey. Eating too **(14)** ..... sugar is part of the problem. Among **(15)** ..... , 12-year-olds have on average only three missing, decayed or filled teeth.

11	**A** getting	**B** got	**C** have	**D** having
12	**A** their	**B** his	**C** them	**D** theirs
13	**A** from	**B** of	**C** among	**D** between
14	**A** much	**B** lot	**C** many	**D** deal
15	**A** person	**B** people	**C** children	**D** family

Christopher Columbus and the New World

On August 3, 1492, Christopher Columbus set sail from Spain to find a new route to India, China and Japan. At this time most people thought you would fall off the edge of the world if you sailed too far. Yet sailors such as Columbus had seen how a ship appeared to get lower and lower on the horizon as it sailed away. For Columbus this **(16)** ..... that the world was round. He **(17)** ..... to his men about the distance travelled each day. He did not want them to think that he did not **(18)** ..... exactly where they were going. **(19)** ..... , on October 12, 1492, Columbus and his men landed on a small island he named San Salvador. Columbus believed he was in Asia, **(20)** ..... he was actually in the Caribbean.

16	**A** made	**B** pointed	**C** was	**D** proved
17	**A** lied	**B** told	**C** cheated	**D** asked
18	**A** find	**B** know	**C** think	**D** expect
19	**A** Next	**B** Secondly	**C** Finally	**D** Once
20	**A** as	**B** but	**C** because	**D** if

Questions 21 – 40

- In this section you must choose the word or phrase which best completes each sentence.
- For questions **21** to **40**, mark **one** letter **A**, **B**, **C** or **D** on your Answer Sheet.

21 The children won't go to sleep ..... we leave a light on outside their bedroom.

A except **B** otherwise **C** unless **D** but

22 I'll give you my spare keys in case you ..... home before me.

A would get **B** got **C** will get **D** get

23 My holiday in Paris gave me a great ..... to improve my French accent.

A occasion **B** chance **C** hope **D** possibility

24 The singer ended the concert ..... her most popular song.

A by **B** with **C** in **D** as

25 Because it had not rained for several months, there was a ..... of water.

A shortage **B** drop **C** scarce **D** waste

26 I've always ..... you as my best friend.

A regarded **B** thought **C** meant **D** supposed

27 She came to live here ..... a month ago.

A quite **B** beyond **C** already **D** almost

28 Don't make such a .....! The dentist is only going to look at your teeth.

A fuss **B** trouble **C** worry **D** reaction

29 He spent a long time looking for a tie which ..... with his new shirt.

A fixed **B** made **C** went **D** wore

30 Fortunately, ... from a bump on the head, she suffered no serious injuries from her fall.

A other **B** except **C** besides **D** apart

31 She had changed so much that ..... anyone recognised her.
A almost **B** hardly **C** not **D** nearly

32 ..... teaching English, she also writes children's books.
A Moreover **B** As well as **C** In addition **D** Apart

33 It was clear that the young couple were ..... of taking charge of the restaurant.
A responsible **B** reliable **C** capable **D** able

34 The book ..... of ten chapters, each one covering a different topic.
A comprises **B** includes **C** consists **D** contains

35 Mary was disappointed with her new shirt as the colour ..... very quickly.
A bleached **B** died **C** vanished **D** faded

36 National leaders from all over the world are expected to attend the ..... meeting.
A peak **B** summit **C** top **D** apex

37 Jane remained calm when she won the lottery and ..... about her business as if nothing had happened.
A came **B** brought **C** went **D** moved

38 I suggest we ..... outside the stadium tomorrow at 8.30.
A meeting **B** meet **C** met **D** will meet

39 My remarks were ..... as a joke, but she was offended by them.
A pretended **B** thought **C** meant **D** supposed

40 You ought to take up swimming for the ..... of your health.
A concern **B** relief **C** sake **D** cause

Part 2

Do not start this part unless told to do so by your test supervisor.

Questions 41 – 50

- In this section you must choose the word or phrase which best fits each space in the texts.
- For questions **41** to **50**, mark **one** letter **A**, **B**, **C** or **D** on your Answer Sheet.

CLOCKS

The clock was the first complex mechanical machinery to enter the home, **(41)** ..... it was too expensive for the **(42)** ..... person until the 19th century, when **(43)** ..... production techniques lowered the price. Watches were also developed, but they **(44)** ..... luxury items until 1868, when the first cheap pocket watch was designed in Switzerland. Watches later became **(45)** ..... available, and Switzerland became the world's leading watch manufacturing centre for the next 100 years.

41	**A** despite	**B** although	**C** otherwise	**D** average
42	**A** average	**B** medium	**C** general	**D** common
43	**A** vast	**B** large	**C** wide	**D** mass
44	**A** lasted	**B** endured	**C** kept	**D** remained
45	**A** mostly	**B** chiefly	**C** greatly	**D** widely

Dublin City Walks

What better way of getting to know a new city than by walking around it? Whether you choose the Medieval Walk, which will **(46)** ..... you to the Dublin of 1000 years ago, find out about the more **(47)** ..... history of the city on the Eighteenth Century Walk, or meet the ghosts of Dublin's many writers on the Literary Walk, we know you will enjoy the experience.

Dublin City Walks **(48)** ..... twice daily. Meet your guide at 10.30 a.m. or 2.30 p.m. at the Tourist Information Office. No advance **(49)** ..... is necessary. Special **(50)** ..... are available for families, children and parties of more than ten people.

46	**A** introduce	**B** present	**C** move	**D** show
47	**A** near	**B** late	**C** recent	**D** close
48	**A** take place	**B** occur	**C** work	**D** function
49	**A** paying	**B** reserving	**C** warning	**D** booking
50	**A** funds	**B** costs	**C** fees	**D** rates

Questions 51 – 60

- In this section you must choose the word or phrase which best completes each sentence.
- For questions **51** to **60**, mark **one** letter **A, B, C** or **D** on your Answer Sheet.

51 If you're not too tired we could have a ..... of tennis after lunch.
A match **B** play **C** game **D** party

52 Don't you get tired ..... watching TV every night?
A with **B** by **C** of **D** at

53 Go on, finish the dessert. It needs ..... up because it won't stay fresh until tomorrow.
A eat **B** eating **C** to eat **D** eaten

54 We're not used to ..... invited to very formal occasions.
A be **B** have **C** being **D** having

55 I'd rather we ..... meet this evening, because I'm very tired.
A wouldn't **B** shouldn't **C** hadn't **D** didn't

56 She obviously didn't want to discuss the matter so I didn't ..... the point.
A maintain **B** chase **C** follow **D** pursue

57 Anyone ..... after the start of the play is not allowed in until the interval.
A arrives **B** has arrived **C** arriving **D** arrived

58 This new magazine is ..... with interesting stories and useful information.
A full **B** packed **C** thick **D** compiled

59 The restaurant was far too noisy to be ..... to relaxed conversation.
A conducive **B** suitable **C** practical **D** fruitful

60 In this branch of medicine, it is vital to ..... open to new ideas.
A stand **B** continue **C** hold **D** remain

Appendix 3 Backward Digits Span Test (English Language)

This auditory task aims to examine participants' complex working memory capacity (storage and processing). Participants are required to (1) listen to sets of digits that are increasing in ascending order, total of the 9 sets. (2) Upon hearing the digits in each set, the participants are required to repeat them in a backward order. The numbers are recorded by the researcher's own voice at one digit per second. The test will be terminated, and participant's working memory span will be determined upon last digits set she has repeated successfully twice (two trails).

Instructions:

You are going to listen to different sets of digits in random order. I will say the numbers and you have to repeat each set **backwards**. Digits will be in increased sets sizes. I will start with sets of three digits and you have 3 attempts each set. When you have two successful attempts, you will move to the next set (4 digits, 5 digits), and so on. The test ends when you fail two times repeating any of the sets.

For example:

When I say: "**4 6 8**"
You say: "**8 6 4**"
Let me know when you are ready.

Set	**Trial 1**	**√/ X**	**Trial 2**	**√/ X**	**Trial 3**	**√/ X**	**Total**
three	456		498		736		
four	5678		2986		5479		
five	43567		21876		46790		
six	456778		256899		780329		
seven	1238765		5601938		0654378		
eight	45609876		04859060		02847565		
nine	876098234		85038365		16829374		
Student name:					**Backward score:**		

Arabic Version of the WM Backward Digit Test

اختبار الذاكرة العاملة (العد العكسي)

تم تصميم هذه الاختبار لقياس سعة الذاكرة العامله من ناحيه (التخزين والمعالجة) لدى المشاركين في البحث. يُطلب من المشاركين الاستماع إلى مجموعات من الأرقام المتزايدة وتكرارها بشكل عكسي. سوف تقدم من خلال ملف صوتي وبفاصل زمني ثانيه واحده بين كل رقم. يتم تحديد سعة الذاكرة العاملة لكل متعلم استنادًا إلى اخر مجموعه من الأرقام التي قام بتكرارها بنجاح مرتين.

تعليمات:

سوف تستمع إلى مجموعات مختلفة من الأرقام. سأقول الأرقام ويجب عليك تكرار كل مجموعة بشكل عكسي. سوف يزداد عدد الأرقام في كل مجموعه بشكل طردي. سنبدأ بمجموعه من ثلاثة أرقام. عندما تكون لديك محاولتان ناجحتان ، تنتقل إلى المجموعة التالية (4 أرقام) وهكذا بشكل متزايد حتى ٩ ارقام كحد اعلى. ينتهي الاختبار عندما تفشل مرتين لتكرار أي من المجموعات

فمثلا:

عندما أقول "٦٥٣"

أنت تقول"٣٥٦"

المدى	المحاولة الاولى	√/ X	المحاولة الثانية	√/ X	المحاولة الثالثة	√/ X
ثلاثة	٥٢٤		٩٣٦		٧١٥	
اربعه	٧٩١٣		٥١٤٦		٩٧٦٢	
خمسه	٤١٥٢٧		٦٤٩٥١		٤١٥٣٩	
سته	٦٣٩٥١٤		١٩٧٢٤٩		٢٦٩٧٢١	
سبعه	٣٩١٥٣٧٢		٩١٧٢٦٣١		٤٩٦٢٤١٣	
ثمانية	٧٢٥٢٩٤١٦		٥٣٧١٩٢٣١		٦٢٩١٦٤٧٣	
تسعه	١٧٣٩٥٦٤٣١		٩٧١٤٩٢٥٦٤		٣١٦٤٩٧٦٢٥	
اسم الطالبة:						
مدى الذاكرة العامله:						

Appendix 4 English Version of Operation Span Test (OST)

You have to complete two tasks in order to answer the test in the next Power Point slide:

1) Read aloud a mathematical problem (operation) and decide if each operation is correct or incorrect (by saying correct or incorrect).

2) Read aloud each word that will appear in the middle of the screen after each mathematical problem and try to remember it.

Operation span test

(9-2)=3 correct? society

8+4=12 correct? bad

2x1=3 correct? plan

4-10=6 correct? cute

(6+1) x 1= 7 correct? fun

(5x2) - 5= 9 correct? test

Arabic Version of Operation Span Test (OST)

اختبار مدى الذاكرة العامله

سيظهر لكِ في الصفحات التالية مجموعات من العمليات الحسابية التي يجب عليكِ

أولا: قرائتها بصوت مسموع ثم تحديد اذا كانت صحيحه عليك القول بصوت مسموع كلمه (صح) اما اذا كانت خاطئة عليك قول (خطاء)

ثانيا: تذكري الكلمة التي سوف تظهر بعد كل عمليه سيطلب منكِ كتابتها في الورقة المرفقة لكِ.

هل ٢٠٠-١٠٠=١٠٠؟

هل ٤-١=٣؟

هل ١٥+١= ١٨ ؟

Answer Sheet of OST

	المجموعة ١	المجموعة٢	المجموعة ٣	المجموعة ٤
	؟؟	؟؟	؟؟	؟؟
١				
٢				
	؟؟؟	؟؟؟	؟؟؟	؟؟؟
١				
٢				
٣				
	؟؟؟؟	؟؟؟؟	؟؟؟؟	؟؟؟؟
١				
٢				
٣				
٤				
	؟؟؟؟؟	؟؟؟؟؟	؟؟؟؟؟	؟؟؟؟؟
١				
٢				
٣				
٤				
٥				
المجموع				

Appendix 5 Monologic Task

Would you agree or disagree that social media, such as **snap chat, twitter,**

Instagram are the current and the future of marketing in Saudi Arabia?

Why and why not?

You can use the following key words in your answer.

advertisement	famous	followers	celebrities	fashionista
attract	popular	products	influencers	accessible

You have 2 minutes to think and 2 minutes to speak. Use *note*s to plan your speech.

Appendix 6 Dialogic Task

With your friend discuss the following question:

Do you agree/disagree with allowing women to drive in Saudi Arabia?

Give your own opinion with reasons. Use the examples below:

1- I agree, but I won't drive, why? I agree and I will drive, why?
2- I have a driving license. I do not have a driving license.
3- Give an experience example of a female in your family/friends who drives a car.
4- Pro and cons / advantage and disadvantage. (e.g. the price of gas, car, driving school)

You have 2 minutes to think and 2 minutes to speak.

Appendix 7 Research Ethical Approval

The Secretariat
University of Leeds
Leeds, LS2 9JT
Tel: 0113 343 4873
Email: ResearchEthics@leeds.ac.uk

UNIVERSITY OF LEEDS

Nada Dafir Alsheehri
School of Education
University of Leeds
Leeds, LS2 9JT

ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee
University of Leeds

15 November 2018

Dear Nada

Title of study: **Examining the relationship between L1 fluency, L2 fluency and working memory capacity in language learners during dialogue performance**

Ethics reference: **AREA 18-038**

I am pleased to inform you that the above research application has been reviewed by the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee and following receipt of your response to the Committee's initial comments, I can confirm a favourable ethical opinion as of the date of this letter. The following documentation was considered:

Document	Version	Date
AREA 18-038 Ethical_Review_Nada alsheehri.doc	2	02/11/18
AREA 18-038 Participant Information Sheet.docx	2	02/11/18
AREA 18-038 university jeddah consent form .docx	2	02/11/18
AREA 18-038 background questionnaires.docx	1	02/11/18
AREA 18-038 Bacward digit test English.docx	2	02/11/18
AREA 18-038 Oxford quick placement test.pdf	1	02/11/18
AREA 18-038 backward digit test Arabic.png	1	18/10/18

Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval as all changes must receive ethical approval prior to implementation. The amendment form is available at http://ris.leeds.ac.uk/EthicsAmendment.

Appendix 8 Institution Information Sheet

The researcher's name: Nada Alsheehri

Research title: Examining the relationship between L2 fluency and working memory capacity in dialogic and monologic performance

The reason of this paper: this paper aims to get the approval of English department administration to conduct the study at the University of Jeddah.

The study purpose: this study aims to examine the relationship between students' second language fluency and their working memory capacity in dialogue performance.

The procedure of the study: if the students decide to contribute in this study, these are the procedures to carry on the study:

- Students will answer background questionnaires, and quick proficiency test both tests will take approximately 30-45 minutes.
- With the researcher in a quiet classroom, each student will receive two working memory tests. The first test (operation span test) will be presented on a laptop screen (PowerPoint) software, and the participant must read sentences and produce answer directly. The second test (backward digit test) is listening to list of numbers and then repeat them in reverse order. It is not a written test.
- Students in pairs, each pair with the researcher in a quiet classroom will be asked to answer a dialogue task: (agree/disagree) statement in Arabic and English and give reasons for your answer.

The general rule of the research

- The participants have the right to ask about the nature of the study and the role of their participations.
- The participation in this study is optional, and participants may withdraw at any time before the researcher has finished data collection and returned to the UK without giving any reasons and with no negative consequences.
- The collected data will be used in the research and will be securely kept on a password-protected computer or in a locked drawer. Only the researcher, the supervisors, and the examiners will have access to the data.
- The dialogue will be audio- recorded and the participants may refuse to participate if they do not want it to be recorded.
- Names will be anonymized and the results will be used in the researcher's thesis.

Appendix 9 Participants' Information Sheet

The researcher's name: Nada Alsheehri

Research title: Examining the relationship between L1 fluency, L2 fluency and working memory capacity in dialogic performance

The reason of this paper :this paper aims to get the approval of the participants to conduct the study at the University of Jeddah.

The study purpose: this study aims to examine the relationship of students' first language fluency, second language fluency and their working memory capacity in dialogue performance.

The procedure of the study: if you decide to contribute in this study, these are the procedures of the study:

- Answer background questionnaires, and quick proficiency test both tests will take approximately 30-45 minutes.
- With the researcher in a quiet classroom you will receive two working memory tests. The first test (operation span test) will be presented on a laptop screen (power point) software and the participant needs to read sentences and produce answer directly. The second test (backward digit test) is listening to list of numbers and then repeat them in a reverse order. It is not a written test.
- Students in pairs, each pair with the researcher in a quiet room will be asked to answer a dialogue task :(agree/disagree) statement in Arabic and English and give reasons for your answer.

The general rule of the research

- The participants have the right to ask about the nature of the study and the role of their participations in the study.
- The participation in this study is optional, and participants have the right to withdraw at any time before the researcher finished data collection and returned to the UK without giving any reasons and without any negative consequences.
- The collected data will be used in the research and will be securely kept on a password-protected computer or in a locked drawer. Only the researcher, the supervisors and the examiners will have an access to the data.
- The dialogue will be audio- recorded and the participants have the right to refuse to participate if they do not want it to be recorded.
- Names will be anonymized and the results will be used in the researcher's thesis.

Appendix 10 Informed Consent form for Students

University of Leeds

School of education

Consent for taking part in the study

I confirm that I have read and understand everything written in the information letter which explains the nature of the study and its procedures and I have the right to ask any questions about my participation or about the study.	
I understand that my participation is optional, and I have the right to withdraw at any time before the researcher finished data collection and returned to the UK without giving any reasons for this and without any negative consequences.	
I agree that the collected data can be used in the research.	
I understand that my name will be anonymised and the results of the working memory tasks, and dialogue will be used in the researcher's thesis and other published paper in the future.	
I understand that the interviews will be audio- recorded.	

Name of student

Student's signature

Date

The Arabic Version of the Consent Form

اقر بأنني قرأت و فهمت كل ما كتب في رسالة التعريف بالبحث والتي توضح طبيعة البحث وطريقة تطبيقه وانه لدي الحق في ان اسال عن دوري في البحث وعن البحث نفسه.	ضع اول حرف من اسمك في حاله موافقتك للمشاركة
اقر بانني فهمت ان اشتراكي في البحث هو اختياري و تطوعي ولي الحق في الانسحاب في اي وقت وذلك قبل ان ينتهي الباحث من جمع بياناته والعودة للملكة المتحدة من دون اعطاء اسباب لذلك ومن دون اي عواقب سلبية.	
اوافق على ان البيانات المجموعة سوف تستخدم في بحث الباحث.	
اقر بانني علمت بان اسمي و بياناتي ونتائج التمارين والمقابلات سوف تستخدم في بحث الدكتوراه للباحث و في اي مطبوعات في المستقبل.	
.فهمت بأن الباحث سوف يقوم بتسجيل المحادثات الحوارية	

اقرار من المعلم و الطالب

اسم المعلم او الطالب

التاريخ

توقيعه

Appendix 11 Coding Symbols

:: double colon used to mark a clause boundary within an AS-unit.

| upright slash to mark an AS-unit boundary

[/] Repetitions

[///] reformations

[/-] false starts

[&] replacement

(1.29) Duration of Mid clause Silent pause of 0.25 or more.

(1.29) Duration of End clause Silent pause of 0.25 or more.

[uh/um/ah] filled pauses

(1.29) Duration of timed mid clause filled pause of 0.25 or more.

(1.29) Duration of timed end clause filled pause of 0.25 or more.

Appendix 12 Samples of Coding Data

A- Sample 1 of monologue coded data

| *I do agree:: with the (0. 34) (uhh) I do agree with that (0.27)*| *[///]*. |*When you saythat social media (0.55):: such as of all its kinds:: such as snap chat twitter Instagram (0.68):: is the current and future of marketing in Saudi Arabia*|. | *I totally agree :: because (uhh) at the main time:: we use social media a lot (0.6)*| . |*We buy:: from the apps the applications (1.06) [&] :: so we do not need to get out (0.46):: from our homes to (uhh) to (0.55) [/] look for clothes or (uh) or [/] other items:: whether is for home or ourselves (0.89)* |. |*(Ummm) (0.93) (0.61) this is a positive side*|. | *But if we were talking about if we will talk about [///] the (uhh) (0.49) negative side of the of this [&] social media (0.74)*|. | *Uhh (0.28) that some celebrity will take advantage of others :: that may (0.64) that may [/] uhh (0.26) advertise about some products:: that maybe harmful for people (1.02)*| and they only do that :: because they wants (0.51) (uhhh) (0.38) because they want [///] money in return:: and they do not care they don't care [/] about the health (0.41) of others| |and some people are immature:: or (0.54) let's say ignorant (0.62) [&] | so we.. they [&] may follow those people (0.74) | |so I really see that:: as (uhh) (0.42) a very bad side (0.25):: of social media (0.63)|. |Also rumours people can use it (uhhh) (0.45) (0.41):: on other people (0.49) (uhh) (0.38) :: (0.31) may be to make them more fame (1) :: or to make a bad repetition sorry a bad reputation [&] of them (0.61)|. |(uhh) (0.62) so (0.5) yes (0.63) :: and (yea)|. |and not to forget the fashionista:: who came (0.47) recently:: who (0.65) (uhhh) (0.47) advertise for an expensive market:: and you may find:: an like more cheaper (0.38) alternative:: in other applications or (0.6) in the markets (0.84) |. |So (yea).|

Participant: 1		**Monologue task**	**Time:**135 sec
Silent pauses	13	End-clause FPs	0
Filled pauses (um/uh)	8	Repetitions [//]	4
Mid-clause SPs	4	Reformations [///]	2
End- clause SPs	9	False starts [/-]	0
Mid-clause FPs	4	Replacements [&]	1

B- Sample 2 of monologue coded data

|Today:: I'm going to talk:: about social media (0.46):: and it's (uhhh) (0.41) the future of marketing:: in Saudi Arabia|. |Nowadays many ca.. (0.27) companies [/] :: and business owners (0.46):: rely on social media influencer (0.39):: (uuhh) (0.42) to show their products:: and what they produce:: and their man.. manufactures (0.38) [/] manufactures [/] (0.57)| . |However, (0.29):: (uhh)(0.46) (0.52) not all media influencers:: are being honest with their fans:: and telling the truth (0.36):: about what they advertise for (0.26)|. |Many of products may not be good:: for personal use (0.41) :: and the company has a lot of [/-] |. |(Uuh) we are talking:: about (0.68) a huge numbers of products:: that are not good uhh (0.51) for personal use:: and are damaged (0.52) | |and they hire (0.5) [/-] and how they can get rid:: of this damaged products|. |They hire: media influencer (0.43):: and they them they [/] [///] have to do this| and (0.46) (uhhh) (0.97) you know (0.98)| |(Ummm) (0.39) (0.53) they have to (0.59) speak all:: about the positive aspects of the product (0.61)| |and they advertise and then and [/-] with a blink of with a blink of an eye (0.45) [///]:: all the damaged products

are being sold out|. |And this is (0.51) increase:: the fraud and all of social media influencer are being in prison (0.33):: uhhh (0.39) because of that (1.02) |. |Thanks.|

Participant: 2		**Monologue task**	**Time:**81 sec
Silent pauses	25	End-clause FPs	0
Filled pauses (um/uh)	8	Repetitions *[//]*	4
Mid-clause SPs	10	Reformations *[///]*	2
End- clause SPs	15	False starts [/-]	3
Mid-clause FPs	6	Replacements *[&]*	0

C- Sample 3 of monologue coded data

Marketing on uhh (0.47) **(0.96)** the **(0.35)** electronic marketing or advertising **(0.70)**| |uhh (0.60) because uhh (0.55) **(0.57)** it's uhh (0.56) sometimes it's [/] affordable **(0.62)** :: and it save time. **(0.67)**| |And umm (0.45) **(0.80)** I'll share my experience:: uhh (0.64) **(0.56)** one time I bought umm (0.52) **(0.57)** a jacket from some website **(0.94)**| |uhh (0.76) **(0.82)** uhh (0.39) I liked it:: but it was uhh (0.43) a little bit expensive **(0.78)**:: but umm (0.46) **(0.59)** [/-] so I bought it because I like it **(0.58)**.| |Then uhh (0.62) I found it in another website:: it was more cheaper **(0.39)**| |and it was the same quality and same colour and everything **(0.55)**:: but the price was uh (0.48) more affordable and suitable for me **(0.58)**.| |And uhhh (0.73) here is the point :: that **(0.73)** **uhh** (0.30) you have to look **(0.29)** and not just buy anything from any website. **(0.37)**| |You have to look :: and you have to search **(0.32)** about the suitable uhh (0.47) website.| |Because there is some website like trick **(0.40)** people in prices (0.64)|. |so uhh **(0.54)** if you will search :: and be [/] uhh (0.47) **(0.82)** and be [/] (2.62) like plan uhh (0.46) on **(0.34)** whether website you will buy **(0.71)** :: uhh (0.35) that will be so uh (0.38) **(0.37)** benefit

for you **(0.77)**| |and uhh (0.64) it [&] you will agree with it :: if you **(0.34)** know :: how to get it ri.. right [/] **(0.93)**| |and umm (0.61) **(2.8)** that it is it|.

Participant: 3		**Monologue task**	**Time:**99 sec
Silent pauses	32	End-clause FPs	0
Filled pauses (um/uh)	23	Repetitions *[//]*	4
Mid-clause SPs	19	Reformations *[///]*	0
End- clause SPs	13	False starts [/-]	1
Mid-clause FPs	23	Replacements *[&]*	1

D- Sample 4 of monologue coded data

|In the social media (0.29):: in this I.. I called the phenomen (uhh) the phenomena (0.58) [/]| |because it is something new| |and that (0.75) control:: and in all the society all the (0.35) cata..age categories [/]:: from adults, from (0.34) children, mothers (0.37):: even the (uhhh) (0.5) old (0.38) (uhh) (0.55) (0.66) the people old (0.27) [///]| |and have (uhh) (0.95) old (0.37) :: like in fifty fifty [/] or (0.63) forty (0.68)| . |(ummm) (0.72) I like the social media:: because it is (umm) (0.48) (0.34) as I said something new (0.41)|. |(Uhh) (0.74) I.. I (0.52) [/] have some comments:: that making uhh (0.44) (0.4) this (0.39) is have or has [&] ummm (0.48) effective side (0.75)|. |(Ummm) (0.69) (0.44) everything (0.55):: when we use in the correct (0.59) use (0.47):: it uuhh (0.54) (0.27) will be some positive (0.54)| |buz ..but [&] I think I I [/] show (0.33):: like my family or umm (0.6) (0.38) friends (0.35):: maybe they (2.03) become addiction:: with (0.28) social media (0.3)| .

Participant: 4		**Monologue task**	**Time:**62 sec
Silent pauses	31	End-clause FPs	0
Filled pauses (um/uh)	12	Repetitions *[/]*	5
Mid-clause SPs	16	Reformations *[///]*	1
End- clause SPs	15	False starts [/-]	0
Mid-clause FPs	12	Replacements *[&]*	2

E- Sample 5 of monologue coded data

|My opinion about (ummm) (1.06) marketing in social media:: so I think it is (1.04) in general (0.52):: (uhhh) (0.84) social media is really helped us :: to reach to a lot (0.43) of (0.75) items that we (0.25) we [/] can't buy it in the normal (0.33) shop| |(uhh) (0.37) when I go to the mall:: or anything like that (0.6):: it is easier to me to buy from social media or anything like it (0.99) |but (uh) (0.76) I have really bad experience| |that is the thing that make me now (0.48) feel scared:: to buy anything| |it is when I bought from Instagram (0.58)::, it was really sucks:: the size was smaller ::and the colour was way different| |because I ordered black ::and it came to me red (0.93)|. | And uhh (0.69) but it is fare [/-] ::because it makes us ::to reach the (0.28) items easier to us (1)|. |Celebrities they use their names:: to (0.25) get money (0.39):: but there is no honesty ::in what they (1.01) marketing for (0.33)|. |Like there is some people (uhh) (0.53) (0.48):: there is someone [///] ::(uhh) there is a celebrity male (0.53) [&]| |He (1.03) (uhhh) (0.95) he [/] gave an advertisement:: for an (0.51) oil (0.4):: for hair (0.42) growing|. |He said it is really good (0.6):: and lab lab [/] lablaaa:: but it was really sucks:: because a lot of people try it:: and said it was a lie

just any normal (0.7) hair oil|. |Like it does not make any different (0.43):: because It is just make you're her wet (0.95)|. |(Ummm) (1) some people they are just (0.67) over (0.25) reacting:: and they are over acting (0.27)| either |like (uhh) this celebrity in social media| |like (0.82) three or four days ago:: her name is Shoug (0.56)| |she said:: she is Saudi (0.63)|. |She was [///] (umm) (0.78) (0.8) she (0.43) she [/] is like being main face for (0.35) golden brownies|. |She started talking:: said put chips between (0.45) your piece of brownie|. |It is just a heaven in your mouth|.

Participant: 5		**Monologue task**	**Time:**128 sec
Silent pauses	48	End-clause FPs	0
Filled pauses (um/uh)	11	Repetitions *[/]*	4
Mid-clause SPs	20	Reformations *[///]*	2
End- clause SPs	18	False starts [/-]	1
Mid-clause FPs	11	Replacements *[&]*	1

F- Sample 6 of monologue coded data

|(Uhh) (0.51) (0.36) I think:: the (0.59) [/-] i'm agree (0.52):: with Umm (0.67) (0.44) snap chat twitter Instagram (0.45)| |(umm) (0.45) (0.27) they are consider:: as (0.82) standards (0.39):: or spaces to attract the people (0.63):: to buy (0.81) (uhh) (0.39) different product (0.34)| |but I think (0.86):: (uh) (0.56) that (0.82) we should have some (1.02) (uhh) like terms or conditions (0.46):: when we deal with these (0.41) (uhh) (0.64) social media (0.48)| |because sometimes the products (uhh) (0.79) its quality is not good (0.39) :: or the price is very high (0.34):: (ummm) (0.48) (0.4) when we compare to another (0.35) place (0.98)| |(uhhh) (0.6) and I think:: it's

(0.31) sometimes bad| |because (0.36) (uhhhh) (0.64) (0.41) some people don't have the (0.77) nnn... (0.35) uhh (0.6) knowledge [/]:: about products (uhh) ::about (0.6) the same things qualities (0.47):: and these things (0.89)|. |Yeas that is it|.

Participant: 6		**Monologue task**	**Time:**68 sec
Silent pauses	30	End-clause FPs	0
Filled pauses (um/uh)	13	Repetitions *[//]*	1
Mid-clause SPs	18	Reformations *[///]*	0
End- clause SPs	12	False starts [/-]	1
Mid-clause FPs	4	Replacements *[&]*	0

G- Sample 7 of monologue coded data

| My opinion:: about the [/] the (0.49) (uhhh) (1) about this about marketing [///]:: in Saudi Arabia:: about (uhh) in online (0.46)|. | I agree and disagree:: in (uh) at [&] the same time:: because it's (0.35) it is [/] does not mean:: if (0.4) this celebrities:: have huge followers (0.5):: or a cheap price|. | It is mean that a good quality (0.31)| . | It does not mean:: that because you can't see it:: in front of yours eyes (0.45) :: you can't touch it:: and (uhh) (0.66) feel (0.66) it|. |so (ummmm) (0.75) (1.02) yes [/]| | it is good and not good in that (0.27) (uhh) the same time (0.41)| | we can buy some (0.38) of (0.35) these:: but not all of it|. |you can decided:: your own (0.27):: if it's (0.34):: ok to buy it (0.37):: online or not (0.44)|. |About me:: I don't think uhh (0.36):: I can buy everything online (0.39)|. |(uhhh) (0.78) especially if it comes:: to my skin and my body (0.31)|. | I can't use anything abaaa about [/] it (0.83)|. |so I'm not that kind of person:: that if I see:: it everybody use it (0.26):: I will use it with them|.

|No. (0.26)| |I prefer (ummm) (0.59) (0.57) touch it:: and (0.91) feel it:: myself and read:: about it a lot|.

Participant: 7		**Monologue task**	**Time:**82 sec
Silent pauses	13	End-clause FPs	0
Filled pauses (um/uh)	9	Repetitions *[/]*	4
Mid-clause SPs	10	Reformations *[///]*	1
End- clause SPs	14	False starts *[/-]*	0
Mid-clause FPs	9	Replacements *[&]*	1

H- Sample 1 of dialogue coded data

B: I'm [/] I'm [/] I'm **(0.51)** **[/-]** every like every uhhh few days:: I will ask her **(0.40)** **::**, are you sure **(0.50)** are you sure [/] this summer? [19.65]

K: |umm (0.71) (0.42) umm (0.51) me (1.19) I want I want [/] :: to uhh (0.37) to [/] drive (0.72)|. |Just try (0.99) :: Sa (0.32) some [/] kilometres just.| [11.88]

B: | uhh (0.37) I think **(0.33)::** it's (0.88) it's [/] ridiculous:: how much they are asking **(0.50)** for **(0.41)** **price::** to **(0.43)** to teach [/] **(0.49)** in the dra in the driving [/] school. | [9.63]

K: yea yea sure. [*]

B: | 2600 I think **(0.35)**:: too much.| [3.34]

K: too much. Sure

B: |Comparing to men **(0.45)**,:: it is not **(0.43)** it is not the same[/]| . |I think **(0.53)::** 400 or something like that|.|For **(0.48)** **[/-]** **and** **(0.36)** when you're going ::to get your license**(0.27)**.| |When it is finish **(0.31)** you also have to pay **(0.71)**.| |It is not..[&] it is ridiculous how much.| [13.90]

B: |I think:: also there umm (0.26) gas stations and **(0.51)** like that**(0.75)**|. |I think:: because women don't know how **(0.60)** how [/] much it actually cost **(0.38)** | |yea:: maybe they will **(0.36)::** do some scheme or something like that| [12.35].

K: |yeah,:: maybe|. [0.41]

B: |my friend's aunt **(0.34)** she drives **(1)**:: but uh she just got her license **(0.66)** and the first day she got out the street::, she drove straight:: to the house to their stairs **(0.91)**|. |Then her brother **(0.48)** said:: you will never let you drive again.| [18.86]

B: I actually **(0.58)** **[/-]**, I thought:: ma maybe [/] like **(0.53)** two years ago or something like that.| |When all the boy in my family:: try drive starting to drive [///] **(0.75)** |. |They got they got [/] into [/-]**(0.71)** |. |We have a farm outside **(0.43)::** and we got also ..all in the cars:: and **(0.80)** start practicing **(0.31)**| |and I was so **(0.41)** jealous, I want:: to go with them:: but I was like so scared ::they are not gonna let me **(0.57)**|. |I was like:: I don't know **(0.50)** 16 ::or something like that **(0.68)** |. |And I don't know:: how to drive:: so it is out of question **(0.70)** not going to drive.| |But I was so envies from them:: because I want ::to drive **(0.60)**| |because I think:: it's it's [/] not something.. yeah **(0.58)** it is not big of of dealing **(0.61)** **[///]** I mean. | | It is ok:: we drive now:: it is no problem. | |yeah:: and there were so many people in social media against it **(0.36)::** and I don't

know:: why it is not that big of deal just driving **(0.91)**.| |It's better to ..than to go [///] with driver **(0.64)**:: And someone you don't know. | [64.95]

K: |it is easy easy yeah [/] :: they are like ::what's happened.| [14.95]

K: |uhh when when [/] I drive,:: it was like a party :: ohh I'm drive I'm drive [/] finally.| [10.80]

B: yeah it is true.

Student	Speaking time	fluency measures	N	duration	student	Speaking time	fluency measures	N	duration
BA.	171	**MidSPS**	27	14.28	**KO.**	47.14	**MidSPS**	3	1.93
syllable	570	**EndSPS**	30	16.73	**syllable**	87	**EndSPS**	5	3.47
Refor.	4	**Filled P**	7	0.96	**Refor.**	0	**Filled P**	7	2.03
Fasle S	5	**Silent P**	57	31.01	**Fasle S**	0	**Silent P**	8	5.4
Repetiti.	12	**MidFPS**	6	0.96	**Repetit.**	6	**MidFPS**	5	2.03
Replace.	0	**EndFPS**	0	0	**Replace**	0	**EndFPS**	0	0

I- Sample 2 of dialogue coded data

F: |Uhh (1.17) when you ask me about umm (0.61) do I agree about umm (0.35) **(0.38)** allowing women ::to drive in Saudi Arabia **(0.62)**. | Uhh (0.38) for me in my opinion I do agree:: that women should uhh (0.54) drive **(0.57)** ::whether in Saudi Arabia or other countries or other societies **(0.41)**.| |I find it very acceptable **(0.62)**:: because the women will be uhh (0.67) women [/] (0.86) will be more uhh (0.29) independent (0.60) [///] on their own.| |They will do their own tasks uh (0.56) **(0.37)**:: if they have uhh (0.56) **(0.35)** if they have [/] school university, or whatever as a job,

or something like that|. |They will go:: to it themselves without uhh without [/] the need for a driver or uhhhh (0.56) (0.55) uhhh (0.58) :: let say (0.67) their father or father [/] or husband (0.49) or etc (0.88). | |So I do agree (0.49).| |And uhh (0.63) and uhh (0.60) for family side.| [58.47]

S: |Do you have a driving license? | [2.92]

F: |Not for me:: I don't have uhh (0.46) a driving license :: and I'm not (0.66) thinking of driving although I find it acceptable Uhh (0.48) (0.30)|. | I do uhh (0.31).. [/-] I find someone it is uhh (0.33) like more available:: to do the things..the uhh to take me [///] for where I go (0.50):: without any lost of money or uhh anything, (0.61) | |so|. [24.23]

S: |Uhh you surprise me|.[1.60]

F: |No:: if I would need if I would like say (0.48) [///] :: uhh (0.51) if there is no one available for me [///] ::to hold me to take me [&].| [8.61]

S: |you will go out yourself|? [1.99]

F: |I will do that of course:: but at the meantime I'm not thinking of it (0.33) | |and I'm thinking I'm too young to do it:: I'm and the [/-] .. although there is no rela relationship [/] or it's uhh (0.30) (0.34) irrelevant that I'm too young (0.32) ::but I don't have the courage to do it in the meantime.| [17]

S: |yes:: me too. | |I will think:: of it uhhh (0.43) after five years maybe or three (0.35). |Or maybe :: when I'm uhh (0.40) (0.75) employ (0.31) or something. | [8.99]

F: |Have you drove before? | [0.94]

S: |Uhhh (0.43) .. (0.34) I try :: but not uhhh (0.37) (0.37) excellent no:: faraway from excellent. | |Uhh (0.64) (0.71) do you have example in your life (0.52) female or family friends ::who drive?| [13.63]

F: |Uhh a friend of my mother:: uhh (0.4) she is … she uhhh (0.35) [/] she is [/] a teacher:: she has a job.| |She has to go:: to to [/] the school and go back. | [12]

S: yes, I love the idea of... [*]

F: |She will lost money:: if she uhh (0.87) take foreign driver ::to take her and etc.| |And if she relate on her husband:: he may busy at sometimes :: and she may get late and (0.71).| |Although she said:: she got lots of accidents, Umm like uhh (0.68) three or two uhh (0.80) like three or uhh (1.15) two [/] or third times (0.36) uhh (0.42) at one day|. |Uhh (0.49) although it's uhh it's [/] for her it's fine [///], :: it is not a big deal or problem.| [30.69]

S: |ok:: Do you agree the price of the gas?(0.83)| |do you agree (0.50):: or find it (1.84) uhh suitable (0.34):: that the price of the gas is so expensive.| [13.10]

F: |well well [/] I don't have experience with it. | [3.72]

S: |don't you go with your (0.50) father or brother? | [2.72]

F: |uhhh (0.57) .. (0.42) as I hear them :: to say it is expensive:: but for me, I don't have experience| |so I can't..| [5.38]

S: |If you see before and after::, you will see the big difference.| [3.56]

F: |so for you:: do you see it a big problem:: that the gas is expensive?| [4.46]

S: |the gas or the cars?| [1.44]

F: |the gas|. [0.40]

S: |No.. yes [&] :: I found it so expensive (0.38) ummm (0.37) for college student (0.56)|. |Maybe if I (0.41) uhh (0.38) drive by myself (0.37)::, I don't think that I will every day (0.73)|. |And from I live:: in the farms.| |The way from there to here thirty minutes:: so take a lot of time and a lot of money.| |Thirty minutes:: I go and back (0.34) |. [26.84]

F: |so:: you said earlier umm (0.43) I think:: you have experience that your uhh (0.32) |. [5.69]

S: |Yes:: Uhh (0.49) (0.54) my mom is Kuwaiti:: so I have three aunties (0.51) uhh (0.53) all of them:: drive (0.61) from (0.87).| [8.30].

F: |a long time. | [1.03]

S: yes. I find it so (*)

F: |you feel you are proud of that. All women and aunties. | (3.29)

S: |yes yes [/]:: I enjoy (0.31) being with females.| |I don't like men (0.33) at all (0.31):: uhhh (0.60) (1.11) but ::I agree.| [8.58]

F: |why you don't like men?| [0.91]

S: |umm (0.60) (0.71) from our cul culture [/].| [1.30]

F: uhh the society. [*]

S: |yes:: I will wear my Abaya and my ankle and my voice and (0.62) a lot of things I think:: about it with men| |but with my.. | [10.14]

F: |but they like give you a proof:: uhh (0.33) for driving:: they did not uhh (0.39) (0.32) they did not [/] say disproof. | [6.53]

S: |my family? |

F: |yea. | **[0.27]**

S: |My father agree:: but not in the meantime.| |He say (0.42):: after 2 years or three:: because I'm young (0.99)|. |I see myself. He always see [///]me:: young no matter what.| [16.90]

Student	Speaking time	fluency measures	N	duration	student	Speaking time	fluency measures	N	duration
FA.	**183.62**	**MidSPS**	**12**	**6**	**SA.**	**122.01**	**MidSPS**	**14**	**10.12**
syllables	**598**	**EndSPS**	**14**	**7.21**	**syllables**	**383**	**EndSPS**	**12**	**6.61**
Refor.	**5**	**Filled P**	**33**	**17.09**	**Refor.**	**1**	**Filled P**	**11**	**4.64**
Fasle S	**2**	**Silent P**	**26**	**13.21**	**Fasle S**	**0**	**Silent P**	**26**	**16.73**
Repetiti.	**12**	**MidFPS**	**32**	**17.09**	**Repetiti.**	**2**	**MidFPS**	**11**	**4.64**
Replace.	**0**	**EndFPS**	**0**	**0**	**Replace.**	**0**	**EndFPS**	**0**	**0**

J- Sample 3 of dialogue coded data

MA: |My name is Mawadah| |and today my friend (0.75) Nawal we are going:: to discuss the uh (0.68) the ex.. the topic [&] umm (0.46) (0.67) allowing women:: to drive in Saudi Arabia (0.37) |. |My friend::, do you agree or disagree:: with this new law in Saudi Arabia? | [17.06]

NA: |yes::, I agree:: and uhh I **(0.49) [/-]** but I won't drive. | [4.69]

Ma: |but you won't drive|. |Ok but for me agree:: I totally agree [///]:: with this new law in Saudi Arabia| |but unfortunately I won't be driving my own car:: because I don't know how to drive| |and I don't have my driving license|. |And also I'm a little afraid of touching the wheel. | [16.78]

NA: |I also I ha.. **[/]** **[/-]** I don't have a driving license **(0.88)** | |but I I'm agree :: to drive women**(0.54)** |. **[7.93]**

MA: |OK (0.62)::, Uh give me an experience of a female in your family or friends:: who drive thier car. | [4.90]

NA: | Uh (0.29) like mother maa [/] mother my friend (0.49) ::she drive|. [6.25]

MA: |For me::, I know many of my friends and family :: who drive their own car (0.35) | |and this help them so much :: that they won't need uhh Uber and and [/] personal drivers anymore | |and they won't face any problem:: like sexual harassment (0.40) or any (0.38) dis.. discomfort [/] with the drivers.| [19.35]

MA: |in your point of view:: do do [/] you know [/-]. can you list some pros and cons| you know the advantages and disadvantages :: of (0.74) uhh the women driving the car|. [10.33]

NA: |it is ummm (0.46) advantages **(0.30) ::**because it is it is important decision **(0.41)** for women| |and it is a right of her right **(0.41) | |**and this help them a lot in their daily work **(0.26)::** such as she can uhh (0.29) go ::to his work **(0.52)** and his son's school|. [19.53]

MA: |uh yes:: that's correct|. |For me I think:: uhh (0.29) the advantages are (0.49) [/-] umm we can't uhh you know uhh list the advantages :: because they because they [/]are a lot (0.51) |. | Uh (0.33) one advantages is ::that the woman is umm (0.27) independent (0.52) |. |She can do ::whatever she wants with her wheel (0.41) |. |For the disadvantages::, there is uhh (0.4) tremendous (0.92) amount of uhh (0.44) car accidents (0.57) uhh (0.49) the beginning of uhh (0.44) 2019 :: because women are allowed to| |and this uhh (0.51) (0.63) because women don't know ::how to drive the car correctly| |that's why the car accident rate uhh (0.40) has grown up (0.57) very fast|. |Also the price of the gas||umm (0.41) yes as we know:: from the 2018 (0.41) uhh (0.45):: you know uhh the (0.47) you know [/] the gas and (0.29) petrol uhh (0.33) has the price is (0.46) uh re re [/] raised up | |and this cause (0.68) uhh a lot of women:: who wants to drive| |but they don't have like uhh (0.37) monthly income (0.58):: to drive their own car |. [65.69]

NA: |Ok **(0.28)::,** also disadvantage the price of gas|. [2.66]

MA: yeah that's correct. Thank you.

Student	Speaking time	fluency measures	N	duration	student	Speaking time	fluency measures	N	duration
MA	**134.11**	**MidSPS**	**13**	**7.53**	**NA**	**41.06**	**MidSPS**	**3**	**1.42**
syllables	**461**	**EndSPS**	**7**	**2.99**	**syllables**	**119**	**EndSPS**	**7**	**3.16**
Refor.	**1**	**Filled P**	**15**	**6.27**	**Refor.**		**Filled P**	**3**	**1.04**
Fasle S	**2**	**Silent P**	**20**	**10.52**	**Fasle S**	**2**	**Silent P**	**10**	**4.58**
Repetiti.	**6**	**MidFPS**	**15**	**6.27**	**Repetitions**	**2**	**MidFPS**	**3**	**1.04**
Replace.	**1**	**EndFPS**	**0**	**0**	**Replace.**		**EndFPS**	**0**	**0**

K- Sample 4 of dialogue coded data

R: |I agree:: but I don't think to drive soon:: because uhh (0.37) I have uhh (0.39) **(0.46)** my own driver **(0.84)** |.|Uhh (0.46) I don't have any problems uhh (0.53) of driving **(1.70)** or learning ::how to drive|. |I'm planning ::to get a license soon:: but not to drive| |It is just for necessary cases|. |What about you? | [20.14]

J: |Uhh (0.4) also I I agree [/] :: but I won't to drive (0.58) right now:: because I'm also have my own uhh (0.49) driver (1.19) |. |uhhhh (1.43) and I agree:: with alla (0.34) al [///] allowing women to drive in Saudi Arabia:: because there is a lot of families in Saudi Arabia need to do their own activities or works|. [21.72]

R: |yes:: by their selves:: without asking anyone in their house|. [2.67]

J: | yes,:: to be uhh (0.39) like umm (0.28) to be like [/] uhh (0.43) independent woman. | [3.80]

R: |uhh (0.27) yes:: and uhh (0.30) some families don't have men **(0.52)** ::so they need to.. | [4.90]

J: |or they have:: but they are selfish (0.40) |. |Do not look:: at uhh (0.44)their families|. [5.77]

R: |Maybe (0.49). | |Uhhh (0.42) two of my friends are driving| |and they don't:: face any problems (0.74) except the high prices of everything:: such as gas and taxes (0.70) |.| What do you think: :of the disadvantages of uhh (0.37) driving? | [15.40]

J: |uhh (0.52) disadvantage umm (0.52) the price and the buy of the cars (0.52):: and uhh (0.67) the accidents (0.82) and uhhhh (0.89) (0.28) yeah (1.46) |. |And also (0.30) all of my friends (0.55) uhhh (0.39) are driving:: and one of my sisters|. |And uhh (0.58) (1.08) yeah:: that's it. | [23.52]

R: |ok::, some of the advantages (0.47) according to my opinion:: is independence|. |You will not have:: to ask anyone ::to go with you anywhere (0.70) |. |Uhh (0.36) and the disadvantages as I said before:: it is the high costs. | [15.05]

J: |for me the advantage is uhh (0.29) to be (0.27) ummm (0.67) [/-] I can uhh (0.54) (0.5) do my own works (0.31):: to myself (0.44) uhh (0.31) (0.5) not asking anyone (0.36) :: "Are you going there? are you going there? [/], I want the car".| |You know:: So I can be uhh (0.38) independent woman (1.20) | |Uh (0.37) and I said the disadvantage before |. [21.61]

R: |yes:: and as a student we can't buy car by ourselves|. |So we need help ::from our family of course. | [8.51]

J: yes. Yeah yeah.

R: |so right now:: I can't buy a car (0.29) ::to be honest (0.38). | [3.70]

j: | so you want:: to be uhh (0.56) a driver (0.76) woman? | [3.92]

R: |no:: (0.25) uhh (0.61) I don't like :: to drive| |but I will learn (0.59) :: just to be uhh able to drive in necessary cases.| [8.07]

Student	Speaking time	fluency measures	N	durat ion	student	ST	fluency measures	N	duration
J.	**80.34**	**MidSPS**	**12**	**6.41**	**R.**	**78.44**	**MidSPS**	**4**	**3.16**
syllables	**232**	**EndSPS**	**7**	**5.44**	**syllables**	**257**	**EndSPS**	**8**	**3.81**
Refor.	**1**	**Filled P**	**25**	**9.97**	**Refor.**	**0**	**Filled P**	**10**	**4.08**
Fasle S	**1**	**Silent P**	**19**	**11.8**	**Fasle S**	**0**	**Silent P**	**12**	**6.97**
Repetition	**3**	**MidFPS**	**25**	**9.79**	**Repetit.**	**1**	**MidFPS**	**10**	**4.08**

Replace.	**0**	**EndFPS**	**0**	**0**	**Replace.**	**0**	**EndFPS**	**0**	**0**

L- Sample 5 of dialogue coded data

F: |Uhh (1.17) when you ask me about umm (0.61) do I agree about umm (0.35) **(0.38)** allowing women ::to drive in Saudi Arabia **(0.62)**. | Uhh (0.38) for me in my opinion I do agree:: that women should uhh (0.54) drive **(0.57)** ::whether in Saudi Arabia or other countries or other societies **(0.41)**.| |I find it very acceptable **(0.62)::** because the women will be uhh (0.67) women [/] (0.86) will be more uhh (0.29) independent (0.60) [///] on their own.| |They will do their own tasks uh (0.56) **(0.37)::** if they have uhh (0.56) **(0.35)** if they have [/] school university, or whatever as a job, or something like that|. |They will go:: to it themselves without uhh without [/] the need for a driver or uhhhh (0.56) **(0.55)** uhhh (0.58) :: let say **(0.67)** their father or father [/] or husband **(0.49)** or etc **(0.88)**. | |So I do agree **(0.49)**.| |And uhh (0.63) and uhh (0.60) for family side.| [58.47]

S: |Do you have a driving license ? | [2.92]

F: |Not for me:: I don't have uhh (0.46) a driving license :: and I'm not **(0.66)** thinking of driving although I find it acceptable Uhh (0.48) **(0.30)|.** | I do uhh (0.31).. [/-] I find someone it is uhh (0.33) like more available:: to do the things..the uhh to take me [///] for where I go **(0.50)::** without any lost of money or uhh anything, **(0.61)** | |so|. [24.23]

S: |Uhh you surprise me|.[1.60]

F: |No:: if I would need if I would like say **(0.48)** **[///]** :: uhh (0.51) if there is no one available for me [///] ::to hold me to take me [&].| [8.61]

S: |you will go out yourself|? [1.99]

F: |I will do that of course:: but at the meantime I'm not thinking of it **(0.33)** | |and I'm thinking I'm too young to do it:: I'm and the [/-] .. although there is no rela relationship [/] or it's uhh (0.30) **(0.34)** irrelevant that I'm too young **(0.32)** ::but I don't have the courage to do it in the meantime.| [17]

S: |yes:: me too. | |I will think:: of it uhhh (0.43) after five years maybe or three (0.35). |Or maybe :: when I'm uhh (0.40) (0.75) employ (0.31) or something. | [8.99]

F: |Have you drove before? | [0.94]

S: |Uhhh (0.43) .. (0.34) I try :: but not uhhh (0.37) (0.37) excellent no:: faraway from excellent. | |Uhh (0.64) (0.71) do you have example in your life (0.52) female or family friends ::who drive?| [13.63]

F: |Uhh a friend of my mother:: uhh (0.4) she is … she uhhh (0.35) [/] she is [/] a teacher:: she has a job.| |She has to go:: to to [/] the school and go back. | [12]

S: yes, I love the idea of... [*]

F: |She will lost money:: if she uhh (0.87) take foreign driver ::to take her and etc.| |And if she relate on her husband:: he may busy at sometimes :: and she may get late and **(0.71)**.| |Although she said:: she got lots of accidents, Umm like uhh (0.68) three or two uhh (0.80) like three or uhh (1.15) two [/] or third times **(0.36)** uhh (0.42) at one day|. |Uhh (0.49) although it's uhh it's [/] for her it's fine [///], :: it is not a big deal or problem.| [30.69]

S: |ok:: Do you agree the price of the gas?(0.83)| |do you agree (0.50):: or find it (1.84) uhh suitable (0.34):: that the price of the gas is so expensive.| [13.10]

F: |well well [/] I don't have experience with it. | [3.72]

S: |don't you go with your (0.50) father or brother? | [2.72]

F: |uhhh (0.57) .. (0.42) as I hear them :: to say it is expensive:: but for me, I don't have experience| |so I can't..| [5.38]

S: |If you see before and after::, you will see the big difference.| [3.56]

F: |so for you:: do you see it a big problem:: that the gas is expensive?| [4.46]

S: |the gas or the cars?| [1.44]

F: |the gas|. [0.40]

S: |No.. yes [&] :: I found it so expensive (0.38) ummm (0.37) for college student (0.56)|. |Maybe if I (0.41) uhh (0.38) drive by myself (0.37)::, I don't think that I will every day (0.73)|. |And from I live:: in the farms.| |The way from there to here thirty minutes:: so take a lot of time and a lot of money.| |Thirty minutes:: I go and back (0.34) |. [26.84]

F: |so:: you said earlier umm (0.43) I think:: you have experience that your uhh (0.32) |. [5.69]

S: |Yes:: Uhh (0.49) (0.54) my mom is Kuwaiti:: so I have three aunties (0.51) uhh (0.53) all of them:: drive (0.61) from (0.87).| [8.30].

F: |a long time. | [1.03]

S: yes. I find it so (*)

F: |you feel you are proud of that. All women and aunties. | (3.29)

S: |yes yes [/]:: I enjoy (0.31) being with females.| |I don't like men (0.33) at all (0.31):: uhhh (0.60) (1.11) but ::I agree.| [8.58]

F: |why you don't like men?| [0.91]

S: |umm (0.60) (0.71) from our cul culture [/].| [1.30]

F: uhh the society.

S: |yes:: I will wear my Abaya and my ankle and my voice and (0.62) a lot of things I think:: about it with men| |but with my.. | [10.14]

F: |but they like give you a proof:: uhh (0.33) for driving:: they did not uhh (0.39) **(0.32)** they did not [/] say disproof. | [6.53]

S: |my family? |

F: |yea. | **[0.27]**

S: |My father agree:: but not in the meantime.| |He say (0.42):: after 2 years or three:: because I'm young (0.99)|. |I see myself. He always see [///]me:: young no matter what.| [16.90]

Student	Speaking time	fluency measures	N	duration	student	ST	fluency measure	N	duration
F.	**183.62**	**MidSPS**	**12**	**6**	**SA.**	**122.01**	**MidSPS**	**14**	**10.12**
syllables	**598**	**EndSPS**	**14**	**7.21**	**syllables**	**383**	**EndSPS**	**12**	**6.61**
Refor.	**5**	**Filled P**	**32**	**17.09**	**Refor.**	**1**	**Filled P**	**11**	**4.64**
Fasle S	**2**	**Silent P**	**26**	**13.21**	**Fasle S**	**0**	**Silent P**	**26**	**16.73**
Repetiti.	**12**	**MidFPS**	**32**	**17.09**	**Repetitions**	**2**	**MidFPS**	**11**	**4.64**
Replace.		**EndFPS**	**0**	**0**	**Replace.**		**EndFPS**	**0**	**0**

M- Sample 6 of dialogue coded data

AM: |OK, mmm (0.72) (0.73) give me your opinion, uhh (0.62) with reasons of course (0.78), uhh (0.67) of uhh (0.49) allowing women:: to drive in Saudi Arabia (0.66), ::do you agree or disagree? | [12.11]

RA: |Yes| |I agree:: (0.47), but uhh (0.45) (0.59) I won't drive.| [4.28]

AM: |Why? | [1.0]

RA: |Because I am scared from (0.39) driving.| [3.01]

AM: |OK:: to me, uhh (0.72) (I) (I) I [/] agree, of course I agree [/]:: and will drive someday (0.27)| |but I'm just waiting until:: I graduate and I have the time :: to go to driving school | [10.91]

RA: |take the lessons?| [1.09]

AM: …|yeah (0.44) | |take the lesson and go ::to driving school, (0.76):: so I get my licence and drive (0.30)|. |I hope:: my Dad will agree on driving (0.85)| |I think he will agree :: because now he's training me :: to drive (0.51)| | like some day like [/] umm(0.52) (1.05) he's putting this kind of oil in the gas thin:: I don't know what you call it:: yeah, so so [/] the car will be clean from the inside (0.66).| |So, we go to …we (0.36) went to [///] Amman road (0.75) :: yeah, so I took all the road:: I was driving:: it was so exciting:: I love it (1.33).| |Uhhh (0.51) So ummm (0.74) (1.94) of course you don't have driving license| |mm (1.39) oK (0.44) is there any one in your family have drive before? | [46.43]

RA: |No| (0.77). |I don't|. [2.24]

AM: |Have you ever drive? | [2.17]

RA: |ever |. [2.30]

AM: |Never? :: Even in desert? | [1.36]

RA: |Just my sister. | [3.0]

AM: |Oh, your sister, OK, (0.28) OK [/]:: I have drive on the road.| [4.44]

RA: |even my Mum…| [0.59]

AM: |OK..OK [/]:: your Mum have drive, {I think} I (0.32) {I think} [/] I thought [///]:: Mum was the one :: who was scared (0.46) | |Not [/-], I thought (0.28) you were the one :: who is very ok|,|awkward (0.75)|. |Do you think there is advantages (0.34)? | |Give me the advantages and disadvantages in driving a car (0.96)? | |anything from any side|. [17.84]

RA: |Some women **(0.48)** don't have a man **(1.0):: in (0.54) in [/] he..**? | [7.30]

AM: |In their family? | [0.78]

RA: |Yeah.| [0.34]

AM: |Yeah.| |So:: do you think:: this is advantage:: to drive?| [2.83]

RA: |Yeah|. [0.66]

AM: |And do you think ::there is disadvantages? | [2.59]

RA: |No **(0.50)** |, |I don't think **(0.64)** :: it has **(0.56)** uhh disadvantages? | [6.11]

AM: |I think:: OK, yeah, me too| |I think :: there is no disadvantage of driving for women here| |but I think:: the only thing that I…um (0.75) (1.21) I didn't drive now:: because (0.44) uhh (0.35) the guys in Saudi Arabia they didn't get used:: to (0.45) women driving on the road|, |so when they saw a women:: and they cry out… ::'Oh my God, there's a woman in the car, follow her'|. |I see that:: I swear to God I see that [/-].| |Like, uh (0.47), my Dad has a friend::, he is he is [/] (0.36) really old that man:: so he went he came [///] to my Dad| |and he said,:: they

sitting in the same place in the coffee shop (0.49) |, |so he went :: to my Dad,:: he said (0.80):: 'I swear to God I see a woman that drive, :: I followed her all the way to her house'. | [41.56]

RA: …|my brothers,:: he say that.| [3.05]

AM: |so I think:: the men here, they're not ready ::to see a woman drive (0.44).| | uhh (0.37) I'll wait maybe (1.0) five years (0.65):: until I drive (1.28) | |two years (0.28)|. |I'll get my own uhh (0.40) chauffeur:: to drive me ::to where I work later (0.30) enshallah. (0.94) | |uhh (0.57) until I see women driving :: and then I will drive (0.90) enshallah| |I hope:: my Dad will agree with that (0.68) |. |Yeah| |that's it|. [27.88]

AM: |uhhh (0.72) (0.87) do you think| |OK| |the price of gas (0.67) some people say:: that women (1.41) they will think:: that the price of the gas is really high|, |and they will not able ::to buy a car, ::and the price of the gas is real high for them (1.04) |, |so what do you think:: (1.05) what do you think [/] about that?| [19.39]

RA: |The price of uhhh (0.55) gas,:: it's (0.51) not high.| [4.53]

AM: |in Saudi Arabia! | [0.67]

RA: |Yeah.| [0.61]

AM: |What about the car?|. [0.84]

RA: |even car::but the uhh (0.33) price| [5.23]

AM: |yeah| |it is true| .. |yeah| |The uhh (0.38) (0.46) [/-] when my uhhh (0.38) dad (0.36) (no his friend) (0.39) [&], they they [/] went to Jordan (0.26):: I think:: his family is actually from Jordan (0.63) |, |so he took my Dad car (0.43) my dad'car

[///]|, |so he said:: 'I will fill it up (0.33) full for gas'(0.55) |, |and he take a bout huge bottle bottle [/]of gas:: like he full it (0.32) with gas (0.63):: they keep filling with gas|, |so he went there:: with that gas|, |and every time the car is out of gas:: he pull up again with the same bottle|, |because their gas like (0.45) [/-]…uhh (1.05) my Dad's car here is the price :: to make a full take 62 (0.66) to 100 Riyal (0.33):: but there 600| |so it is really high.| [52.02]

AM: |I think:: there is no woman will buy a car ::unless she is employed as well.| |I'm OK with that|

RA: |Oh my God. | |Yeah **(0.38)** |, |I think that as well. | [4.96]

AM: |You would buy your own car someday (1.18),:: or will you get one :: if someone will buy

you a car? | [8.98]

RA: |Yes|, |I will|. [1.20]

AM: |What kind of car are you gonna buy? (1.26) :: Seriously be honest|. [3.83]

RA: |Maybe **(0.34)** Mercedes.| [2.87]

AM: |What kind of job hahahaha…?| |What kind of job:: that will get you Mercedes? | [5.0]

RA: |you don't know|. [0.79]

AM: |To me:: I might (0.57) at first,:: I'm going to be very broke (1.05) | |when I I [/] just get employed (0.31):: so I think:: I'm going :: to buy Prius (1.34)| |and then I'm going :: to be richer, just a little bit (0.66):: and I'm going :: to buy a Range Rover.| [18.82]

Student	Speaking time	fluency measures	N	duration	student	ST	fluency measure	N	duration
Am.	**281.45**	**MidSPS**	**31**	**17.69**	**Ra**	**48.59**	**MidSPS**	**8**	**3.88**
syllables	**925**	**EndSPS**	**19**	**23.54**	**syllables**	**115**	**EndSPS**	**5**	**3.29**
Refor.	**4**	**Filled P**	**20**	**7.69**	**Refor.**	**0**	**Filled P**	**3**	**1.33**
Fasle S	**4**	**Silent P**	**50**	**41.23**	**Fasle S**	**0**	**Silent P**	**13**	**7.17**
Repetitions	**14**	**MidFPS**	**20**	**7.69**	**Repetitions**	**1**	**MidFPS**	**3**	**1.33**
Replace.	**1**	**EndFPS**	**0**	**0**	**Replace.**	**0**	**EndFPS**	**0**	**0**

N- Sample 7 of dialogue coded data

LA: |Uhh (0.61) give me your opinion and reasons :: uhhh (0.97) and if you agree or disagree:: with allowing women:: uhhh (0.61) to drive in Saudi Arabia.| [9.06]

AM: |uhhhh (0.48) I agree:: uhhh (0.57) because ummm (0.60) **(0.51)** **{**I want to take}…I want to take [/] uhhhh (0.29) **{**lesson} license [&] :: because uhhh (0.32) **(0.92)** I want to bring my needs without waiting for someone uhhh or having to uhhhh (0.57) ride with a uhhh (0.57) stranger taxi driver **(1.10)** | |What about you?| [23.14]

LA: |Uhhh (0.55) I actually agree:: , but I won't drive for the moment:: because ummm (0.68) I don't own a license| |and most importantly I don't even own a car (0.53), :: to begin with (0.63).| |Ummm (1.52) (0.82), and I don't even know:: how (0.69) to drive a car|, |I don't even (0.81) I mean (1.02), I don't know:: what to do if I sit in front of the uhh (0.38) (8.14) …yeah, I just don't|. |Ummmm (1.07) (0.69) so (1.22) are you planning on getting your license anytime soon? | [38.41]

AM: |Uhhh (0.38) No uhhhhh (0.57) **(0.47)|**, |uhhhh (0.51) now I don't have uhhh (0.64) a driver's lesson.. license [&] ::but in the future uhhhh (0.38) I want uhhh **(0.31) ca ..**catch [/] it uh:: I want take it, uhhh (0.72) InshaAllah|. [14.92]

LA: |As for me:: I want to take uhhhh (0.73) a license:: but I don't know when or (0.39). Uh (0.87) how| |because (0.53) I think:: I should get a car first,:: then I'll get a license (0.79) so I could ummm (0.57) (0.71) drive… drive [/] it with no worries (0.56).| | Uhh (0.61) Have any of your relatives or family or friends uhhh (0.81) drove a car before?| [22.01]

AM: | Umm (0.59) no||umm (0.52) I think no **(0.49)**:: but Inshallah in the future ummm (0.48) **(0.59)::** we will uhhh (0.31) try to drive.| **[8.46]**

LA: |Yes|, |same as me (0.30) |, |none of them are driving now,:: but they're willing and planning to| |but they want to get their license first (0.78), uhhh (0.56) and cars too (0.78)|. |I think (0.84):: they won't be taking their husbands or (0.63) their fathers cars, :: because they also need it| , |so yeah| |they need their own car (0.43) | |so yeah| |but in the future this thing is gonna happen.| [23.16]

AM: |Uhh (0.29) do you think uhh (0.58) :: dri…. women driving [///] has uhh (0.50) advantages (0.27) or disadvantages? | [5.77]

LA: |Yes|, |of course|, |like anything else (0.48)|. | As for the advantages (0.40):: you can go and get whatever you want at any time you want (0.35) |. |If you

wanted this thing, right now:: you can just (0.90) go {into car} … into your car [/] and go and get it|, |you don't even need:: to wait for someone for a specific time or days for that matter (0.72)|. |As for the dis ..disadvantages [/] yeah (0.62) the gas:: is (0.91) becoming high in price| |and uhhh (0.62) (1.58) uhh (0.84) the driving school (0.37) tuition few is (0.82) fee [///] I mean :: is uhh (0.57) a bit high at the moment|, |so I think:: we need to wait a bit before we can do anything (0.55) |, |but I think:: in a matter of five years:: there are going to be a lot of women driving uhhhh (0.72) around on the streets (0.33) |. |What about you?| [49.97]

AM: |For me|, | I uhhhh (0.38) think:: not advantages (0.98) but umm (0.64) (0.50) {women…women} [/] [/] women driving {has} [/-] (0.60) umm (0.43) women driving (0.55) umm (0.53) I think:: it is (0.33) necessary things (0.32)| |because ummm (0.45) (0.59) to make uhhh (0.55) life easy (0.31) easier [///] :: to women.| [19.60]

LA: |But not always it's necessary|, |I mean, some of them (0.59) she just wants:: to drive for fun (1.24) |, |you know|. [6.92]

AM: |No, but some uhhh (0.35) women uhhhh (0.56) (0.60) there is no man in house (1.09)|. … |he needs | [6.38]

LA: …|yeah|, |those those [/] (0.72) driving for them is necessary :: because she wouldn't uhh (0.81) need a man to do her needs, and a strange a man for {the ma} for the fact [&](0.94).| |Uhhhh (0.49) but I talk .. I'm talking [///] about those girls :: who just want to have fun (0.25) :: and they actually (0.31) make it look bad in the eyes of the fathers and brothers. | [21.63]

AM: | yeah…maybe **(0.73)** |, |but some women uhhh (0.35) want **(0.67)** uhh needs [&]:: to do this **(0.65)**, uhhhh (0.42) :: because bring her needs without **(0.49)** uhhh (0.47) waiting for someone :: to **(0.54)** uhhh (0.36) give her lift. [11.61]

LA: |She wouldn't even need:: to pay him an extra money for the things she want.| [4.43]

AM: Yeah.

Student	Speaking time	fluency measures	N	duration	Student	Speaking Time	fluency measure	N	duration
LA	**123.5**	**MidSPS**	**21**	**22.83**	**AM**	**75.83**	**MidSPS**	**16**	**8.82**
syllables	**560**	**EndSPS**	**16**	**9.33**	**syllables**	**232**	**EndSPS**	**7**	**4.79**
Refor.	**2**	**Filled P**	**21**	**15.21**	**Refor.**	**2**	**Filled P**	**33**	**15.36**
Fasle S	**0**	**Silent P**	**37**	**32.16**	**Fasle S**	**1**	**Silent P**	**23**	**13.61**
Repetitions	**3**	**MidFPS**	**21**	**15.21**	**Repetitions**	**5**	**MidFPS**	**31**	**14.36**
Replace.	**0**	**EndFPS**	**0**	**0**	**Replace.**	**3**	**EndFPS**	**2**	**1**

Appendix 13 Test of Normality Kolmogorov-Smirnov and Shapiro-Wilk

Variables	Kolmogorov-Smirnov		Shapiro-Wilk	
	Statistic	Sig.	Statistic	Sig.
Monologue speed				
Length of sample	.124	.016	.951	.013
Syllables/60	.166	.000	.895	.000
Speech rate/60	.103	.087	.969	.112
Articulation rates	.092	.200*	.973	.179
PTR	.131	.008	.898	.000
MLS	.181	.000	.851	.000
Dialogue speed				
Length of sample	.147	.002	.900	.000
syllables	.169	.000	.830	.000
Speech rates	.080	.200*	.982	.495
Articulation	.099	.190	.974	.191
PTR	.118	.028	.905	.000
MLS	.186	.000	.777	.000
Monologue breakdown				
SPs/60	.111	.047	.847	.000
Midcsps/60	.124	.016	.786	.000
Endcsps/60	.065	.200*	.988	.787
M.Leng. Mid_csps	.238	.000	.557	.000
M.Leng End_csp	.283	.000	.571	.000
FP/60	.157	.000	.747	.000
Mid-FPs/60	.129	.010	.825	.000
M.Length Mid_cfp	.413	.000	.239	.000
Dialogue breakdown				
SPs/60	.064	.200*	.982	.484

Midcsps/60	.075	.200*	.944	.006
Endcsps_60	.188	.000	.871	.000
M_Leng_Mid_csps	.255	.000	.543	.000
M_Leng-End_csps	.251	.000	.554	.000
FP/60	.088	.200*	.976	.236
Mid-FPs/60	.069	.200*	.983	.539
M.Leng-Mid_cfp	.482	.000	.143	.000
Monologue Repair				
Total Repairs/60	.102	.095	.959	.031
False Starts/60	.318	.000	.734	.000
Reformations/60	.258	.000	.828	.000
Repetitions/60	.080	.200*	.951	.013
Replacements/60	.303	.000	.793	.000
Dialogue repair				
Total Repairs/60	.135	.006	.956	.024
False Starts/60	.270	.000	.815	.000
Reformations/60	.136	.005	.902	.000
Repetitions/60	.138	.004	.926	.001
Replacements/60	.337	.000	.665	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction