Self-monitoring behaviour of L2 learners: proficiency level, dual task paradigm and working memory capacity

Ghadah Ahmad Albarqi

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

School of Literature and Languages

Department of English Language and Applied Linguistics

University of Reading

September 2019

Declaration

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

Ghadah Ahmad Albarqi

19th September 2019

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

A-repair	Appropriateness-repair
ANOVA	Analysis of Variance
BDST	Backward Digit Span Test
CEFR	Common European Framework of Reference
D-repair	Different-information-repair
E-repair	Error-repair
EFL	English as a Foreign Language
EIT	Elicited Imitation Test
H	Hypothesis
L1	First Language
L2	Second Language
LST	Listening Span Test
MANOVA	Multivariate Analysis of Variation
PL	Proficiency Level
PLT	Perceptual Loop Theory
PPP (3Ps)	Presentation, Practice, and Production
PSTM	Phonological Short-Term Memory
RQ	Research question
SLA	Second Language Acquisition

SPSS	Statistical Package for the Social Sciences
TBLT	Task-Based Language Teaching
TC	Task Condition
WMC	Working Memory Capacity
WRT	Word Recall Test

Abstract

Speaking a second language involves complicated processes that draw on cognitive resources to function efficiently. One of the important speech processes is self-monitoring which entails checking internal and external speech against the existing linguistic system. Although L1 research has provided a detailed account of L1 self-monitoring behaviour, there are few studies which actually investigate L2 self-monitoring behaviour. Also, there have been few studies examining the role of individual learner differences such as proficiency and working memory on L2 self-monitoring behaviour. The present study aims to address these gaps in the literature by investigating the extent to which L2 self-monitoring behaviour can be influenced by proficiency level (PL) and task condition (TC). The relationship between working memory capacity (WMC) and self-monitoring is also examined. PL (Elementary and Intermediate groups) in the current study is assessed using a C-test and an Elicited Imitation Test. TC entails performing either in single or dual task conditions. The single task condition involves describing oral narrative picture prompts whereas the dual task condition entails a simultaneous performance of the oral narrative prompts and a secondary task. WMC was assessed using Back Digit Span Test in L2 English.

The participants were sixty-six Saudi L2 learners of English from a state University in Saudi Arabia. The participants' oral performances were coded and analysed in terms of disfluency, repair types, temporal phases of repair, and accuracy. The study employed a 2x2 between-subject factorial design with PL and TC as independent variables, and WMC as a continuous variable. To examine the effects and interaction effects of PL and TC on the dependent variables, MANOVA was run. The data showed that PL had a significant effect on the combined measure of dependent variables whereas TC did not. The results also showed a significant interaction effect between the two variables on the combined dependent variable. Two-way between-subject design ANOVAs were run to locate the significant differences.

The data analysis indicated that almost all aspects of self-monitoring were significantly affected by the development of proficiency. This suggests that proficiency development is likely to have led to changes in L2 self-monitoring behaviour in terms of disfluency, repair type, temporal phases of repair, and accuracy. On the other hand, task condition only slowed down two temporal phases of A-repair which suggests that L2 self-monitoring aspects might be robust to the effects of the increased demand of task condition. The interaction analyses indicated that there were differences between the two proficiency groups when they

performed in the single task condition, but these differences disappeared when they performed in the dual task condition. The follow-up analyses demonstrated that some aspects of self-monitoring considerably increased in the dual task condition which suggests that the increased demand of the dual task condition pushed the learners out of their comfort zone in using their L2 ability, with a particular impact on their self-monitoring behaviour.

Correlation analyses were used to examine the relationship between WMC and self-monitoring aspects in the two task conditions. The analysis indicated that in the single task condition there were significant negative correlations between WMC and repetition on the one hand, and WMC and lexical-repair, on the other, but these correlations were weak. However, there was no correlation between the WMC and self-monitoring aspects in the dual task condition. The data suggest that during the single task condition, available working memory resources might enhance pre-articulatory self-monitoring, resulting in slightly less repetition and lexical-repair in L2 overt speech. The data in the current study has important implications for the Perceptual Loop Theory, as some premises of this theory were supported in L2 context.

Acknowledgements

First, I would like to express the deepest appreciation and thanks to my first supervisor, Dr. Parvaneh Tavakoli, who introduced me to the interesting field of Applied Linguistics and whom I consider as my role model in academia. Her patience, wisdom, encouragement and careful comments made a great difference in my PhD journey, and laid the foundations for my work as a researcher. Without her guidance and help, this thesis would not have been possible. I would like to extend my thanks to Dr. Erhan Aslan, my second supervisor, for his careful comments on my work, and for his insightful questions that have extended my thinking. In addition, thanks go to Dr. Jacqueline Laws, the previous Director of Postgraduate Research Studies, for her continuous encouragement and reassurance throughout my study in DELAL.

I am extremely thankful to Taif University for granting me the chance to pursue my studies and funding all my expenses during the PhD journey. I am sincerely grateful to the students and staff at Taif University in Saudi Arabia for their help and contribution to data collection. I would like also to thank the staff of DELAL and University of Reading for making my study enjoyable and constructive. Many thanks to all my friends for helping me survive the stress throughout my journey. I am also thankful to members of L-SLARF (London Second Language Acquisition Research Forum) for their insightful comments on my work during the annual PhD conferences which were of great help in shaping my thinking.

Most importantly, my sincerest gratitude goes to my parents whose love and prayers were of great inspiration and motivation. I am also grateful to my sisters and brothers with whom I shared my ups and downs. I want to especially thank my loving husband, Abdullah, for believing in me and sacrificing a great deal to make my dreams come true. Great thanks go to my adorable two sons (Azooz and Soso) who are a source of great inspiration and joy in my life. They have taught me the boundlessness of myself. Thanks to all of you!

Chapter 1: INTRODUCTION

1.1 Motivation of the study

As an English language teacher teaching university students at Taif University (a state university in Saudi Arabia), I was always fascinated by the extent to which learners at different stages of language learning engaged in the process of monitoring their speech, making attempts at repairing their utterances, and feeling concerned about the forms in which they put their intended meanings. I could easily see that when my students aimed at producing a complex sentence, this was usually associated with hesitations, self-corrections, and reformulations. One key question shaping in my mind was why L2 learners at different levels of proficiency feel obsessed with correcting their errors during speech production. English in Saudi Arabia is taught in a foreign language context, which means that students have a very limited opportunity to practice English outside the classroom. Lack of real communication in English probably made L2 learners reluctant to speak English in the classroom, and they become concerned about checking and correcting their linguistic errors in an attempt to produce grammatically accurate utterances.

The prevailing teaching method of English in Saudi classrooms is Presentation, Practice, and Production (PPP). This method consists of three stages. The first stage is the presentation of an aspect of the target language in a context that students are familiar with. The second stage involves practice, where students are provided with activities to help them practise and familiarize themselves with the new aspects of the target language. The third stage is production, where the students produce the language in activities in the syllabus. This approach mainly focuses on introducing and highlighting new structures and vocabulary and providing good practice before production. L2 learners endeavour to practise and produce the correct grammatical forms and accurate pronunciation of new words. This approach is easy to plan and implement especially in large classes. Perhaps more importantly, it provides a clear structure to the lesson: a good beginning, a clear middle and a productive end. Despite all these advantages, as a teacher using this method, I had noted that it was not as successful as one could expect. L2 learners, particularly those at higher proficiency levels, got bored during the practice stage of the lesson. This approach has also been criticised for denying L2 learners meaningful communication (Willis, 1996), as learners mainly attempt 'to demonstrate to the teacher their control of the target form' (Willis, 1996, p. 4). The approach focuses on the accurate production of the target structure and vocabulary. This could partly explain why L2 learners in EFL classrooms become obsessed with correcting their errors.

The attention towards accuracy has been reinforced by the testing policy in the country which values accurate production of the target language. In such a context, L2 learners are expected to be concerned with correcting their errors. In other words, they may monitor their speech for linguistic errors. Taken together, it seems that teaching and assessment practices could turn learners' attention towards a certain aspect of their speech. Although some researchers have argued that L2 learners often focus their attention on meaning rather than linguistic forms when monitoring their speech (Lennon, 1984; Van Hest, 1996), this argument might not hold for all L2 learners as Kormos (2006) argues that:

> Formally instructed foreign language speakers in countries where explicit grammar teaching plays a significant role in the curriculum, everyday teaching practice, and state-level language testing, can allocate their attentional resources and make decisions concerning error corrections in a different way from learners in a second language environment or from students instructed with communicative methods. (Kormos, 2006, p. 132)

This argument suggests that in classrooms where the emphasis is directed towards linguistic forms, it is expected that L2 learners will pay attention to accuracy of grammatical, lexical, and phonological items rather than to the communicative content of their utterances. The issues of self-correction and monitoring are complicated since L2 learners may have different attitudes towards self-correction. For example, L2 learners may notice their errors but choose not to correct them (Mackay 1992). The debate about self-correction triggered my curiosity about this phenomenon.

My interest in self-monitoring was also motivated by my work on a twenty-thousand-word dissertation, which was an essential requirement in a taught-track PhD programme at the University of Reading, triggered a greater motivation for further research into this area. I was examining the effects of +/- immediacy of information on L2 self-repair and disfluency (Albarqi, 2016). This study examined the extent to which disfluency and self-repair in L2 speech can be influenced by the presence and the absence of the picture prompts. While I was reviewing the literature on self-monitoring, I observed some gaps between L1 and L2 research in this area. An extensive bulk of L1 literature have studied different areas of L1 self-monitoring such as functioning of self-monitoring processes (Levelt, 1983, 1989, 1999; Seyfeddinipur, Kits, & Indefrey, 2008); different aspects of monitoring (Blackmer & Mitton, 1991; Brédart, 1991; Broos, Duyck, & Hartsuiker, 2016; Levelt, 1983, 1989, 1999; Oomen & Postma, 2001); and the impact of the increased cognitive demand on self-monitoring behaviour (Jou & Harris, 1992; Oomen & Postma, 2001, 2002). Furthermore, L1 research has particularly been interested in testing the assumptions of the existing theories of L1 self-monitoring (e.g., Dell, 1986; Laver, 1980; Levelt, 1983). The empirical examination of different assumptions of monitoring theories in L1 literature helped to validate different premises of these theories (e.g., the Perceptual Loop Theory, see Section 2.3.3).

However, although extensive research has examined L2 self-monitoring in relation to language impairment such as stuttering or aphasia (e.g., Eichorn, Marton, Schwartz, Melara, & Pirutinsky, 2016), there is still a need for research on L2 self-monitoring behaviour in EFL and ESL (English as a Second Language) classrooms. In this regard, it has been argued that L1 speech production model can be applied to the L2 context without major changes (Kormos, 2006). This suggests that the Perceptual Loop Theory (PLT), which was originally developed to explain self-monitoring behaviour of L1 speakers, can be used to explain L2 self-monitoring behaviour as well. One of the main premises of this theory is that self-monitoring is largely drawing on cognitive resources during its functioning (Levelt, 1983, 1989, 1999; Levelt, Roelofs, & Meyer, 1999). However, 'the fact that the L2 speakers' system of knowledge is typically incomplete and L2 speakers' production mechanisms are not fully automatic' (Kormos, 1998, p. 334), is expected to result in variations in L2 self-monitoring behaviour. Furthermore, the fact that the L2 linguistic system is in a constant state of development suggests that L2 self-monitoring behaviour would vary according to the different levels of proficiency. As such, L2 self-monitoring is defined as checking inner and overt speech against the existing linguistic system (Kormos, 2006). Thus, it is important for L2 research to study the changes of L2 self-monitoring behaviour at different levels of proficiency. The rarity of research in this area constitutes a motivation for the current study.

1.2 Rationale of the study

Speaking a second language involves making errors, self-correction and hesitations among other speech features. These speech features are used by researchers as a window to provide a view into the underlying speech processes. Among these processes is self-monitoring which is a very interesting phenomenon because it relates to different areas of a second language such as language processing, production, acquisition and development. When we speak a second language, we tend to revise the utterances in our minds before articulation, which is considered one of the processes of self-monitoring (Levelt, 1983). Speaking a second language involves repeated occurrences of self-correction and hesitations due to the limited knowledge of the second language (Kormos, 2006). Researchers contend that self-monitoring can facilitate L2 learning and development in several ways (DeKeyser, 2007, 2010; Kormos, 1999; Swain, 1995; Swain & Lapkin, 1995). For example, repeated monitoring can enhance fluency and automaticity of L2 speech production (DeKeyser, 2007). It can also direct learners' attention to gaps in their knowledge of the target language which can trigger further learning processes (Izumi, 2003; Kormos, 1999, 2006; Swain, 1985). Furthermore, some aspects of self-monitoring such as self-repair can be regarded as a useful predictor of language development (DeKeyser, 2010; Kormos, 2006; Tavakoli, Nakatsuhara, & Hunter, 2017).

An important premise of the PLT states that self-monitoring is sensitive to contextual effects (Levelt, 1983). This principle follows from the assumption that self-monitoring draws on cognitive resources, as discussed above. As L2 processing largely consumes cognitive resources (Kormos, 2006), self-monitoring behaviour is expected to be affected by the influences of different contexts. Some L2 studies have investigated the extent to which task characteristics or task conditions can affect L2 self-monitoring behaviour (e.g., Ahmadian, Abdolrezapour, & Ketabi, 2012; Ahmadian & Tavakoli, 2014). It has been shown that types of self-repair were affected by task structure and task condition. However, L2 studies have rarely employed a dual task condition in examining the effects of resource limitation on L2 self-monitoring behaviour. Resource limitation entails depleting available cognitive resources. The dual task paradigm requires participants to perform two tasks simultaneously which is expected to consume available cognitive resources. This technique is preferred in studying the impact of resource limitations on self-monitoring in the field of cognitive psychology (e.g., Jou & Harris, 1992; Oomen & Postma, 2002). However, previous studies (both in L1 and L2 contexts) which employed dual task paradigm in studying self-monitoring lack systematic operationalisation of the concurrent tasks. That is, the degree of similarity between the two concurrent tasks in previous studies has not been based on theoretical foundation. This is regarded as a shortcoming in the operationalisation of the dual task condition. As such, there is a need to systematically operationalise the dual task condition.

One of the rationales of the current study is to address the methodological shortcomings found in previous research. It has been recommended that operationalisation of variables under investigation should be theoretically driven, and that shortcomings of previous studies need to be taken into consideration (Awwad, 2017). The operationalisation of the dual task condition is one of these shortcomings as discussed above. Wickens (2007) has offered insights into the operationalisation of the dual task condition by proposing several dimensions for assessing the degree of similarity between concurrent tasks. The degree of similarity between the concurrent tasks can determine the extent to which the two tasks draw on the same resource pool (Wickens, 2007). Once the two tasks draw on the same resource pool, performance is expected to deteriorate because available cognitive resources would largely be consumed (Wickens, 2007). Hence, dual task condition needs to be systematically operationalised according to this theoretical principle. Moreover, some studies on self-monitoring have used network description task which entails describing the directions of coloured dots and lines (e.g., Declerck & Kormos, 2012; Levelt, 1983; Oomen & Postma, 2002). This task has been criticised because it requires participants to correct every single error, mainly lexical errors, in their speech (Van Hest, 1996). That is, the task mainly directs learners' attention towards one aspect of self-monitoring which is correcting lexical errors. The influence of such a task might not resemble real life conditions in which a speaker may have to correct different types of errors. As such, the current study intends to address the methodological drawbacks in self-monitoring research, namely, the dual task condition and the task type.

Another rationale of the current study is concerned with the extent to which L2 self-monitoring behaviour is dependent on working memory capacity. According to the Perceptual Loop Theory (the PLT), monitoring is a conscious process which is influenced by the processing constraints of working memory (Levelt, 1983). Due to the limited capacity of working memory, self-monitoring can deal with few items of internal or external speech at a time (Levelt, 1983). In L2 speech processes, cognitive resources are already exhausted through several processes (grammatical, lexical, phonological) as these processes are not yet automatized, especially in lower proficiency learners (Kormos, 2006). This means that L2 speakers are likely to prioritize certain aspects of their speech during the process of monitoring. Although it has been advocated that L2 self-monitoring studies need to investigate the association between monitoring aspects and working memory (Kormos, 2006), only few studies have attempted to do so (e.g., Georgiadou & Roehr-Brackin, 2016;

Mojavezi & Ahmadian, 2014). These studies have, nevertheless, focused on few aspects of monitoring behaviour (i.e. self-repair and pauses) and the findings of these two studies are inconsistent. As such, there is still a need for further studies to look at other aspects of monitoring such as temporal phases of repair, accuracy, and disfluency which might provide important information about the association between working memory and self-monitoring behaviour.

Finally, it has been noted that L2 research mainly focuses on self-repair as a proxy of self-monitoring behaviour (e.g., Georgiadou, 2016; Gilabert, 2007; Kovač & Milatović, 2012). Furthermore, only few studies have employed disfluency in their investigation of self-monitoring (e.g., Ahmadian, Abdolrezapour, & Ketabi, 2012; Ahmadian & Tavakoli, 2014). According to the PLT, disfluency is useful in examining self-monitoring since disfluency features are produced as corrective actions to expected errors (Levelt, 1983). That is, speakers' disfluency is often aimed at buying time while checking the correctness or appropriateness of their speech. Other aspects of monitoring that have been rarely examined in L2 research are the temporal phases of repair. According to the PLT, making repair involves three phases: the first phase includes the erroneous or inappropriate utterance, the second phase is characterised by making hesitation, and the third phase entails correcting the error or inappropriateness. Calculating the duration of these phases of repair can provide insights into the functioning of self-monitoring (Kormos, 2006). The only studies that examined temporal phases of L2 repair were conducted by Van Hest (1996) and Declerck and Kormos (2012), but these studies were implemented on a small scale. Also, the ratio of error correction, which is considered a reliable measure in examining self-monitoring (Oomen & Postma, 2001), has not been employed in many L2 studies. Assessing accuracy is useful in studying self-monitoring since self-mentoring behaviour is regarded as an attempt of L2 learners to promote accuracy in their speech (Gilabert, 2007; Kormos, 1999). Therefore, the current study intends to study different aspects of L2 self-monitoring behaviour, namely, disfluency, self-repair types, temporal phases of repair, and accuracy.

To address the gaps discussed above, the current study examines L2 self-monitoring behaviour according to three variables: proficiency level, task condition, and working memory capacity. The rationale for including proficiency level is to examine the extent to which L2 self-monitoring behaviour changes with the development of the linguistic system. Concerning task condition, the study is designed along dual task paradigm which means that participants will perform two tasks simultaneously in the dual task condition, and only one task in the single task condition. Furthermore, the interaction effects between individual differences in proficiency level and task condition is examined because interaction analysis is regarded as an adequate tool in exploring ‘the black box’ of language processes (DeKeyser, 2012, p. 190). This study also examines the relationship between working memory capacity and different aspects of L2 self-monitoring such as disfluency, repair types, temporal phases of repair, and accuracy.

The current study offers original contributions to the field of L2 self-monitoring in terms of theory, research, and practice. This study endeavours to test some assumption of the PLT into the L2 context. As the PLT assumes that self-monitoring is dependent on cognitive resources, the current study aims to investigate the extent to which resource limitation can affect different aspects of self-monitoring behaviour. Furthermore, one of the purposes of the current study is to examine the extent to which self-monitoring behaviour can be influenced by proficiency development. Investigating the influence of proficiency level on L2 self-monitoring behaviour could reveal useful information about the relationship between self-monitoring and other speech processes. In other words, self-monitoring as a controlled process depends on cognitive resources (Levelt, 1983), yet the automatization of other L2 speech processes is expected to have an impact on the functioning of L2 self-monitoring. Also, according to the PLT, self-monitoring draws on the limited-capacity cognitive system which is working memory. Hence, the examination of the relationship between WMC and self-monitoring aspects, is expected to enhance our understanding of the L2 self-monitoring functioning. As a final product, this study is expected to provide a reference for future L2 research on L2 self-monitoring behaviour through testing some assumptions of the PLT into L2 context, besides introducing novel methods and practices in examining L2 self-monitoring behaviour.

1.3 Originality of research design

This section highlights the originality of the research design and the novel practices used in the current investigation of L2 self-monitoring. The present study has provided an original operationalisation of the dual task condition which was based on Wickens’ (2007) proposal of the limited-capacity multiple resources model. The two concurrent tasks are designed to consume the available cognitive resources as they potentially draw on the same resource pool as described in Wickens’ (2007) model (see Section 4.5.2). These resource pools are:

perceptual modality (visual or auditory); coding format (spatial or verbal); and processing stage (the data involve perception, cognition, and responding) (Wickens, 2007) (see Section 4.5.2). The degree of similarity between the concurrent tasks moderates the extent to which the tasks draw on the same resource pools (Wickens, 2007). When the two tasks draw on the same resource pool, it is likely that cognitive resources are largely consumed. Moreover, the secondary task of the dual task condition is designed specifically for the purpose of the current study using E-Prime Psychological Software (3.0). It involves showing continuous words on a computer screen and requires participants to decide whether these words relate to animate or inanimate objects by pressing a certain button in each case (see Section 4.5.2). Another original aspect of the current study is the novel design of two comparable narrative picture prompts on the basis of the guidelines proposed by De Jong and Vercellotti's (2016). The two picture prompts are made similar in terms of task structure, story line complexity, vocabulary, and the number of elements. The oral narrative picture prompts are used as a primary task in the dual task condition. The two tasks (primary and secondary) are comparable along the three dimensions described in Wickens' (2007) model. The two tasks are visual, verbal, and entail perception, cognition and responding. These novel aspects have been introduced to address the methodological limitations found in previous research on L2 self-monitoring (for further details, see Section 2.11.2).

1.4 Definitions of the key terms

The aim this section is to introduce definitions of the key terms used in the current study. Further discussion about these terms will be presented in different chapters of the thesis. These terms are self-monitoring, the PLT, self-repair, disfluency, temporal phases of repair, accuracy, and dual task paradigm.

- **Self-monitoring** is the construct that the current study aims to investigate. Kormos (2006) defined L2 self-monitoring as checking the inner and overt speech against the existing linguistic system. This suggests that L2 self-monitoring behaviour would change with the development of the linguistic system.
- **The Perceptual Loop Theory (The PLT)** is a theory of self-monitoring which is based on the monolingual speech production model developed by Levelt and colleagues (e.g., Levelt, 1983, 1989, 1999; Levelt, Roelofs, & Meyer, 1999). It has been assumed that this model can be applied to L2 context without major changes (Kormos, 1998), yet the

assumption that L2 linguistic system in a constant state of change makes it necessary to test the assumptions of the PLT in L2 learners at different levels of proficiency.

- **Self-repair** is an aspect of L2 self-monitoring which entails retracing an utterance to make a correction or adjustment to what has been said before. The current study follows self-repair classification developed by Levelt (1983) and Kormos (1998).
- **Disfluency** is an aspect of self-monitoring which includes filled, silent pauses, and repetition. Disfluency has a special status in the PLT as disfluency features are supposed to be produced as corrective measures to anticipated errors. That is, these features are used as strategies to buy time during monitoring utterances before and after production.
- **Temporal phases of repair** are aspects of self-monitoring which involve three stages of repair. The first phase includes making an error, or inappropriateness; the second phase entails hesitation; and the third phase involves making a correction or an adjustment to what has been produced.
- **Accuracy** is employed as an aspect of self-monitoring since promoting accuracy in L2 speech is considered the goal of self-monitoring (Gilabert, 2007; Kormos, 1999). Accuracy measures in the present study include the ratio of error correction (repair divided by total number of errors), and the percentage of error-free clauses (error-free clauses divided by total number of clauses multiplied by 100).
- **Dual task paradigm** is a model used in experimental psychology to examine processing and cognitive resources. The use of the dual task paradigm involves performing two simultaneous tasks: single and dual task conditions. Single task condition entails performing only one task whereas the dual task condition requires participants to perform two tasks simultaneously. It is used in self-monitoring research to test the assumption that cognitive resources are limited (e.g., Jou & Harris, 1992; Oomen & Postma, 2002).

1.5 Organization of the thesis

This section will present an outline of the content and structure of the thesis. This thesis consists of seven chapters: the introduction, literature review, pilot study, methodology, results, discussion, and conclusion. The first chapter has four sections. The first section has already introduced the background and inspiration of the current study. The second section has presented a description of the importance of self-monitoring in L2 learning and development. It has also identified the gaps in the body of L2 research in terms of theory and methodology. It then introduced an overview of how to address such gaps, and an outline of how the investigation is carried out. It concluded with a short account of the anticipated contribution of the current study to L2 self-monitoring theory and research. The final section introduced definitions of the key terms in the present study.

The second chapter reviews and evaluates previous literature on self-monitoring. The review includes existing theories of L1 and L2 speech production and models of self-monitoring. As self-monitoring has been studied extensively in L1 literature, L1 research is reviewed to highlight methods of examining self-monitoring behaviour. This is followed by looking at a range of L2 studies that examined L2 self-monitoring for the purpose of tracking development and identifying gaps in the field. This chapter also highlights how research in the area of automaticity can help us understand self-monitoring behaviour in second language. Finally, studies relating to the three variables of the current study, that is, proficiency level, task condition and working memory are reviewed and evaluated in terms of both theoretical and methodological factors.

The third chapter presents the details and results of the pilot study which is mainly conducted to test the design and operationalisation of the picture prompts and dual task condition. The chapter starts by presenting the aims of the pilot study. This is followed by presenting a detailed description of the procedures of designing the picture prompts and dual task condition. Further explanation is provided about the recruitment of participants, the data collection and analysis. The findings of the pilot are discussed in relation to the theoretical principles and previous studies. This chapter concludes by discussing the limitations of the pilot study that need to be considered in the main study.

The purpose of the fourth chapter is to present the objectives and the experimental materials of the current study. In Section 4.3 the research questions are introduced; they are formulated on the basis of the literature study in Chapter 2. The design of the study is discussed in Section 4.4. The recruitment of participants is described together with the experimental materials. This chapter also introduces tests of proficiency and working memory capacity (Sections 4.6.3 and 4.6.4). This is followed by presenting a discussion of the procedures and methods applied in collecting the data (Section 4.8). Detailed explanation of coding procedures and measures of monitoring aspects will also be discussed in Section 4.9. This chapter concludes by describing the procedures of checking inter-rater reliability.

The fifth chapter introduces the results of statistical analyses. It presents answers to research questions in relation to the hypotheses stated in the methodology chapter. The chapter has ten sections: First, it starts with an introduction that gives an overview of other sections. Section 4.2 presents the descriptive statistics of L2 self-monitoring aspects in relation to two proficiency levels and two task conditions. The next section discusses the use of the Multivariate analyses of variance (MANOVA), and presents the results obtained from the MANOVA. This is followed by introducing the two-way between-groups univariate analysis ANOVAs and presenting the data of interaction effects of proficiency level and task condition on L2 self-monitoring aspects. Section 4.8 introduces an explorative analysis which is carried out to examine the impact of proficiency level and task condition on the temporal aspects of L2 self-monitoring (i.e. temporal phases of repairs). The results of the Pearson product-moment correlations between working memory capacity and L2 self-monitoring aspects in the two task conditions are then presented in Section 4.9. This chapter concludes by outlining the key findings of the analyses in relation to the research questions.

The sixth chapter of the thesis is dedicated to a discussion of the findings. This chapter starts by reviewing the aims of the current study, then presents the results of the data analysis in relation to the four research questions of the study. In this chapter, the results of the data analysis are interpreted in relation to previous studies, the Perceptual Loop Theory (the PLT), and L2 speech production model. Visual presentations of the results are provided to communicate the main findings. Section 6.2 of the chapter presents discussion about L2 self-monitoring aspects in relation to proficiency level. Section 6.3 discusses the data of L2 self-monitoring behaviour with respect to task condition. This is followed by discussing the interaction effects between proficiency and task condition in Section 6.4. Finally, the relationship between working memory capacity and self-monitoring behaviour in the two task conditions is discussed in Section 5.5.

The seventh chapter draws conclusions from the findings of the current study. It then highlights the significance of the results obtained and their contribution to advancing the theory of self-monitoring behaviour, namely, the Perceptual Loop Theory. The chapter then discusses the implications of the findings for research, assessment and pedagogy. The limitations of the study are acknowledged, and the chapter concludes by proposing some potential areas for future research in the area of L2 self-monitoring.

Chapter 2: LITERATURE REVIEW

2.1. Introduction

The goal of this chapter is to review and evaluate previous literature on self-monitoring. This includes reviewing some existing theories of L1 and L2 speech production and models of self-monitoring. As self-monitoring has been extensively studied in the L1 context, L1 research is reviewed to understand the different parameters of monitoring which can be useful in exploring L2 self-monitoring. This is followed by looking at a range of L2 studies that have examined L2 self-monitoring for the purpose of tracking development and identifying gaps in the field. This chapter also highlights how research in the area of automaticity can help us understand L2 self-monitoring behaviour. Finally, studies relating to the three variables of the current study, that is, proficiency level, task condition and working memory are reviewed and evaluated in terms of both theoretical and methodological factors.

2.2. Models of speech production

This section will present and evaluate selected models of L1 and L2 speech production. The selection of these models was based on theoretical considerations discussed in the following sections. The L1 speech production model (e.g., Levelt, 1983, 1989, 1999; Levelt, Roelofs, & Meyer, 1999) provides a detailed account of the underlying psycholinguistic processes of speech production. In the field of L2 speech, the model of L1 speech is adopted given that there is still no established model in this field so far. This chapter also presents and evaluates Kormos's (2006) model of L2 speech production.

2.2.1. The monolingual speech production model

Speaking is a complex cognitive process that involves many parts of the body and brain to create and coordinate the processing and delivery of speech. Researchers have developed speech production models in an attempt to understand how speech is prepared and executed. The most widely referred to model is the one developed by Levelt and colleagues (e.g., Levelt, 1983, 1989, 1999; Levelt et al., 1999) which involves various components that operate incrementally to turn communicative intention into overt speech. This model has received the most extensive empirical support in the field of cognitive psychology and psycholinguistics. For this reason, it has been adopted as a theoretical framework of the current study. This blueprint is modular, that is, it involves modules that work together in processing utterances. Although there is some disagreement among researchers on the details of how this model works (Levelt, 1989; Laver, 1973; Nooteboom, 1980), there is a consensus on the existence of four main processing components in this model which are the Conceptualiser, the Formulator, the Articulator, and the Monitor (see Figure 2.1 below). This section briefly presents these components. This model also proposes three knowledge stores: a store containing knowledge of the world, a mental lexicon, and the syllabary. These three stores will be discussed in relation to the corresponding speech components. For example, the mental lexicon store will be discussed in relation to the Formulator.

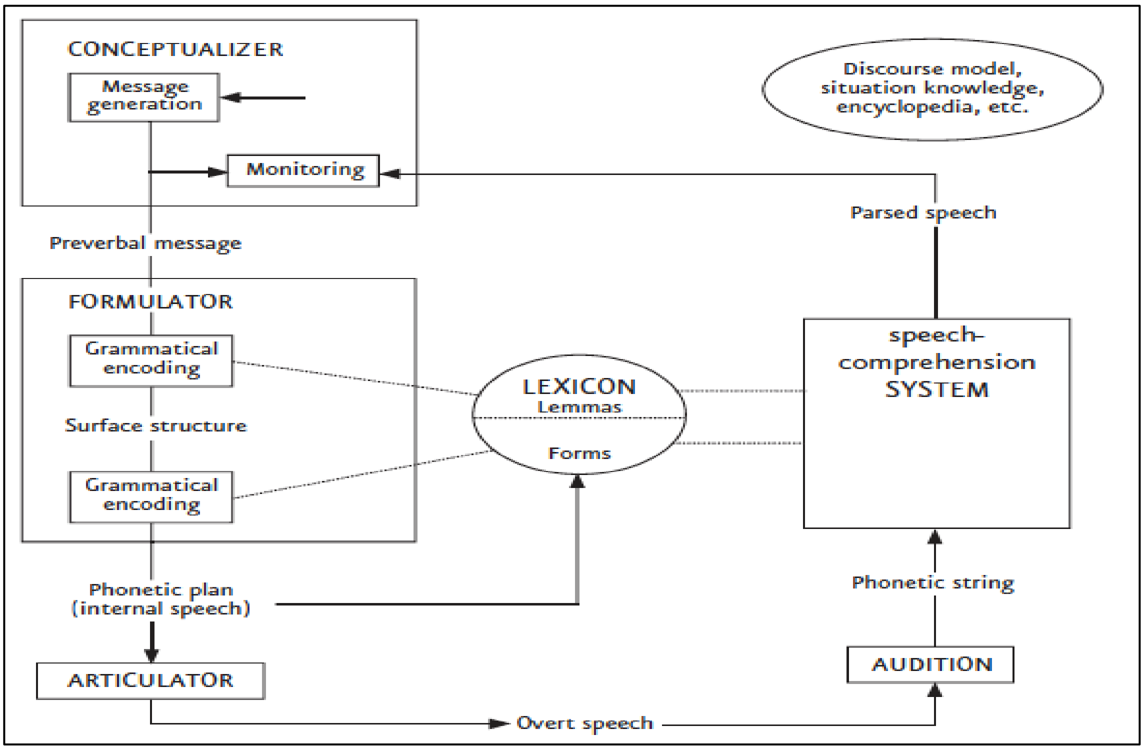


Figure 2.1. Monolingual speech production model (Levelt, 1989)

A. The Conceptualiser

The Conceptualiser is the first processing component which involves two stages: macro-planning and micro-planning (Levelt, 1989). At the conceptualiser, the communicative intention of speakers is turned into a preverbal plan. **Macro-planning** entails elaborating the communicative intentions into a series of sub-goals, namely, speech acts which can satisfy these goals such as assertion, question, instruction, command, etc. (Levelt, 1989, p. 158). **Micro-planning** entails giving perspective to the speech act, and also providing these speech acts with particular structures. It had been suggested that this stage is more concerned with conveying the speaker's standpoint in relation to the prepared message (Levelt, 1989). In other words, no words or syntactic structure are planned at this stage, although the preverbal message determines, at least to some extent, what will be required in the next stage. The final product that emerges from this stage is the preverbal plan which includes the necessary elements for the next stage, that is, the grammatical encoding (Levelt, 1989). The conceptualiser is widely assumed to be the slowest component of speech production (Boland, Hartsuiker, Pickering, & Postma, 2005; Felker, Klockmann, & De Jong, 2019; Hartsuiker, Catchpole, De Jong, & Pickering, 2008; Levelt, 1989) due to its controlled nature, that is, it largely draws on cognitive resources (for more details, see Sections 2.6.2, and 2.9.2).

B. The Formulator

The formulator draws on a mental lexicon store which has two parts of lexical representations: lemma which includes the syntactic features of lexis; and lexeme which contains the morpho-phonological features. The Formulation stage involves two main components which are grammatical encoding and phonological encoding. Grammatical encoding is mainly concerned with retrieving lemmas which are corresponding to the semantic information in the preverbal plan. The output of the grammatical encoding is 'the surface structure' which is defined as 'an ordered string of lemmas grouped into phrases and sub-phrases' (Levelt, 1989, p. 11). The surface structure is the input to phonological encoding which entails accessing lexemes that have the morph-phonological features of lexis. The phonological encoding results in the phonetic plan or the internal speech which contains the essential linguistic features of the utterance such as stress, pitch, phonemes, and morphemes. This stage is supposed to be largely automatic in L1 processing (Levelt, 1983, 1989).

A. The Articulator

After preparing the morpho-phonological features of the internal speech, the articulation stage starts by accessing the articulatory gestures in the syllabary store. This stage is mainly responsible for turning morpho-phonological representations to motor plan. The articulatory buffer is supposed to hold internal speech for about 100-200 seconds before speech execution (Levelt, 1989. Then, successive chucks of internal speech are retrieved by the Articulator from this buffer and sent for production (Levelt, 1989, p. 13). That is to say, the internal speech is turned into overt utterances through this last processing component.

B. The Monitor

Self-monitoring generally refers to the speaker's effort to check inner and overt speech, and when a speaker detects an error or inappropriateness in the utterance, the speech is halted and the correction is made before or after articulation (Levelt, 1983, 1999). Detailed discussion of monitoring will be presented in Section 2.4. According to this model, self-monitoring is a consciously controlled process, that is, it depends on cognitive resources in its functioning (see Section 2.6.2 for further discussion).

It is worth mentioning that in the latest update of Levelt's (1999) model, slight modifications were proposed to the structure of this model. That is, the speech processes were categorised to a dichotomous system which includes: semantic/syntactic component, and phonological/phonetic part as displayed in Figure 2.2 below.

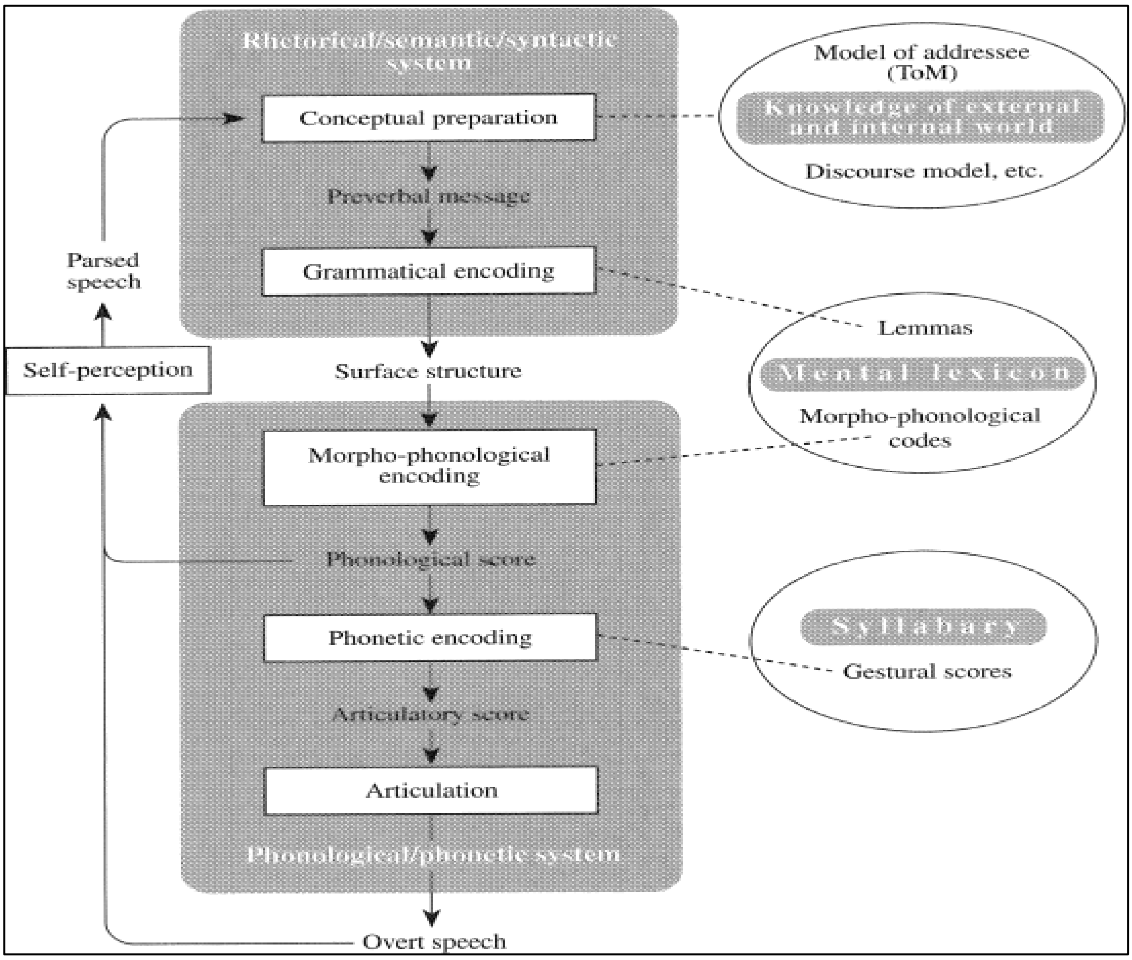


Figure 2.2. An updated version of Levelt's (1999) blueprint of L1 speaker

The semantic/syntactic part involves the conceptualisation of the message. This was referred to in the old version of this model as the Conceptualiser which involves the preparation of the preverbal plan. This system also includes the grammatical encoding which is concerned with retrieving lemmas and matching the semantic information to the preverbal plan to produce the surface structure. **The phonological/phonetic part** entails turning the surface structure into a phonetic plan. This system includes morpho-phonological encoding, phonetic encoding, and articulation. In the older version of this model, this system has been referred to as the formulation and articulation stage. The outcome of the phonological/phonetic system is overt speech. It can be noted that no principal changes were added to this latest update. Interestingly, many SLA researchers still refer to Levelt's earlier model in their studies as it seems to be a more clearly represented model of speech production (e.g., Ahmadian, Abdolrezapour, & Ketabi, 2012; Declerck & Kormos, 2012; Georgiadu & Roehre-Brackin, 2017).

This model has provided valuable information about L1 speech production with plausible details at each processing component. For this reason, this model has been adopted to the L2 speech production context. However, the differences between L1 and L2 speech processing and production cannot be ignored. The first difference between the two is that most L2 processing components are dependent on cognitive resources especially at lower levels of proficiency; thus, attentional resources are supposed to be largely consumed when speaking a second language. As self-monitoring is dependent on cognitive resources (Levelt, 1983), it is expected to be greatly affected by the consumption of cognitive resources during L2 processing. To provide explanation about how L2 speech processing and production work, a number of L2 speech production models were proposed (e.g., De Bot, 2003; Kormos, 2006). The next section presents the working model of L2 speech production developed by Kormos (2006).

2.2.2. The bilingual speech production model

Kormos (2006) proposed a working model of L2 speech production which was based on both Levelt's (1989) blueprint of L1 speech production and previous models of L2 speech production (e.g., De Bot, 2003; Poulisse & Bongaerts, 1994). Kormos's L2 speech production model shares the theoretical underpinnings of the L1 speech production model. For instance, it has the same processing components of the Conceptualiser, Formulator, Articulator and Monitor. In both models, the Conceptualiser and the Monitor involve controlled processing, that is, they draw on cognitive resources of working memory (see Section 2.10 for further details). Self-monitoring also draws on the comprehension system in both models.

This model shares the same three knowledge stores proposed by Levelt (1989): the store which contains the knowledge of the internal and external world; the mental lexicon store; and the syllabary. However, some modifications were proposed in the L2 speech model to account for the differences between L1 and L2 speech processing. First, the three knowledge stores are included in a large store which is called long term memory, and this store is shared between L1 and L2 (Kormos, 2006) (see Figure 2.3). That is, the long-term memory includes the episodic store (knowledge of the world), the lexicon store, and the syllabary. A fourth store has been proposed to specifically account for L2 processing, that is, the declarative knowledge store which includes syntactic, lexical, phonological elements which have been recently acquired but not yet automatized. The rationale for adding this store is that research from the neuroimaging field (e.g., Abutalebi, Cappa, & Perani, 2001) found that recently acquired rules are stored in a separate store. Therefore, a separate store for the newly learned rules was proposed, and named the declarative knowledge store. This store does not exist in the L1 model because L1 rules are repeatedly used and retrieved without much awareness, unlike L2 rules that need a conscious effort to be correctly executed (Kormos, 2006). This assumption, however, has not yet been empirically validated in L2 research.

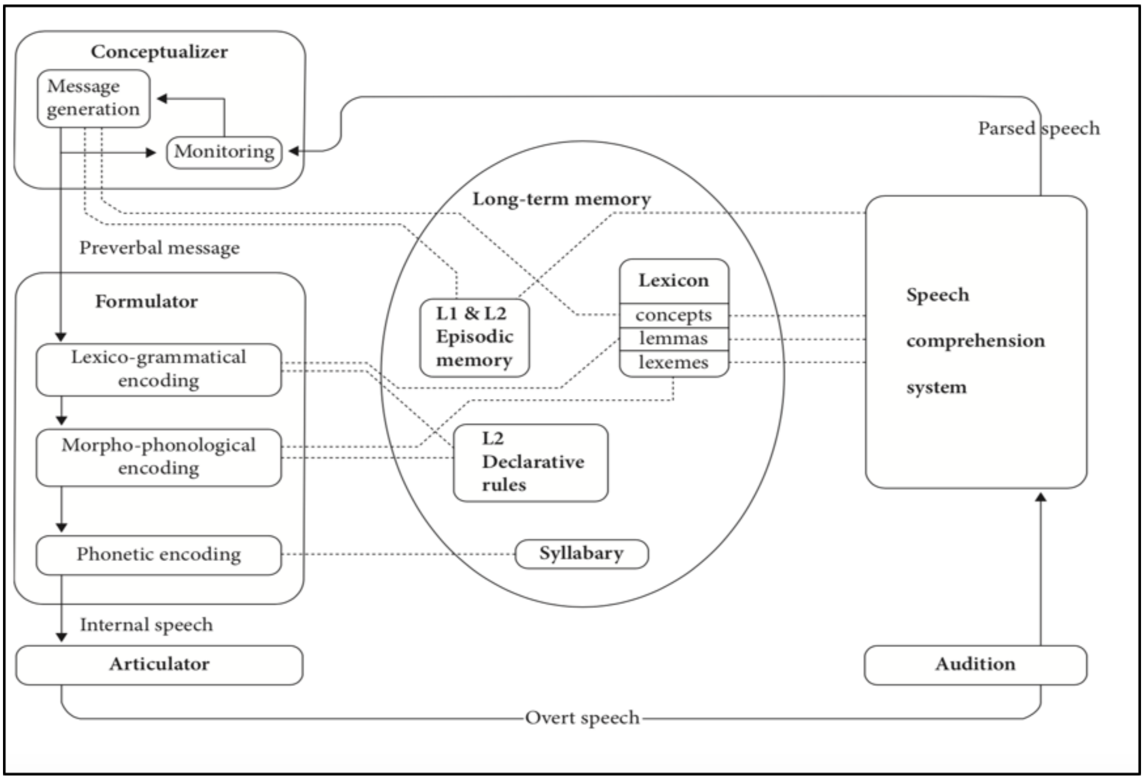


Figure 2.3. The model of bilingual speech production (Kormos, 2006)

Unlike the monolingual model, where most processes are automatic, L2 speech processes are largely controlled especially at the lower levels of proficiency. That is, L2 processing depends on cognitive resources especially at the lower levels of proficiency. Moreover, monitoring in this model is expected to be greatly influenced by resource limitations due to the fact that available cognitive resources are engaged by other L2 processes such as lexical, syntactic and phonological encoding (Kormos, 2006). Thus, limited cognitive resources are left for monitoring. However, it is still not clear to what extent resource limitations might influence L2 self-monitoring at different proficiency levels. Despite the fact that many theoretical considerations of Levelt's (1989) model have been empirically validated in the field of psycholinguistics (Kormos, 2006), L2 speech production model has not yet received much support from L2 research. In fact, there is still a need for further studies to empirically test various components of this model (e.g., self-monitoring) on L2 speakers at different levels of proficiency. Despite the importance of L2 self-monitoring (see Section 2.5), little attention has been given to this construct in L2 research. To understand the theoretical underpinnings underlying self-monitoring, the next section will look at the existing models of self-monitoring with emphasis on the Perceptual Loop Theory (PLT), because it is compatible with Levelt's model of speech production (1983, 1989, 1999; Levelt et al., 1999).

2.3. Theories of monitoring

Speaking involves people's endeavour to scrutinize their speech in an attempt to avoid the occurrence of errors or inappropriateness in their utterances. To understand how monitoring operates during speech processing, some theories of monitoring have been proposed. Understanding self-monitoring is crucial to any theory of speech production. It has been argued that 'any theory of language production is incomplete without a theory of self-monitoring' (Hartsuiker, 2014, p. 417). This section will look at three important theories of self-monitoring: The Node Structure Theory, the Production-Based Theory, and the Perceptual Loop Theory. There are some differences between these theories including the location of the monitor, its capacity, and its relationship with the cognitive systems, that is, the working memory and the comprehension system.

2.3.1. Node Structure Theory

The Node Structure Theory was based on the Spreading Activation Model of speech production which postulates that there are four levels of knowledge in L1 speech which are semantic, syntactic, morphological, and phonological. According to this model, speech production takes place as a result of the functioning of interactive networks of units (words, phonemes, morphemes) and interactive nodes that connect these units (Dell, 1986). The level of activation of a node determines whether it can be selected for processing (Dell, 1986). Following the Spreading Activation Theory, the Node Structure Theory presumes that monitoring is inherent in the production of speech, that is, if a node formed a connection with the wrong node, this could trigger prolonged activation which likely results in awareness leading to error detection. In other words, error detection occurs as a result of activation that engages networks of nodes within the system. This means that there is no particular component for monitoring because it is considered to be a distributed process that can take place at different levels (Postma, 2000). In this theory, the mental node system has seven levels: propositional nodes, conceptual compound nodes, lexical nodes, syllable nodes, phonological compound nodes, phonological nodes and feature nodes (MacKay, 1992). According to this theory, the monitoring process involves no resource limitations (Postma, 2000). This theory receives serious criticism because of the assumption that all errors will be detected, whereas in fact some errors might not be detected or corrected (Levelt, 1992). Also, in his review of this theory, Postma (2000) provided empirical evidence that has challenged the claims of the Node Structure Theory. His evidence includes experiments which used dual task condition and indicated that there was an effect of task demand on self-repair (Postma, 2000). Despite the fact that the effect was relatively small, it suggests that self-monitoring is consciously controlled, an idea which challenges the assumptions of the Node Structure Theory (Postma, 2000)

2.3.2. Production-Based Theory

The Production-Based Theory was based on the modular conception of speech production. This theory postulates that there are several monitors which are distributed throughout the speech production system, that is, there is a monitor for each processing component and sub-process (Laver, 1980). According to this theory, the monitoring processes do not need to be consciously controlled as these monitors operate within speech components (Laver, 1973; Nooteboom, 1980); thus, the monitoring system in this theory is not restricted by cognitive resources. In this model, the monitor can inspect inside the processing components rather than the end-products of these components. For example, the monitor that is located within the Formulator can detect an error in any of the sub-processes of the Formulator (i.e. syntactic, lexical, and phonological encoding), rather than the final outcome of the Formulator (i.e. the phonetic plan).

According to this theory, each local monitor needs to approve the output of each processing component before transferring it to the next stage (Laver, 1980). It needs to be understood that each stage can be activated only if it receives the output of the previous stage (Laver, 1980). This assumption has been seriously questioned as it is believed that such a strict function of the monitor is likely to have a detrimental impact on the speech flow (Blackmer & Mitton, 1991; Postma, 2000). Another shortcoming of this theory is that there is reduplication of knowledge as a result of locating several monitors within the processing components (Levelt (1983, 1989). In other words, as this theory allows the monitor to reach the inside of the processing components, different representations might be accessed by two processors at the same time. For example, during the formulation stage, retrieving lemma could be accessible to the Formulator and the Monitor at the same time. Such reduplication of knowledge during monitoring is considered a serious objection against the Production-Based Theory (Postma, 2000).

2.3.3. Perceptual Loop Theory (PLT)

To avoid the shortcomings identified in other theories of monitoring, the Perceptual Loop Theory (PLT) proposed a single central monitor that is located within the Conceptualiser and receives feedback from three channels which are called loops (Levelt, 1989). Unlike the Production-Based Theory, the PLT suggests that the monitor could only inspects the end-products of the processing components via its loops as displayed in Figure 2.4 below.

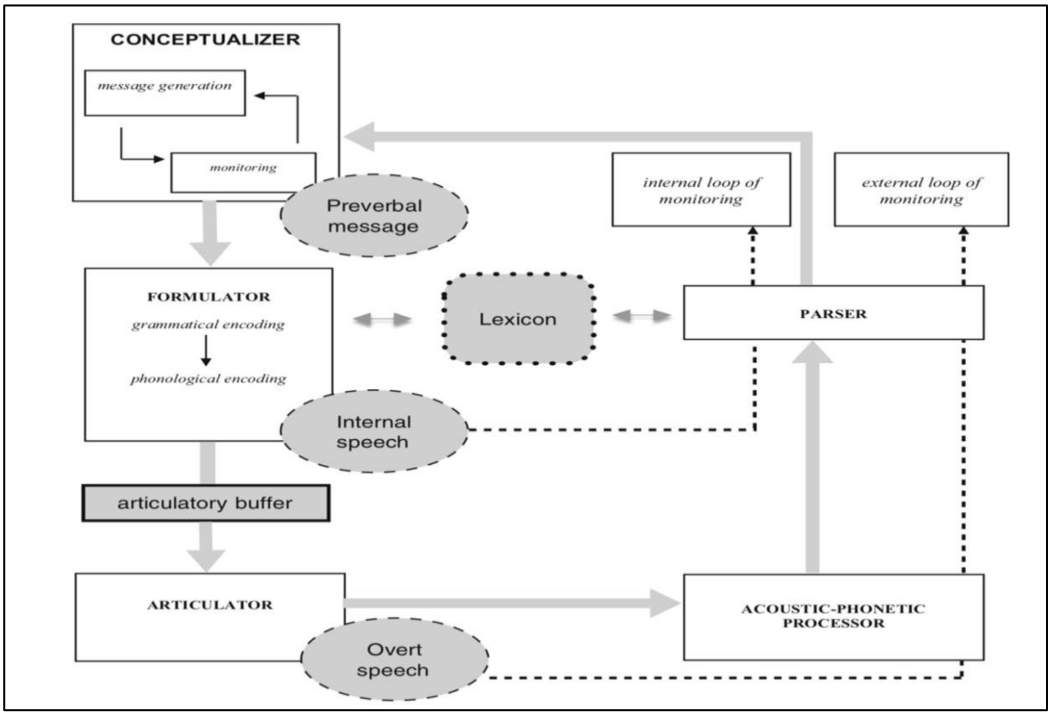


Figure 2.4. The Three Loops of Monitoring. Adapted from Levelt (1989)

Each loop inspects the outcome of a processing component. In other words, the perceptual loop checks the preverbal plan which is the end-product of the Conceptualiser. The inner loop is responsible for inspecting the phonetic plan which is the outcome of the Formulator. These levels of monitoring (i.e. the perceptual and inner loops) lead to repair, which is covert, that is, it cannot be examined directly. The auditory loop scrutinizes the output of the Articulator (i.e. the overt speech) resulting in overt repair. According to the PLT, **Covert repair** entails disfluency (filled pauses, hesitations and repetitions) which is considered as a corrective reaction to anticipated errors. That is, when a speaker anticipates an error, s/he produces pauses or repetition to buy time for correcting the error before articulation. This is called Covert repair. **Overt repair** will be discussed in detail in Section 2.4.1. The latest update of this model suggests that the monitor can reach inside particular intermediate levels of the Formulator (Levelt et al., 1999). However, it is still not clear how the processing components allow the monitoring to reach inside its sub-processes.

According to the PLT, an important characteristic of monitoring is that it is a conscious process, which means that the monitoring resources are limited (Postma, 1997). It has been stated that self-monitoring draws on working memory (Levelt, 1989) which is recognised as a limited-capacity cognitive system (see Section 2.10). Furthermore, the PLT postulates that the monitor draws on the comprehension system in checking one's own speech as well as the speech of others (Levelt, 1983, 1989). Despite the fact that the PLT has the most plausible premises among other theories of monitoring, it has been criticised due to a number of issues. For example, researchers who conducted their investigation on people with language impairment argue that the association between self-monitoring and comprehension might not be as strong as depicted in the PLT (for details, see Hartsuiker & Kolk, 2001; Postma, 2000). Also, some researchers challenge one of the premises of the PLT, namely, the Main Interruption Rule (MIR) which states that the speaker immediately interrupts the utterance upon detecting an error (Levelt, 1989). However, it has been found that some errors might not be immediately detected (e.g., Blackmer & Mitton, 1991; Broos, Duyck, & Hartsuiker, 2018; Oomen & Postma, 2001; Postma & Kolk, 1993; Postma & Noordanus, 1996; Seyfeddinipur, Kita, & Indefrey, 2008). Further discussion on this will be provided in Section 2.4.2. However, among the monitoring theories discussed above, the PLT seems the most viable approach with the most empirical support in the field of psycholinguistics; therefore, this model will be adopted as the theoretical framework for the current study. The next section will look at different aspects of self-monitoring in light of the PLT.

2.4. L1 Self-monitoring

According to the PLT, monitoring involves inspecting inner and overt speech so that when an error is detected, the utterance is halted and the correction is made (Levelt, 1989). There are **two main functions** for monitoring which are matching and issuing of instructions. Matching entails comparing the inner or outer utterance with the speaker's communicative intention and with the linguistic rules. In other words, the monitor compares what a speaker says with his/her original intention of the message and also with the linguistic rules of a certain language. This matching process takes place in the working memory (see Section 2.10). The second function is giving instructions for adjustment upon detecting an error or inappropriateness. That is to say, when the monitor detects a problem in the speech, an alarm signal is sent to the working memory and the speech is halted; then instructions are issued to speech components to make repair. This shows that there is an interactive relationship between monitoring and other speech components (i.e. the Conceptualiser, the Formulator, and Articulator). Types of repair vary according to the feedback received from the three loops. The next section will discuss different types of repair according to the classification proposed by Levelt (1983).

2.4.1. Taxonomy of L1 Self-repair

Levelt (1983, 1989) proposed a classification of L1 self-repair which has been widely used in a range of self-repair studies (see Section 2.4.3). Monitoring checks the speech before and after articulation, resulting in covert and overt repairs respectively (see Section 2.3.3). Overt repair includes three main types of self-repair: different-information repair, appropriateness-repair, and error-repair (Levelt, 1983, 1989). This section discusses these three main types of self-repair (see Table 2.1 below for summary). The first type of L1 self-repair is **different-information repairs** (D-repair) which occurs at the Conceptualiser level and entails re-arranging, modifying or replacing the pre-verbal message. The following examples are taken from Levelt's data (1983, 1989).

Example 1. We go straight on, or . . .we come in via red, go then straight on to green.

In this example, the speaker describes the direction of network pattern by saying 'we go straight on'; then he likely realises that there are more details that the listener needs to know thus he halts his speech and produces a new message with different information, 'we come in via red, go then straight on to green.' According to Levelt (1983), speakers choose to produce different-information repair **for two reasons**: First, when speakers intend to express some complex ideas, they may realise that the pre-verbal message needs to be re-arranged in a different way. This type of repair indicates that there is a problem at the conceptual level, where different parts of the message need to be differently re-ordered. Therefore, the initial plan is abandoned before completion, and a new one is then encoded. **Second,** speakers replace the current plan when it contains inappropriate information, thus they abandon the current preverbal plan and encode a different one. This type of repair shows that speakers monitor their messages after the processing of the message has started in the Formulator (Levelt, 1983). When deviation from original intention is realised, the speaker interrupts the flow of speech and produces a different message. This type of repair might be more cognitively demanding than the other types discussed below because it involves restructuring the whole message rather than part of it. This may explain why there are few examples of this type in Levelt's corpus, that is, there is only 1% D-repair reported in Levelt’s data. In other words, L1 speakers may only produce this type of repair when they think that a serious communication problem may occur. This is the only type of self-repair that results in replacing a speaker’s entire message.

Table 2.1. Levelt's (1983, 1989) classification of L1 self-repair

Type of repair	**Examples**
Different-information repair	We go straight on, or . . .we come in via red, go then straight on to green
Appropriateness-repair	
a. Ambiguity repair	We start in the middle with . . .. in the middle of the paper with a blue disc.
b. Appropriate-level repair	with a blue spot, a blue disc at the upper end
c. Coherence repair	To the right is yellow, and to the right –further to the right is blue
Error-repair	
a. Syntactic-repair	And black . . . from black to right to red
b. Lexical-repair	Well, let me write it back - er, down

c. Phonological-repair	A unut, unit from the yellow dot

The second main type of L1 repairs is **Appropriateness-repairs** (A-repair), which is triggered by a problem at the Conceptualiser level (Levelt, 1983). Unlike D-repair, appropriateness-repair indicates that the originally intended message is not changed, but the way in which the message is to be communicated needs to be modified. In other words, the pre-verbal message is correct, but the way in which the message is to be communicated may be inappropriate, incoherent, or ambiguous. There are three sub-categories which relate to appropriateness-repairs: ambiguity, appropriate-level, and coherence repair.

First, **the ambiguity repair** which is produced when a speaker realises that the utterance might be ambiguous for the interlocutor. Ambiguity repair may occur when information is insufficient, thus a speaker interrupts the speech flow to reproduce a preverbal plan in a clearer way (Example 2). In this type of repair, the needs of the listeners and the way they may interpret the message is taken into account. This type of repair usually occurs with the use of demonstratives or deictic expressions (e.g., from that one, the blue one) (Levelt, 1983).

Example 2. We start in the middle with . . .. in the middle of the paper with a blue disc.

In this example, the speaker is describing a network pattern, then he may realise that the listener would think that ‘in the middle’ refers to the middle of the pattern rather than the paper, and thus the speaker interrupts his speech and makes the repair ‘in the middle of the paper’ (Levelt, 1983). That is, by adding ‘the paper’ to the utterance, ambiguity is likely clarified for listeners.

The second sub-category of A-repair is **appropriate-level repair** which entails modifying the way of communicating an utterance to be more appropriate for the context of the conversation, the interlocutor, or the social standards (i.e. standards of formality and decency) (Levelt, 1989). This kind of repair usually makes the utterance more accurate and precise with respect to a particular context (Example 3).

Example 3. with a blue spot, a blue disc at the upper end

It has been explained that ‘a blue spot’ and ‘a blue disc’ refers the same concept when describing a network pattern (Levelt, 1983). However, the speaker chooses to replac**e** ‘spot’ with ‘disc’ in an attempt to be more precise and accurate in the given context (Levelt, 1983).

The third subcategory is **coherence repair** which entails changing and modifying the utterance to make it coherent with previous discourse (Example 4).

Example 4. To the right is yellow, and to the right –further to the right is blue

In this example, the speaker describes the first network pattern which is 'To the right is yellow.' After that, he starts expressing another pattern which is also 'to the right'. The speaker may then realise that the same expression was used in the previous sentence, thus he adds 'further' so that the utterance becomes coherent with a previously used utterance (Levelt, 1989). According to Levelt (1983), the sub-categories of Appropriateness-repair imply that the informational content of the message is correct, yet speakers choose to produce repair in an attempt to make their speech clearer, more coherent or appropriate. Brédart (1991) proposed a fourth sub-type to the Appropriateness-repair which is **repair for good language**. This sub-type includes pragmatic repair and repair for good language (Brédart, 1991). That is, speakers might find the utterance correct but they choose to make repair so that their speech becomes more appropriate for particular contexts or topics. The definition of this type of repair seems to have a considerable overlap with appropriate-level repair and coherence repair. In short, these two types of repair (D-repair and A-repair) are concerned with checking the message at the Conceptualiser level, which is different from the third type of repair (i.e. Error-repair).

Error-repair (E-repair) is the third type of repair introduced by Levelt (1983). This type of repair entails a different speech component, namely, the Formulator. There are three sub-categories to E-repairs which are syntactic, lexical, and phonological repairs (Levelt, 1983). **lexical-repair** entails correcting errors in lexical elements. This sub-category includes all groups of words: content words, function words, auxiliaries, prepositions, and modals (Levelt, 1983, 1989). That is, any trouble during lexical access is supposed to result in lexical repair (Example 5).

Example 5. Well, let me write it back - er, down, so that (Levelt & Cutler, 1983)

In this example, the speaker corrects 'back' because he may consider it wrong to say, 'write it back'. As such, he produces 'down' so that the utterance 'write it down' is correct. Levelt (1983, 1989) explains that lexical-repair occurs when a correct preverbal message is transferred to the Formulator, and during lexical encoding an erroneous lexical item becomes activated. It has been assumed that deciding on lexical-repair may not be easy because it may often involve some conceptual reference (Levelt, 1989, 1999). That is, speakers may replace a lexical element when they realise that the conceptual preparation of lexical input might be erroneous. According to Levelt (1983, 1989), this category includes words relating to colour, direction, prepositions, articles, morphologies, auxiliaries, and verbs. It could be assumed that adding morphologies and auxiliaries to this category might be confusing as these elements seem to have grammatical connections (see Section 2.5.1).

In addition to lexical-repair, E-repair includes another sub-type which is **syntactic-repair** that entails correcting any deviation in syntactic structure (Example 6).

Example 6. <u>And black</u> . . . <u>from black to right to re</u>d

In this example, the speaker describes a network pattern. He starts a syntactic construction (i.e. 'and black'), then he likely finds this syntactic structure incorrect, thus he reproduces another utterance with an accurate syntactic structure (i.e. from black to right to red). Syntactic-repair, as the title shows, is exclusive to troubles of syntactic structure rather than other grammatical errors (e.g., derivational, inflectional morphology, auxiliaries, prepositions).

The last sub-category of E-repairs is **phonetic-repair** which indicates that an error can be produced at the phonological encoding stage (Example 7).

Example 7. A <u>unut,</u> <u>unit</u> from the yellow dot.

In this example, the speaker might initially intend to say 'unit' but later hears himself producing 'unut' instead, therefore he interrupts the speech and produces the repair. This sub-type lacks detailed explanation and description of how phonological lapses occur. To summarise, these are the three sub-categories of E-repair which are produced to correct an error in the speech whether this error is lexical, syntactic or phonological. This is different from appropriateness-repair where repair is carried out to enhance the way a message is produced by making it more precise, appropriate or less ambiguous.

Levelt's (1983, 1989) taxonomy of L1 self-repair provides a comprehensive framework of L1 self-repairs which gives useful representations of the underlying stages of speech processing. This framework of self-repairs is an attempt to map out different types of repairs according to the compatible speech stages. Notwithstanding the great value of Levelt's classification, some minor issues might be identified. First of all, some examples provided in Levelt's taxonomy seem unclear, often leaving the concepts open to individual interpretation. Given the context of the tasks that these examples have been taken from, that is, network description task, it is difficult at times to establish a connection between certain examples and the category of repair being discussed (see example 6 above). These examples were mainly translated from Dutch which adds to their ambiguity. Moreover, the lexical-repair category in this classification has been criticised for including lexical elements that have grammatical connections (e.g., inflectional morphologies, auxiliaries, prepositions) (Kormos, 1998). It has been argued that the lexical elements would be better included in a different category, that is grammatical repair (Kormos, 1998). Likewise, there has not been enough discussion about the phonological-repair in this framework (Kormos, 1998). Further discussion is provided in Section 2.5.1. The next section presents the stages of making self-repair.

2.4.2. Structure of self-repair

According to the PLT (Levelt, 1983, 1989), there are three main phases for making repair: The Error-to-cut-off (Reparandum); Cut-off-to-repair (Editing term); and making repair (Reparatum). The first and the third phases are essential in L1 repair structure whereas the second phase is optional as there are examples of L1 repair that do not include this phase (Levelt, 1989).

A. Error-to-cut-off (First phase)

This phase involves producing an erroneous or inappropriate utterance which results in halting the speech flow (Levelt, 1989). The erroneous or inappropriate utterance is also referred to as the original utterance or the reparandum (Levelt, 1983, 1989). This interval is presumed to show the time needed to detect an error in one's speech (Levelt, 1989). According to the PLT, speakers interrupt their utterances immediately upon detecting an error or inappropriateness (Levelt, 1989). This is called the Main Interruption Rule which is assumed to be particularly correct in the case of E-repair which involves erroneous information (Levelt, 1989). A-repair is considered an exception to this rule because in this case no error is being made so the speaker may choose to interrupt the speech after the utterance being completed (Levelt, 1989). There is disagreement in the literature about this rule, as researchers argue that this phase is not exclusively used for error detection but for planning the repair as well (e.g., Blackmer & Mitton, 1991; Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008; Oomen & Postma, 2001; Pillai, 2006). One property of this phase which has been identified in previous literature is that it can be slowed down as a result of conceptual difficulty (Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008).

B. Cut-off-to-repair (Second phase)

This phase entails producing silent or/and filled pauses before executing the repair, and it is also known as the editing phase. According to Levelt (1983), this phase is optional in repair structure, that is, some repair might not include this interval (i.e. filled or silent pauses). This phase serves the function of holding the floor while planning repair and giving notification or a signal to listeners that there is a problem in the speech (Levelt, 1989). The PLT states that this phase is used for re-planning, as the interruption frees processing components to prepare the correction for erroneous utterance (Levelt, 1989). It was found that L1 speakers used more editing terms with E-repair than A-repair (Levelt, 1989). Using fewer editing terms with A-repair occurs due to the different nature of A-repair as it is not made specifically for correcting real errors but to add 'further specifications' to the speech (Levelt, 1983, p. 71). Blackmer and Mitton's (1991) findings contradicted the PLT's principle which suggests that Cut-off-to-repair is used for re-planning. Blackmer and Mitton (1991) found that in some cases the Cut-off-to-repair duration was zero millisecond, which was interpreted as meaning that re-planning might occur earlier during the previous phase (Error-to-cut-off phase) (see Section 2.3.4). Others argue that that these two phases are prepared and formulated as a single unit where error-detection and planning start with the phase of Error-to-cut-off, and re-planning continues to the end of the Cut-off-to-repair (Blackmer & Mitton, 1991; Broos et al., 2018; Hartsuiker et al., 2008; Oomen & Postma, 2001; Pillai, 2006).

C. Making Repair (Third phase)

This phase entails the production of repair; it is also known as 'repair proper' (Levelt, 1983, p. 42). It has been assumed that executing repair 'requires the speaker to have access to the structural property of the original utterance' (Levelt, 1983, p. 42). By using the structural property of the erroneous utterance or the original utterance, the speaker may gain in fluency (Levelt, 1983). Despite the fact that speakers endeavour to correct their erroneous utterances, many repairs are not correct themselves (Levelt, 1983). Levelt (1983) proposed that repair either has a regular structure which he referred to as well-formed, or an irregular structure which he called ill-formed. This rule states that repair is well-formed if the structure of the repair mirrors the structure of the original utterance. In other words, with the well-formed repair, the listener can predict the structure of repair from the structure of the original utterance. Another premise of this rule states that there should be a completion (constituent) in the repair (the third phase), and this completion should result in well-formed coordination. The following examples are given in Levelt's (1983) data:

1- To the right is a green, a blue node

2- He conquered Babylon, the great Alexander

Example (1) is considered well-formed because the word 'node' can be placed appropriately after the original utterance (i.e. a green node) and a connector 'and' can be placed between the two utterances (i.e. a green node and a blue node). However, Example (2) is ill-formed because there is no parallelism in this utterance and no connector could be placed correctly between the two parts (i.e. He conquered Babylon and the great Alexander). Levelt (1983) argues that this rule could distinguish between real repair and utterances that are just produced to add elaboration but not real repair. He applied this rule to most of his repair categories. However, this rule is complicated, and for practical reasons, it will not be included in the current study.

2.4.3. L1 self-monitoring research

Research on L1 self-monitoring has tested theoretical underpinnings of the PLT by employing various task conditions. One of the main premises of the PLT that has been tested by L1 researchers is the extent to which resource limitations have an impact on self-monitoring. One of the first studies that manipulated cognitive resources along dual task condition was conducted by Jou and Harris (1992). For further details about dual task condition, see Section 2.9.1. Jou and Harris (1992) examined L1 self-repair in two task conditions (single and dual task conditions). The single task condition involved asking participants to listen to audible texts which were read aloud and then they were asked to recall what they heard. In the dual task condition, participants were required to recall what they heard besides performing a simultaneous mental arithmetic task. The results indicated no differences between self-repairs in the two task conditions. This study was criticized for not investigating the effects of dual task condition on the ratio of error-correction as this measure is supposed to be a reliable predictor of self-monitoring (Oomen & Postma, 2002).

This shortcoming was addressed in a later study conducted by Oomen and Postma (2002). They examined the effects of dual task condition on the ratio of error correction, temporal phases of repair (Error-to-cut-off and Cut-off-to-repair), and disfluency. This study used a controlled network description where participants were asked to describe patterns of dots which mainly emphasized colors and directions of dotted lines. The participants were told that that their performances would be recorded and then presented to a listener who needed to complete a blank network based on their description. In the dual task condition, participants described the movements of the dotted lines as well as performing concurrent random finger tapping on a keyboard. The findings showed that the ratio of error-correction deteriorated in the dual task condition rather than in the single task condition. This finding was in line with the prediction of the PLT. However, the Error-to-cut-off phase was faster in the dual task condition than in the single task condition which contradicted the assumption of the PLT (i.e. cognitive demand should lead to longer duration of repair phases). Disfluency was not affected by the dual task condition. The network description task used in this study was criticized because it required participants to produce simple syntactic structures where the focus is mainly on whether the utterance is wrong or right (Van Hest, 1996). It seems that the simple version of this task greatly encouraged error-correction particularly correction of lexical errors. When cognitive demand increased in the dual task condition, the cognitive resources were divided along the simultaneous tasks, thus the error correction consequently decreased. It could be assumed that when investigating self-repair in different task conditions, researchers need to carefully choose a task that does not direct learners' attention towards one aspect of monitoring rather than another. Performing a network description task might not resemble real life contexts in terms of asking speakers to correct every single error they make, and therefore this task does not depict the foci of monitoring in real life contexts.

Oomen and Postma (2001) also investigated the effect of time pressure (i.e. fast and normal speaking conditions) on some aspects of L1 self-monitoring (the ratio of error correction and temporal phases of repair). They found no differences between the ratio of error-correction in these two conditions. This was interpreted as the sensitivity of L1 self-monitoring possibly increased during performance under time pressure, thus the same number of errors could be detected. They found that temporal phases of repair become shorter in the fast condition which suggests that the monitor could adjust its speed according to the task condition. Blackmer and Mitton (1991) also examined the temporal phases of repairs in the speech of 61 callers on a radio talk show. The temporal phases in this study were: Error-to-cut-off, Cut-off-to-repair, and total time of repair. The aim of this research was to test some premises of the PLT and other theories of monitoring such as the MIR (Main Interruption Rule) and the functions of different phases of repair. The findings showed that in some cases the repair directly followed the Error-to-cut-off with zero milliseconds for re-planning (i.e. Cut-off-to-repair). This suggests that L1 speakers plan corrections while speaking, and that they interrupt their speech only when they are ready to produce repair. This was considered a contradiction to the PLT which assumes that the Error-to-cut-off entails error detection, and that Cut-off-to-repair involves re-planning of repair. However, Blackmer and Mitton (1991) could not find an answer as to whether re-planning commences at the Conceptualiser or the component from which the problem originated (e.g., the Formulator).

Postma, Kolk, and Povel (1990) examined the relationship between errors, disfluencies and self-repair. The aim of this study was to test one of hypotheses of the PLT, that is, whether disfluencies are considered corrective actions to anticipated errors. Speech errors, self-repair and disfluencies were compared in two task conditions. In one condition, accuracy was prioritized by asking participants to pay attention to accuracy of their speech, in the other condition, the participants were told that accuracy was not important. The findings showed that in the higher accuracy condition, the error rate was lower than in the other condition (Postma, Kolk, and Povel, 1990). There were no differences between disfluency and self-repair in the two conditions. It was interpreted that disfluency behaves in the same way as self-repair which lent support to the hypothesis that disfluency may be used as corrective measures for anticipated errors (Postma et al., 1990). One of the explanations provided in this study is that participants may pay great attention to correctness of their speech in the higher-accuracy condition so that fewer errors are produced. However, the assumption that disfluency is corrective measure to errors as presented in this study needs more explanation. If disfluency is a corrective measure, it is expected to substantially increase in the higher-accuracy condition. Disfluency might serve different purposes during spontaneous speech production (Dornyei & Kormos, 1998; Krashen, 1981; Lennon, 1990), so that it could not easily be categorized.

Seyfeddinipur, Kits, and Indefrey (2008) also attempted to test whether L1 speakers interrupt their utterances the moment they detect an error (as predicted by the PLT), or when they are ready to produce the repair. To elicit L1 speech, they asked twelve native speakers of German to describe their houses. In this study, repair was classified as major and minor repair. Major repair includes fresh starts (i.e. abandoning the message and producing a new one), and minor repair involves modifying parts of the message. The data showed that in some instances L1 speakers halted their speech when they detected an error, but they did not do that 'by default,' as there were cases when speakers interrupted their speech only once the repair was ready, so that the duration of pauses was kept to a minimum level (Seyfeddinipur et al., 2008, p.841). It can be seen from reviewing some L1 studies on self-monitoring that these studies mainly focus on testing different theoretical underpinnings of the PLT. Some of the premises of the PLT have been challenged, e.g., the Main Interruption Rule. Also, researchers have disagreed with the PLT on the specific functions of the phases of repair (Error-to-cut-off and Cut-off-to-repair). The association between disfluency and self-monitoring was tested as well. Yet, there is still a need for more studies to provide further empirical evidence on some issues, particularly the function of the phases of repair.

This section (2.4) shows that L1 research laid the basis for examining self-monitoring behaviour in terms of theories and methods of investigation. It has been explained how L1 speakers check their speech for errors or inappropriateness, interrupt their speech upon detecting any deviation from the intended plan, and produce adjustments accordingly. L2 self-monitoring shares the main stages of self-monitoring (see Section 2.5 below). However, there are some areas in L2 self-monitoring that still need to be understood and explained. For example, the extent to which proficiency level can affect self-monitoring behaviour. Proficiency development is an important factor in investigating L2 self-monitoring as it can have influence on the availability of cognitive resources for self-monitoring (see Section 2.2.2 above). L2 researchers employed different methods to study L2 self-monitoring as can be seen in Sections 2.5 and 2.6 below.

2.5. L2 Self-Monitoring

In the field of second language, self-monitoring functions not only as an inspector or corrector of speech errors and inappropriateness, but also as a predictor of L2 learning and acquisition in different ways (Kormos, 1999). First, self-monitoring can lead learners to notice a gap in their knowledge, which in turn can trigger further acquisition (Kormos, 1999; Swain, 1995; Swain & Lapkin, 1995). Also, self-repair is initiated in a way that is similar to confirmation or clarification checks, which have been found to contribute to language learning (DeKeyser, 2010; Kormos, 1999; Pica, Holliday, Lewis, & Morgenthaler, 1989; Swain, 1985, 1995). Furthermore, self-repair is considered a useful predictor of language development (DeKeyser, 2010; Golonka, 2006). However, the efficiency of L2 monitoring has been assumed to depend on having sufficient declarative knowledge of grammar to allow for effective monitoring (DeKeyser, 2010). It has also been argued that repeated monitoring can improve fluency and automaticity of speech production (DeKeyser, 2007). This section will present L2 self-repair taxonomy which is based on Levelt's classification of self-repair. This is followed by reviewing and evaluating L2 research on self-monitoring.

2.5.1. Taxonomy of L2 repair

Building on Levelt's (1983, 1989) taxonomy of self-repair, Kormos has developed a comprehensive framework of L2 self-repair (1998, 1999, 2006). Her classification was supported by retrospective interviews obtained from L2 learners who were studying English in a foreign language context. She modified and refined the three main types of L1 repairs (i.e. A-repair, D-repair, E-repair), and added a fourth category (i.e. Rephrasing repair) to this typology which was found to be specifically produced by L2 learners (see Table 2.2 below for summary). In this section, Kormos's modification of the three types of L1 self-repair is discussed and evaluated. The fourth category that was added to the classification of self-repair is presented and assessed. This section ends with an assessment of Kormos's taxonomy in relation to the purpose of the current study.

Table 2.2. Repair types in Kormos's (1998) classification

Types of repair	Examples
Different-information-repair	we have some er er v may be you have vegetarians in your group
Appropriateness-repair	it doesn't it's not a problem
Error-repair	we could arrange er _ more _ smaller [taɪbɪə] tables if you would like that _ better
Rephrasing-repair	we will er reflect er to you in another letter we will answer you

Kormos (1999) added a third sub-category to Levelt's (1983, 1898) different-information repair (see Section 2.4.1 for details about D-repair). The third sub-category was called **message-replacement repair** which occurs when a speaker abandons one message and produces a different one. On the level of surface structure of this repair, it seems similar to D-repair produced by L1 speakers. However, Kormos (1999) argued that message-replacement repair occurs in L2 speech when L2 speakers realise that they cannot encode the current preverbal message due to their limited L2 competence, therefore they may abandon the preverbal message and encode a different one (Example 8). Examples and retrospections are taken from Kormos's (1999) data.

Example 8. 'we have some er er v may be you have vegetarians in your group' (Kormos, 1999, p. 321).

Retrospection: 'Here the idea of vegetarians suddenly popped up and I abandoned what I was going to say because I would not have been able to list any more types of food anyway' (Kormos, 1999, p. 321).

In this example, the reason for abandoning the message is the speaker's limited linguistic ability that does not enable her/him to continue speaking about different types of food. Identifying this type of repair is difficult without using retrospective interview because its linguistic features seem similar to that of D-repair (i.e. abandoning the message because it is inappropriate or inaccurate).

Next, Kormos (1999) agrees with Levelt's (1983) definitions of the three sub-categories of appropriateness-repairs (i.e. ambiguous, coherent and appropriate-level repairs). However, she adopts Brédart's (1991) proposal of adding repair for good language as a fourth sub-category to appropriateness-repair. It has been further explained that repair for good language includes two sub-categories (i.e. good language repairs and pragmatic repair), and these two sub-classes need to be distinguished (Kormos, 1999). She first suggests that pragmatic repair entails correcting utterances to be more appropriate for certain contexts (Example 9), and then she adds that repair for good language indicates replacing elements of utterances to be more sophisticated (Example 10) (Kormos, 1999). The first part of her argument shows a considerable overlap with Levelt's A-repair (i.e. Appropriate-level-repair). However, the second part explains what is meant specifically by repair for good language which entails speakers correcting their utterance in an attempt to sound more sophisticated (Example 10). It seems that this type of repair likely occurs in L2 context in the case of higher proficiency learners who want to sound like native speakers of the target language.

Example 9. 'it doesn't it's not a problem' (Kormos, 1999, p. 321).
Retrospection: 'First I wanted to say it does not matter but I realised that in a business deal you cannot say it does not matter' (Kormos, 1999, p. 321).

In this example, the speaker produces repair to make the speech more formal. Social standards and repair for formality purposes were mentioned by Levelt (1989) in his explanation of appropriate-level-repair. Therefore, adding pragmatic repair as a subcategory to A-repair might be repetitive and does not offer any novel development to this classification.

Example 10. ‘thirty-five per people’ (Kormos, 1999, p. 321).
Retrospection: ‘First I wanted to say persons but I had used persons several times before, so I said people’ (Kormos, 1999, p. 321).

In this example, the speaker realises that he is repeating some of the words he used before. Although the use of ‘persons’ is correct, he chooses to interrupt his speech and produce the repair (i.e. people). Kormos’s explanation of this sub-category (i.e. repair for sophisticated language) could be considered as a novel contribution which was not discussed in this particular sense even in Brédart's (1991) original proposal (i.e. repair for good language).

Error-repair is the third category of Levelt’s (1983, 1989) framework which is modified by Kormos (1999). There are three sub-categories to this type (i.e. lexical, syntactic and phonological repairs). Concerning lexical-repairs, there are similarities and differences between Levelt’s (1983) and Kormos’s (1999) frameworks. Similar to Levelt (1983), Kormos defines lexical repair as repairs which involve failure in lexical access whereby erroneous lemma might be activated. This repair type includes words, phrases, idioms, and preposition. Also, derivational morphology, such as replacing 'different' with 'difference,' belongs to this category (Kormos, 1999). Unlike Levelt (1983), Kormos (1999) suggests that repair of certain lexical sets should not be added to this category (e.g., inflectional morphologies, auxiliaries) because these sets might have a grammatical reference and need to be added to a different category. For instance, inflectional morphology, auxiliaries and prepositions that have syntactic connections tend to be different from other repairs in the lexical category. She therefore adds these types of repairs to a category called ‘grammatical-repair’. It should be noted that this category (i.e. grammatical-repair) is different from Levelt’s (1983) syntactic-repair (see Section 2.4.2). That is, syntactic-repair entails correcting the syntactic structure of utterances whereas grammatical-repair includes correcting syntactic structures as well as grammatical errors such as inflectional morphologies (i.e. tenses), auxiliaries (e.g., is, are) and prepositions which have syntactic connections (e.g., ‘on’ in ‘decided on’) (see Kormos (1998) for further details about the classification of prepositions). In short, it seems that adding a grammatical-repair category is adequate for analysing L2 speech because, unlike L1 speakers, L2 speakers might produce this type of repair more frequently due to the limited linguistic competence. The third sub-category to E-repair which Kormos (1998) modified is phonological repairs. Kormos (1998) suggests that phonological errors occur as a result of phonological slips of the tongue during the phonological encoding and articulation. This repair type includes correcting lapses that might occur at the intonation, stress, phoneme, allophone, or allomorph level (Kormos, 1998) (Example 11).

Example 11. 'we could arrange er _ more **_** smaller **[tɑɪbɪə] tables** if you would like that **_** better' (Kormos, 1998, p. 62).

This example shows that a phonological error (i.e. [tɑɪbɪə]) occurs during speech production, and upon hearing this error, the speaker interrupts the speech and produces the correct phonological form (i.e. tables). The repair types discussed in the above paragraphs are already present in Levelt's (1983, 1989) framework of self-repair, and Kormos (1998) has attempted to refine them to be more adequate for L2 self-repair research, whereas the repair type presented in the next paragraph is a new type, introduced by Kormos (1998).

In addition to the three main types discussed above, Kormos (1998) proposes a fourth main type of self-repair, **Rephrasing repair**, which is different from A-repair, D-repair and E-repair. Rephrasing-repair is produced when L2 speakers are not certain about the correctness of the utterances they produce due to their low proficiency level (Kormos, 1998). In other words, this repair type is different from D-repair and A-repair as these types of repair entail changing or modifying the preverbal plan or the way of producing the message. On the other hand, in rephrasing repair, the pre-verbal plan is revised without change, and the content of the utterance is slightly modified or paraphrased due to the uncertainty as to whether the utterance is correct or not. This uncertainty occurs because of the low proficiency level of the speaker (Kormos, 1998) (Example 12).

Example 12. 'we will er **reflect er to you** in another letter **we will answer you'** (Kormos, 1998, p. 63).

Retrospection: 'What happened here was that I was not sure whether 'reflect' really means 'answer' I knew what reflect means but I do not know whether you can use it for writing as well as that is whether it means the same in writing as in speech that you 'reflect on something'' (Kormos, 1998, p. 63).

The structure of this example seems to be similar to Appropriateness-repair where a speaker changes an utterance to be more accurate, appropriate or coherent. However, the retrospective interview shows that the speaker makes this repair because there is uncertainty about the correctness of the verb used, due to the limited linguistic knowledge. It is difficult to decide whether this repair is initiated at the Conceptualiser or the Formulator levels or engages both. Also, the structure of this repair could easily be confused with appropriateness-repairs. Therefore, it would be difficult, if not impossible, to reliably classify repairs into this category without retrospective interviews.

To conclude, Kormos's (1998) framework can be regarded as the most reliable classification of L2 self-repair due to the fact that this framework does not draw only on the linguistic characteristics of repair, but also on retrospective interviews which enables the researcher to map out repair types according to the psycholinguistic processes of speech processing. Furthermore, Kormos's (1998) framework adds a detailed explanation and specification to the self-repair framework proposed by Levelt (1983, 1989) which makes it adequate for investigating L2 self-repair processes. However, one of the primary issues that might be encountered when using this taxonomy in experimental studies with a complex design, is the practicality issue. That is, its heavy reliance on retrospective interviews, makes its implementation difficult, especially with a large number of participants. Rephrasing repair in particular cannot be reliably identified without retrospective recalls. Some researchers used practical technique to overcome this issue (see Section 2.6.4 below). The following section presents some L2 studies that employ this classification in the investigation of L2 self-repair and self-monitoring behaviour.

2.6. Research on L2 Self-Monitoring

L2 self-monitoring is defined as checking inner and overt speech against the existing linguistic system (Kormos, 2006). This suggests that functions of L2 self-monitoring likely change according to development in the L2 linguistic system. However, it is still not clear to what extent L2 self-monitoring aspects are affected by changes in the linguistic system. The purpose of the following sections is to review studies that have been conducted on L2 self-repair. To make it easier to navigate through information, this review is divided into four sections. In each one, studies are grouped according to the methodology they share, or repair classification they employ. That is, some researchers examined the effects of proficiency or tasks on self-repair following the taxonomy of Levelt (1983) and Kormos (1998). Others created their own classifications to explore self-monitoring. Few studies attempted to explore self-monitoring using qualitative methods. This review will then present an evaluation of the literature on L2 self-repair in an attempt to identify gaps that exist in this field.

2.6.1. L2 research on self-repair and proficiency

Perhaps the most important factor in distinguishing between L1 and L2 self-monitoring is L2 speakers' proficiency level (Van Hest, 1996). Some researchers examined how monitoring behaviour changes as a result of proficiency development. One of the earliest studies that attempted to bridge the gap between L1 and L2 self-repair in theory and research was conducted by Van Hest (1996). In her study, Van Hest (1996) examined L1 and L2 self-repairs at three levels of proficiency (beginning, intermediate, and advanced). The participants were thirty native speakers of Dutch who were L2 learners of English. Each participant performed picture description and interviews in both Dutch and English. She followed Levelt's (1989) taxonomy of the main types of repair, and added her own sub-categories for each main type of repair. Findings in this study showed that advanced L2 learners produced less E-repair, and more A-repair than the intermediate and lower proficiency learners. Interestingly, she found that the picture description elicited more A-repair than the interviews. It was assumed that picture description is more structured than the interviews, thus participants added specific details about pictures which led to a higher number of A-repair. Van Hest (1996) found that L1 speakers executed shorter A-repair temporal phases as compared to L2 speakers. It was explained that L1 speakers' knowledge of L1 was automatized so that L1 repair processing was faster than L2 self-repair. Also, A-repair was found to be processed significantly slower than other types of repair (lexical and phonological) suggesting that A-repair processing is different from other types of repair (Van Hest, 1996). Unfortunately, this study did not examine the temporal phases of repair in relation to different levels of proficiency or under different speaking tasks.

Examining the temporal phases of repair in different proficiency levels can be found in few L2 studies. One of these studies was conducted by Kormos (2000b), as she examined the effect of proficiency on the frequency of repair types, and the differences in the temporal phases of repairs among repair types. Participants were divided into three groups: advanced, upper-intermediate, and pre-intermediate learners, with ten participants in each group. L2 oral performances were elicited using information gap task. The data from a correlation analysis showed that the higher proficiency learners produced more A-repair and less E-repair than lower proficiency learners. There was no correlation between frequency and temporal phases of D-repair and proficiency. However, it showed that higher proficiency learners spent a shorter time on the Cut-off-to-repair (interruption) of E-repair and A-repair than the lower proficiency learners. It was argued that the PLT's premises that the Error-to-cut off is used for error-detection and that the Cut-off-to-repair is used for planning is applicable to L2 self-repair (Kormos, 2000b). It was also assumed that learners at a higher level of proficiency have some of their processes automatized and thus attentional resources become available for controlled processes. e.g., the Conceptualiser and the Monitor (Kormos, 2000b). In addition, it was suggested that L2 learners at a higher proficiency level focus on the appropriateness of their utterances, and that lower proficiency learners are concerned with the linguistic accuracy of their speech (Kormos, 2000b). This assumption was later supported by other researchers (e.g., Georgiadou, 2016; Kovač & Milatović, 2012).

Kovač & Milatović, (2012) examined self-repair in 101 Croatian L2 learners of English. The data showed that E-repair constituted about ¾ E-repair of all self-repair whereas A-repair was about 8.5% of the repair. The researchers assumed that their participants belong to the lower proficiency level based on the distribution of repair. They suggested that self-repair can be a good indicator of proficiency. That is, when E-repair constituted most of the repair produced, L2 learners can be considered as lower proficiency speakers. It was also found that the ratio of error correction was very low with only 1 out of 5 errors corrected. It was presumed that the monitor does not recognize errors due to insufficient declarative knowledge of linguistic rules and lexical items. However, this assumption needs to be verified through stimulated interviews. Other studies showed that proficiency level is not as effective as some other factors influencing repair frequency (Georgiadou, 2016). He investigated the impact of proficiency, error-tolerance and speaking habits on frequency of self-repair in Emirati L2 English learners. One of the operationalisations of error-tolerance in this study was participants' perceptions of an ideal L2 speaker. The results indicated that elementary learners produced less rephrasing repair than their intermediate counterparts. It was explained that the development of proficiency leads to qualitative rather than quantitative changes in L2 self-repair. However, it was found that participants' perceptions of the ideal L2 speaker contributed to making a greater number of self-repairs. It was thus concluded that personal beliefs about error correction plays an important role in the frequency of L2 self-repair.

The relationship between pause behaviour and self-repair was also studied in the L2 context. Williams and Korko (2019) examined the extent to which pause behaviour (silent and filled pauses) is affected by type of self-repair and proficiency level (lower intermediate and advanced groups). They employed two types of self-repair: false starts and correction which were based on Levelt's (1989) classification of appropriateness-repair and E-repair respectively. They found that the lower intermediate group produced considerably more silent pauses when they made corrections (E-repair) as compared to the advanced group. Unlike previous studies, they did not find differences between the two groups concerning the appropriateness-repair, which is called in this study false starts. However, the lower intermediate group produced twice as many E-repair (correction) as compared to the advanced group. It has been argued that silent pause frequency can be considered as a predictor of self-repair across different L2 proficiency groups. However, this study highlighted only one of the self-repair phases which is pauses that occur after erroneous utterance. According to the PLT, this is known as the editing phase or the second phase of self-repair. To understand self-repair and self-monitoring behaviour, all phases of self-repair need to be examined across different types of repair as discussed above. In fact, most of the studies discussed in this section have reached the same conclusion. That is, at later stages of proficiency development, some speech processes (e.g., the Formulator) can reach a certain level of automatization and thus some resources might become available for monitoring processes (for further discussion, see Section 2.7). It has also been contended that self-repair can be a useful predictor of proficiency.

2.6.2. L2 research on self-repair and tasks

In the field of TBLT, several researchers manipulated task types and task conditions to explore the influence of task complexity on L2 monitoring behaviour (e.g., Ahmadian, Abdolrezapour, & Ketabi, 2012; Ahmadian & Tavakoli, 2014; Gilabert, 2007). For instance, Ahmadian and Tavakoli (2014) examined the impact of pressured and careful online planning conditions on self-repair. Careful online planning entails providing L2 participants with abundant time during task performance whereas pressured planning requires L2 learners to speak in a limited time. Thirty intermediate L2 learners performed an oral narrative task under careful online planning and pressured online planning conditions. Findings showed that L2 learners produced more E-repair and fewer A-repair and D-repair in careful online planning as compared to pressured online planning conditions. This finding was interpreted as meaning that the careful online planning promotes the functioning of both inner and auditory loops of monitoring. However, it could be assumed that the careful online planning may play a facilitative role for the auditory loop (rather than the inner loop); thus, L2 learners detected and corrected more overt errors than in the pressured online planning condition. Qualitative data in this study indicated that planning time was used by L2 learners to monitor their speech in terms of linguistic accuracy and appropriateness of their utterances, and to plan and monitor the utterances they are about to produce. This is one of few studies that discuss how disfluency is used to monitor L2 speech. This finding is in line with Tavakoli (2011) who suggested that mid-clause pauses are associated with online planning and monitoring.

The influence of different degrees of task structure on L2 monitoring was examined by Ahmadian, Abdolrezapour, and Ketabi (2012). Besides examining the main types of repair (i.e. A-repair, E-repair, and D-repair), this study drew attention to the role of covert repair (i.e. hesitation and mid-clause pauses) in L2 monitoring. Findings showed that the structured task led to the production of more E-repair, whereas the unstructured task directed L2 learners to produce more A-repair and D-repair. It was assumed that the structured task directed learners' attention towards the Formulation and the Articulation stages of speech production. The unstructured task, on the other hand, directed the attentional resources towards the Conceptualiser level. This study is important because it showed that monitoring foci could be manipulated via task structure as the less structured task directs learners' attention towards conceptual aspects of their messages, whereas a more structured task leads them to concentrate on the linguistic aspects of their utterances. More importantly, L2 learners in this study commented that they used mid-clause pauses and hesitation to monitor their speech and prepare their self-repair. However, mid-clause pauses and hesitation in this study were examined on a small scale using retrospective interviews. The study could have provided valuable information about self-monitoring if disfluency had been examined under different task structures.

The impact of different types of tasks on L2 self-repair was investigated by Gilabert (2007). In this study, self-repair behaviour was examined under simple and complex versions of different types of tasks. Three different tasks were employed, namely, narrative, direction-giving and decision-making tasks. In the study, it was found that the narrative task triggered a higher rate of self-repair than the other tasks. This was interpreted as being due to the fact that the narrative task was less demanding and thus more attentional resources were available for monitoring. Although it was assumed that the simple versions of the three tasks would impose less attentional demand on the learners, it was not clear why L2 monitoring behaviour was almost the same in the simple and complex versions of the tasks, with the exception of the direction-giving task. However, one of the limitations of this study is that it did not follow Levelt's (1983) classification of self-repair. As such, the findings of the study cannot be compared to other L2 self-repair studies. In fact, this issue is encountered in a number of L2 studies in TBLT field (e.g., Craig, Kormos & Danny, 2016; Tavakoli & Skehan, 2005; Wang & Skehan, 2014). Most of these studies examined self-repair as an aspect of fluency following the taxonomy of L2 fluency proposed by Tavakoli and Skehan (2005). Self-repair in this taxonomy includes repetitions, replacements, reformulations and false starts. Although this taxonomy is valuable in studying L2 fluency, it is not as effective in examining L2 self-monitoring behaviour.

In short, this section has presented how L2 self-monitoring foci could be affected by different task conditions and task types. As self-monitoring is sensitive to contextual effects (Levelt, 1983), the above studies attempted to understand the extent to which self-monitoring can be affected by different task types and task condition (Ahmadian, Abdolrezapour, & Ketabi, 2012; Ahmadian & Tavakoli, 2014; Gilabert, 2007). A variety of contextual effects have been demonstrated in this section such as careful and pressured online planning; structured and unstructured tasks; simple and complex versions of narrative, direction-giving and decision-making tasks. However, dual task condition has rarely been employed in L2 research to examine L2 self-monitoring. Section 2.9 will present a dual task paradigm and its importance to self-monitoring research.

2.6.3. L2 research and qualitative methods

Most L2 self-monitoring studies employ a quantitative method to examine self-repair behaviour. Few studies have used stimulated recalls to explore different aspects of self-repair. Some of the studies discussed above employed stimulated recall on a small scale, e.g., Ahmadian and Tavakoli (2014), and Ahmadian et al., (2012). This section presents studies that involve a large-scale examination of self-repair using retrospective interviews (e.g., Kahng, 2014; Simpson, Eisenchlas & Hughes, 2013). Simpson, Eisenchlas and Hughes (2013) examined the extent to which self-repair relates to monitoring behaviour and proficiency. Their study collected oral performances and retrospective stimulated interviews from four L2 learners of Chinese throughout three months of their study in China. The researchers argued that the relationship between self-repair, proficiency, and monitoring is more complex than other studies have depicted. That is, L2 learners may produce self-repair for a variety of reasons, not merely for linguistic errors and inappropriateness. They argued that self-repair cannot be considered a direct indicator of proficiency; it merely reflects learner's understanding of L2 semantics and syntax. Such a claim, nonetheless, cannot be verified as this study only examined a small sample size (four participants) belonging to the same proficiency group.

Another study which used stimulated recalls was conducted by Kahng (2014) who collected oral performance and retrospective interviews about the cognitive processes that occur while L2 learners speak in the second language. Qualitative data in this study revealed that lower proficiency level learners commented on issues relating to linguistic aspects of their utterances more frequently than the higher proficiency learners who focused on aspects relating to message contents. It was suggested that lower proficiency learners were aware of processing difficulties in grammatical and lexical retrieval. This means that their monitoring behaviour relates to linguistic aspects of the utterance whereas the higher learners monitoring behaviour is linked to the meaning and suitability of their messages. To conclude, employing qualitative methods to explore L2 self-monitoring may yield important information about self-monitoring behaviour when conducted on a large scale, and involving different proficiency levels and different contexts.

2.6.4. L2 dichotomous classification of self-repair

Some L2 studies have attempted to develop their own classifications of self-repair by grouping categories and sub-categories of self-repair into a dichotomous taxonomy. That is, one category typically involves repair that relates to linguistic errors, and another category includes repair that entails appropriateness or conceptual aspects of the utterance. For instance, O'Conner (1988) applied two types of self-repair: corrective repair and anticipatory repair; the corrective repair is used to correct errors and mainly focuses on linguistic accuracy, whereas the anticipatory repair underlines discourse level repairs. The participants were six American speakers of L2 French at two levels of proficiency (three beginners and three advanced). No differences were found between the frequency of repairs in the two proficiency groups. However, the types of repairs in the two groups were different. The lower group produced corrective repair that focused on linguistic accuracy. The advanced learners focused on discourse level repairs. It was concluded that the reason lower learners focused on the linguistic errors is likely due to their limited knowledge of the target language. O'Conner (1988), therefore, suggests that self-repair is a useful indicator of various stages of interlanguage development. This taxonomy seems practical and it was adopted with minor changes in a number of studies that followed.

A number of studies employed similar dichotomous taxonomy to investigate L2 self-repair (e.g., Simard, Fortier, & Zuniga, 2011; Simard, French, & Zuniga, 2017; Zuniga, 2015; Zuniga & Simard, 2019). In these studies, L2 self-repair was divided into Form-repair (F-repair) and Choice-repair (C-repair), where F-repair entails correcting linguistic elements and C-repair relates to the conceptual content and lexical elements of the message. It has been argued that lexical-repair needs to be added to C-repair provided that lexical elements have conceptual preparation that take place at the conceptualisation level according to the latest update of Levelt's classification (Levelt et al, 1999). In Zuniga and Simard's (2019) study, it was found that proficiency level was not a very strong predictor of L2 repair, and that attentional control and L1 self-repair together explained about 35% of variances in L2 repair. One shortcoming of this study is that its finding cannot be compared to previous studies due to differences in self-repair classification, particularly in their categorization of lexical-repair. However, this study showed an interesting explanation for different traits of attention. It has been shown that attention not only maintains learner's focus during task performance, but also regulates resources for parallel processes during L2 oral performance (Zuniga & Simard, 2019). It has been assumed that this latter function of attention might be applicable to self-monitoring. That is, self-monitoring draws on attention to efficiently maintain balanced processing rather than focused processing during L2 speech production (Zuniga & Simard, 2019).

To conclude, using one comprehensive taxonomy of L2 self-repair has the advantage of facilitating comparability between studies, and helping to identify gaps in literature, as well as illustrating how to address such gaps in future research. It can be seen from this review of literature that L1 research mainly focus on testing hypotheses relating to monitoring theories. L2 research, on the other hand, examines monitoring in different task conditions and levels of proficiency. Comparing L1 and L2 self-repair was also a common theme in L2 research. However, L2 self-repair research has rarely attempted to test assumptions of self-monitoring theories or related the findings to monitoring theories. It can also be seen that most L2 studies concentrate on self-repair as the sole indicator of self-monitoring, while little attention is being paid to other parameters of self-monitoring behaviour such as ratio of error-correction, disfluency and temporal phases of repair. According to L1 researchers, dual task paradigm is an important technique in examining the impact of resource limitations on self-monitoring behaviour as demonstrated in Section 2.4 above. This method has rarely been employed in L2 research. More discussion about the dual task condition will be presented in Section 2.9.1. The next section (2.7) will look at automatisation process and how it can affect self-monitoring behaviour.

2.7. Automaticity and automatisation

The term "automaticity" is generally used in cognitive psychology to describe 'properties of behaviour that reflect the individual's ability to perform very rapidly and with little or no (conscious) effort' (Segalowitz, 2012, p.53). Automaticity refers to the reduction of attention when performing a cognitive activity, as attentional control is recognized as involving intention, awareness, and depletion of cognitive resources as a way of dealing with limited processing capacity (Kahneman, 1973). Researchers agree that the reduction or absence of attentional control while a process is being executed demonstrates that performance becomes automatic (DeKeyser, 2001; Segalowitz, 2003). Some researchers have linked parallel processing with automatic behaviour, and serial processing with controlled or attention-based performance (Schneider & Shiffrin, 1977). Other definitions of automatic behaviour show that a process is automatic when it operates in a fast and parallel manner with less interference from other tasks (DeKeyser, 2001; Pashler, 1998; Schneider, Dumais, & Shiffrin, 1984; Segalowitz, 2003). This means that once the process becomes automatic, more attentional resources can be freed up for other purposes (Segalowitz, 2003). Regardless of which approach one takes, the key characteristic of automaticity recognized by most researchers is that when a process reaches an advanced level of automatization, performance is likely to become more accurate and stable (Segalowitz, 2003; Skehan, 2009; Tavakoli, 2019).

Different accounts have been proposed to explain how automatization takes place. Anderson's (1983) model of skill acquisition (i.e. Adaptive Control of Thought ACT), proposes that knowledge starts as explicit (declarative knowledge), and then transfers to specialised procedural knowledge. According to this model, there are five stages for the development of automatic performance: composition, proceduralisation, generalisation, discrimination, and strengthening. Anderson (1983) believes that automatization does not only involve speeding up, but qualitative changes take place in the performance. Cheng (1985) suggests that automatization is a restructuring process where the development of automaticity leads to a qualitative change in task components. It has been assumed that automatization leads to the creation of a new mechanism which results consequently in rapidity of performance (Cheng, 1985). This assumption has been supported by some researchers (e.g., Segalowitz & Segalowitz, 1993) who believe that automatization is a product of qualitative change rather than mere speeding-up of performance. Other theories of automaticity attribute the development of automaticity to memory, e.g., Logan (1988), who presumes that automatic processing is similar to memory retrieval. Mackay (1982) also suggests that automaticity develops with practice, where the association between stimuli and response becomes stronger. Researchers argue that this procedure of the development of automaticity (automatization) is more important than automaticity itself (Segalowitz, 2003). In other words, the changes (qualitative and quantitative) that take place during automatization are more important than the end-product of the process, namely, automaticity (Segalowitz, 2003). These definitions and theories of automaticity have mainly been proposed to describe the development of general cognitive skills.

2.7.1. Automatisation and L2 processing

It is widely held in the field of cognitive psychology that language learning is a form of complex cognitive skills. However, a consensus has not been reached about what automaticity is, or how it can be operationalised in relation to language processing and production (Segalowitz, 2012). The theory of automaticity which has been adopted in SLA is Anderson's (1983) model of skill acquisition (i.e. Adaptive Control of Thought ACT). Based on this model, it has been claimed that L2 learning proceeds through the same stages as other skilled performances (DeKeyser, 2001). In other words, learning a second language in instructed settings is likely to start as declarative knowledge, which is largely dependent on attention (DeKeyser, 2001). Then, with a culmination of frequent use, language knowledge can become proceduralised, that is, less attentional control is needed for oral performance to be executed efficiently (DeKeyser, 2001). However, there is disagreement among L2 researchers about whether and how explicit knowledge can be automatized (Segalowitz & Hulstijn, 2005).

The monolingual speech production model proposed by Levelt (1983) can be regarded as the starting point for studying automaticity in relation to language processing (see Section 2.2). According to this model, the Conceptualiser and the Monitor draw on cognitive resources during its functioning whereas the Formulator and the Articulator are largely automatic in L1 speech processing due to the culmination of use during a lifetime. According to the theoretical underpinnings of automaticity, when cognitive components become automatized, the controlled components can not interfere in their operations (Segalowitz & Hulstijn, 2005). In other words, the consumption of cognitive resources by controlled components (e.g., the monitor) does not affect the function of the automatized components such as the Formulator. However, the question that rises in this context is whether the automatization of speech components (e.g., the Formulator and the Articulator) affects the functioning of controlled components. In other words, it is not clear to what extent the controlled processes (i.e. the Conceptualiser and the Monitor) would benefit from the cognitive resources that are released by the automatized processes (the Formulator and the Articulator) especially in L2 processing. In this respect, Segalowitz and Hulstijn (2005) suggested that L2 processing can be a testing ground for automaticity. The distinction between controlled and automatic processing can be reflected in earlier and later stages of proficiency development (Segaowitz & Hulitijn, 2005). In other words, at the lower levels of L2 proficiency, it is expected that L2 speech processes largely draw on cognitive resources. The consumption of cognitive resources at the lower proficiency level is expected to affect different aspects of L2 speech (Kormos, 2006).

The potential area that can provide novel information about automaticity in L2 processing is the development of L2 speech processes such as lexical retrieval, grammatical encoding, and fluency of oral production. Researchers have found that syntactic processing can reach automatization with repeated practice (DeKeyser, 1997; Robinson, 1997; Rodgers, 2011; Towell, Hawkins, & Bazergui, 1996). This suggests that syntactic encoding as an important sub-process of the Formulator can be automatized. Furthermore, researchers in the field of lexical access and word recognition have contended that some teaching approaches can lead to automatization of lexical access (Hulstijn, van Gelderen & Schoonen, 2009; Pellicer-Sánchez, 2015). These studies imply that the development of the Formulator sub-processes is very important in understanding automaticity in L2 processing. In fact, it has been assumed that the automatization of these two sub-processes of the Formulator is expected to reduce disfluency in L2 speech (Segalowitz, 2010). It has also been assumed that studying fluency of oral production can be a useful predictor of the extent to which production and comprehension have become automatized (Tavakoli, 2019). Examining L2 fluency may reveal important information about the automatization of both the Formulator and the Articulator.

Automaticity in second language has been presumed to be similar to processing efficiency (Segalowitz, 2010). This can be explored through studying processing at different levels of proficiency where it can be useful to track development of the underlying speech processes and oral speech production. Researchers argue that for automaticity to be a useful construct in L2 processing, studies need to assess various aspects of the automatized process rather than merely measuring speeding-up (DeKeyser, 1997; Robinson, 1997). The qualitative changes in L2 speech processing that emerge throughout proficiency development need to be examined and assessed. It would be helpful to look at automatization in L2 processing as a continuum rather than a dichotomous procedure that entails either controlled or automatic processes, especially as certain speech components are inherently controlled. The development of such controlled processes might not entail complete absence of cognitive control such as the monitoring. Perhaps the most important question to ask when investigating automatization of L2 processing is about the extent to which the inherently controlled processes (e.g., the monitoring) can be affected by the automatization of other speech components. The next section will discuss the relationship between monitoring and automaticity.

2.7.2. Automatisation and L2 monitoring

If we assume that automatisation is a continuum, monitoring and automaticity are as the two ends of this continuum, since they entail dissimilar associations with cognitive resources. In other words, automaticity refers to the absence of attentional control whereas self-monitoring is 'hardly ever made without a touch of awareness' (Levelt, 1989, p. 20). Self-monitoring as viewed in Levelt's Perceptual Loop Theory (1983, 1989) is a controlled process. Controlled processing is widely held to be capacity limited (i.e. drawing on available central processing resources), and needs conscious attention to function effectively (Shiffrin & Schneider, 1977). Automatic processing, on the other hand, is not affected by resource limitations, and functions without awareness (Shiffrin & Schneider, 1977). It has been stated that the controlled nature of the monitor makes it subject to limitations of working memory (Levelt, 1983) (For details, see Section 2.10.3). This assumption was contradicted by the production-based monitoring model which predicts that monitoring is relatively less dependent on the available processing resources (see Section 2.3.2 above).

With monitoring system being inherently controlled, a question that might be asked is whether the degree of its dependence on cognitive resources can be changed over time. To answer such a question, researchers have attempted to compare L2 and L1 monitoring (e.g., Broos, Duyck, & Hartsuiker, 2018). It has been assumed that L2 monitoring is more dependent on cognitive resources than L1 monitoring (Broos et al, 2018). It has also been argued that the monitoring system is more active in L2 than in L1 processing, however, it is still not obvious in what ways L2 monitoring is engaged. In other words, it seems that there are aspects or functions of the monitoring system that can be automatized with the culmination of use which means that it probably requires fewer resources in advanced levels of language development. Yet, it is still not clear which aspects of monitoring might develop and how they can be automatized.

A promising area for exploring self-monitoring in relation to automaticity is proficiency development. One important aspect that can be examined is the reduction in error rate which has been predicted to be one of qualitative changes that emerge during automatization of linguistic knowledge (DeKeyser, 2013). To date, studying errors and repair features in L2 speech has attracted increasing research because types of self-repair has been suggested to be a proxy measurement of automaticity (see Section 2.6.1 above). Although research shows that global self-repair frequency is not greatly affected by the development of proficiency (Georgiadou, 2016; Kormos, 2002), different types of self-repair are shown to be considerably influenced by proficiency level. That is, higher proficient L2 learners produced more repair that focuses on appropriateness of their messages (i.e. discourse level), whereas the repair of the lower proficient entails correcting linguistic aspects (e.g., Gilabert, 2007; Kahng, 2014; Kormos, 2000a; Van Hest, 1996) (Further discussion can be found in Section 2.6.2). Such a qualitative change in self-repair has been assumed to relate to the automatization of related speech components. In other words, with the development of proficiency, certain sub-processes of the Formulator become automatized and thus some attentional resources are freed up to be in the service of other controlled processes (Kormos, 2000a; Van Hest, 1996). The automatization of syntactic encoding and lexical access (see Section 2.7.1) could probably minimise the functioning of monitoring in terms of checking utterances against linguistic rules (For more about monitoring functions, see Section 1.4). Researchers have argued that the released cognitive resources are likely used to serve other functions of monitoring such as inspecting utterances against speaker's communicative intention (Kormos, 2000a; Van Hest, 1996). This shows that despite the fact that self-monitoring is a controlled process, the automatization of other processes seems to influence the functioning of monitoring. It is expected that studying proficiency can reveal important information about the development of monitoring functions. Furthermore, researchers have suggested that dual task condition is useful in investigating the relationship between cognitive resources and functioning of the L2 monitoring system (Broos et al., 2018). For further details about the dual task condition, see Section 2.9.1.

2.8. Proficiency level

Proficiency is an important factor in investigating second language speech processing and production. Proficiency is predicted to play an important role in understanding language processing in L2 learners (Dunn & Fox Tree, 2009). Neuroimaging studies argued that language proficiency is likely the most important factor that influences the bilingual language system, more important than the age of acquisition (Abutalebi, Cappa & Perani, 2001). Previous research on monitoring has demonstrated that proficiency influences some aspects of L2 self-monitoring such as self-repair foci (see Section 2.6.1 above). For this reason, proficiency will be investigated as one variable in the current study. The purpose of this section is to present definitions and measurements of proficiency.

2.8.1. Defining language proficiency

Researchers believe that definition and measurement of language proficiency has not been adequately addressed in the L2 field (e.g., Hulstijn, 2012). Some researchers define L2 proficiency as ‘learner’s language ability, which consists of language knowledge and strategic competence. Language knowledge and strategic competence allow speakers to produce and understand discourse (Bachman & Palmer, 1996). Proficiency has also been defined as ‘a person’s overall competence and ability to perform in L2’ (Thomas, 1994, p.330). Others regard language proficiency typically as the ability to use linguistic knowledge and skills in the comprehension and production of the target language (Gaillard & Tremblay, 2016). It is also defined as ‘knowledge of language and the ability to access, retrieve and use that knowledge in listening, speaking, reading and writing’ (Hulstijn, 2015, p. 21). Although a consensus on the definition of L2 proficiency has not yet been reached, these definitions indicate that proficiency refers to the linguistic ability of L2 learners.

However, linguistic knowledge has been classified as only one aspect of language competence according to the Common European Framework of Reference for Languages (CEFR). The CEFR, which is an international standard for describing language ability, states that communicative language competence includes linguistic, sociolinguistic, and pragmatic abilities. This framework has been developed to assess language proficiency in teaching and assessment. It classifies language users into three main categories: First, Basic user which is the lower category and includes two levels which are A1 (beginner) and A2 (Elementary). The second category is referred to as Independent user and includes B1 (Intermediate) and B2 (Upper intermediate). The third category entails description of Proficient user and comprises two levels: C1 (Advanced) and C2 (Proficient). The CEFR describes how learners perform in five language skills (reading, writing, spoken production, spoken interaction, and understanding language). There are further details for all five skills at each level of proficiency in terms of 'can do' statements. One great benefit of this framework is that it facilitates the assessment of learners' proficiency by providing a detailed description of language ability at different levels of development. This framework is useful in providing L2 researchers with reliable guidelines for assessing the existing language ability of L2 learners. The next section will look at measuring language proficiency.

2.8.2. Measuring language proficiency

L2 studies have recently assessed proficiency using the existing standardized proficiency tests (e.g., TOEFL, IELTS, Oxford Placement Test), the Elicited Imitation Test, C-tests or its counterpart the cloze tests. In this section, the discussion will be directed to the C-test and the Elicited Imitation Test as they are used in the current study. More discussion about these two tests will be presented in the Methodology Chapter. C-test is a language testing tool which has been designed to measure participants' general language ability (Dörnyei & Katona, 1992). This test usually consists of four-five texts in the target language to which the rule of two is applied: the second half of every second word of the second sentence is mutilated (Dörnyei & Katona,1992). The underlying principles informing the development of C-test and cloze test are based on the concept of "Unitary Competence Hypothesis" (Oller, 1979; Oller & Conrad, 1971) which proposed that the nature of language ability is a unitary and unified construct, and to test this ability it was possible to develop tools that tapped into this unitary ability. Following from Oller's (1979) work, although some research evidence was provided in support of this hypothesis, others challenged the principles of Unitary Competence Hypothesis (Canale & Swain, 1980; McNamara, 1995).

Although the use of C-test as a measure of language proficiency has been challenged, as have most other tests of proficiency, the existing evidence suggests that the C-test is a reliable measure of L2 proficiency (Babii & Ansari, 2001; Dörnyei & Katona, 1992; Richard, Daller, Malvern, Meara, Milton, & Treffers-Daller, 2009; Sigott, 2004). Its correlations with the four language skills as well as with the Oxford Placement Test (Sigott, 2004) make it a good candidate for assessing participants' proficiency levels. It also appeals to researchers and practitioners because it is practical in implementation, economical in use and administration, and objective in scoring (Dörnyei & Katona, 1992; Feldmann & Stememr, 1987; Klein-Braley, 1997; Klein-Braley & Raatz, 1984). However, this test has been criticised for not assessing language skills in the spoken modality. It has been argued that this is a serious shortcoming for studies that investigate oral modality (Tremblay, 2011).

To compensate for such a limitation, the C-test can be used alongside an aural or oral task such as the elicited imitation test which has been found to be valid in assessing oral language proficiency (Housen & Pierrard, 2005). It has been suggested that using an aural/oral test would provide a better picture of participants' proficiency (Tremblay, 2011). The EIT has been recently used in a number of studies to assess linguistic proficiency in oral modality especially after Ellis (2005) used it as part of L2 assessments. It was found that this test, along with other tasks, namely, a limited grammaticality judgment and oral narration task, load onto a factor that is termed implicit knowledge (Ellis, 2005). This suggests that the EIT taps into the implicit knowledge of L2 learners. Further discussion about the EIT can be found in Section 4.5.3. In short, this section highlights the importance of assessing L2 proficiency through reliable and valid tests. It directs attention to employing tests that are compatible with the modality that is being assessed.

2.9. Task condition

Task condition can be considered a testing ground for examining self-monitoring behaviour as self-monitoring is sensitive to contextual effects (Levelt, 1983). Understanding the impact of different contexts on the functioning of self-monitoring can provide important insights into L2 monitoring behaviour. Different aspects of task types and task condition have been employed in L1 and L2 research to explore self-monitoring (see Sections 2.4.3 and 2.6.3). Dual task condition is a preferred psycholinguistic technique for examining the influence of resource limitations on self-monitoring in L1 (see Section 2.4.3). The next section will discuss dual task condition in relation to self-monitoring.

2.9.1. Dual task paradigm and self-monitoring

It has been recommended that L2 self-monitoring studies need to use dual task condition which is helpful in exploring self-monitoring (Broos et al., 2018). Dual task condition is defined in the psychological domain as the performance of two tasks simultaneously. That is, there should be primary and secondary tasks, and once the performance of the primary task is affected by the interference of the secondary task, this shows that the process being measured is not automatic, that is, highly dependent on cognitive resources; therefore it can be said that resource limitation has an impact on performance. As self-monitoring is inherently dependent on cognitive resources, some researchers tend to use this task condition to examine the functioning of self-monitoring (see Section 2.4.3).

In most of the existing dual task experiments, primary task was operationalised along network description (e.g., Declerck & Kormos, 2012; Oomen & Postma, 2002). Network description entails asking participants to describe patterns of dotted lines which mainly focuses on colors and directions, as discussed in Section 2.4.3. Despite the practicality of its implementation along dual task condition, the network description task has been criticised for a number of reasons (see Section 2.4.3). The secondary task that is commonly used in dual task condition is finger tapping (e.g., Declerck & Kormos, 2012; Oomen & Postma,2002). That is, participants need to press one of tens keys on a keyboard, with a different finger on each key every second, as randomly as possible. The only study that employed dual task condition to examine L2 self-monitoring was conducted by Declerck and Kormos (2012). They investigated the effect of single and dual task conditions on the efficiency and accuracy of L2 monitoring. Participants in this study belong to two proficiency groups (lower and higher proficiency groups). The dual task condition in this study did not affect fluency, speed of error-detection, or the overall frequency of repairs. However, under the dual task, L2 learners made more lexical errors and corrected a smaller proportion of their overall errors. Also, the ratio of error-correction (repair divided by total number of errors) in advanced learners decreased more significantly than intermediate learners in the dual task. It has been assumed that different demands of dual task condition on different proficiency groups might show that monitoring was affected by conscious decisions taken by L2 learners whether to correct their errors or not (Declerck & Kormos, 2012).

Despite the fact that previous studies provided valuable theoretical and methodological information about L2 self-monitoring under dual task condition, some limitations of these studies can be identified. First, the use of network description in monitoring studies might lead to serious inadvertent consequences in terms of monitoring foci, as discussed previously in Section 2.4.3. Declerck and Kormos (2012) also recommended that future studies should make use of different types of speaking tasks along with dual task condition. Importantly, there is a concern that the two concurrent tasks employed in previous studies might be easy to perform as psychological research shows that performing two different tasks simultaneously is not demanding (Duncan, 1980). In other words, the different concurrent tasks might not consume great cognitive resources. For instance, it is easier to simultaneously listen to music and write rather than listen to music and listen to text at the same time. For these reasons, it has been decided that the current study needs to revise the operationalisation of dual task condition taking into consideration theoretical development in the field of simultaneous tasks. It is not an easy task to operationalise dual task condition in a way that is specifically suitable to the aim of the current study, namely, exploring self-monitoring behaviour in oral performance. The next section will discuss Wickens' (2007) model of limited-capacity multiple resourcing pools which involves theoretical guidelines for assessing similarities of the concurrent tasks.

2.9.2. Limited-capacity multiple-resources model

The limited-capacity multiple-resources model developed by Wickens, (2007) provides useful theoretical guidelines for assessing similarities between concurrent tasks according to three dimensions which have been called 'qualitative resource similarity'. According to this model, the degree of similarity between tasks moderates the extent to which the tasks depend on the same resource pools (Wickens, 2007). That is to say, when the two tasks draw on the same resource pool, it is more likely that cognitive resources will be largely consumed than when the two concurrent tasks are different. These resource pools are: First, **perceptual modality** which refers to the distinction between visual or auditory in language, namely, reading and listening. It has been assumed that it is easier to receive input that is different across modalities rather than within a certain modality, that is, it is easier to simultaneously read and listen rather than read two texts at the same time or listen to two simultaneous texts (Wickens, 2007). Second, **processing code** refers to distinction between verbal (linguistic) and non-verbal (spatial) materials. It is assumed that performing between modalities (e.g., listening and driving) is easier than performing within modalities (e.g., listening and reading) (Wickens, 2007). This also includes verbal versus spatial responses (e.g., speech vs continuous manual activity). Third, **processing stages** refers to three stages of processing which are perception, cognition, and responding as displayed in Figure 2.5 (Wickens, 2007).

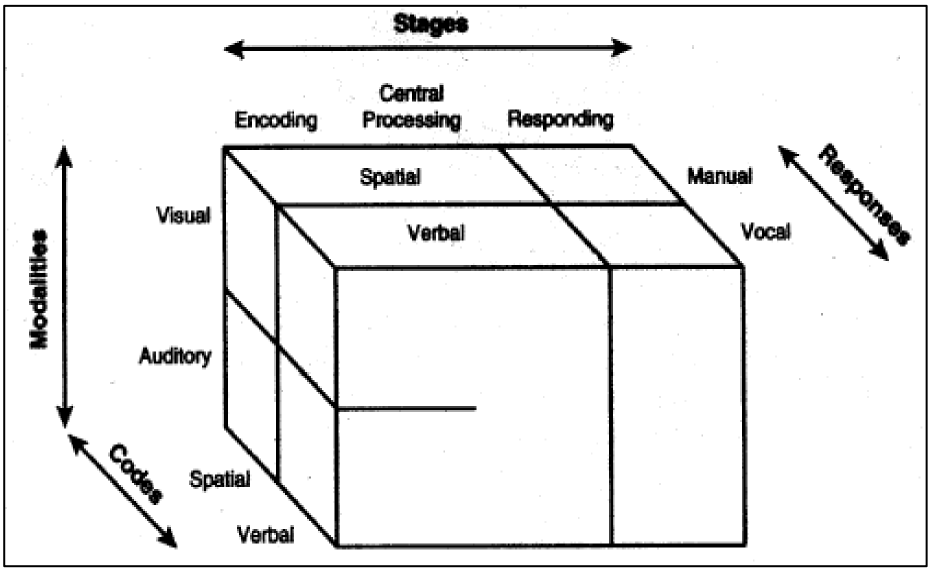


Figure 2.5. Wickens' (2007) model of multiple resources

Similarities between concurrent tasks in previous studies which used dual task condition need to be assessed according to this model. In other words, it is important to assess the similarities between finger tapping and network description to evaluate the degree to which these tasks can deplete cognitive resources in L2 performance. According to the first criterion, that is, processing modality, the input in the two tasks are dissimilar. That is, network description involves visual input whereas finger tapping instruction entails a different modality. The two tasks are also different along the second dimension (processing code) with verbal material in the network task and no-verbal modality in the finger tapping. Finally, the response is auditory in the network task and manual in the finger tapping task. It can be noted from this quick assessment that the network and finger tapping tasks are different along several dimensions. Hence, it could be assumed that the operationalisation of the dual task condition in previous studies might have influenced the results, and as such it needs to be revised. To ensure consuming the available cognitive resources while performing, it is important to theoretically operationalise the two concurrent tasks. This is a very significant step as researchers recommended that the operationalisation of variables under investigation needs to be theoretically driven, and that shortcomings of previous studies should be taken into consideration (Awwad, 2017). Further discussion about the operationalisation of the dual task condition will be provided in the Methodology Chapter.

2.10. Working memory

Measuring individual differences in working memory capacity (WMC) is a very important variable when investigating self-monitoring. According to the Perceptual Loop Theory

(PLT), monitoring is influenced by the processing constraints of working memory (Levelt, 1983). The next section will look at Baddeley's model of working memory to understand how different components function. This is followed by discussing how working memory capacity is implicated in L2 processing and L2 self-monitoring. Finally, measurement of WMC will be presented.

2.10.1. Baddeley's model of working memory

The construct of working memory is defined as a cognitive system that combines storage and attentional control, besides facilitating a range of cognitive tasks such as reasoning, learning and comprehension (Baddeley, 2003). The working memory model developed by Baddeley and Hitch (1974) involves the three components including the central executive, (a control system of limited attentional capacity), which is supported by two slave storage systems: the phonological loop (based on sound and language), and the visuospatial sketchpad (see Figure 2.6).

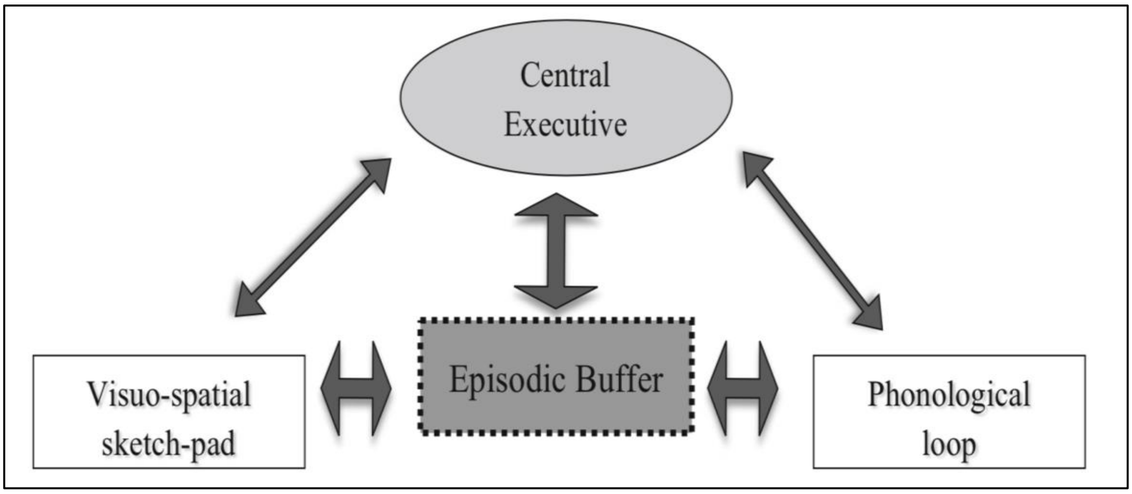


Figure 2.6. Baddeley and Hitch's multi-component WM model (Baddeley, 2012)

This model has gone through several changes over time and most of the changes involve the function and properties of the central executive. The central executive is regarded as important component in working memory although, due to its complex nature, very little is known about its functioning (Baddeley, 2012). The central executive is in control of the other two components (phonological loop and visuspatial sketchpad). It is also recognized as the link between working and long-term memory (Baddeley, 1996). Baddeley (1996, 2012) made some suggestions about the functions that the central executive needs: the ability to direct attention; the capacity to divide attention between two important targets or stimuli; the capacity to switch attention between tasks; and the capacity to interfere with long-term memory. In the older version of this model, the central executive was compared to a little person inside the head (homunculus) that can perform cognitive activities that other working memory components cannot do (Baddeley, 1996). The view of the central executive as homunculus (the component that can perform all activities) brings serious criticism to this model in terms of how it can be tested and how the theory can be developed (Baddeley, 2015). To address this shortcoming, researchers decided to remove one function of the central executive which is the storage capacity (Baddeley & Logie, 1999; Baddeley, 2000). However, once the functions of the central executive were minimised, the model seems to have another serious flaw in terms of the absence of any component for the storage. To solve this issue, a fourth component was proposed which is the Episodic buffer (Baddeley, 2000). The central executive becomes responsible for attentional control, planning, and regulating the flow of information (Gathercole, 1999). It has also been suggested that the functions of the central executive include dealing with competing stimuli and regulating their interference effects (Kane & Engle, 2003). This function can be seen when a person has to do two concurrent tasks. When the resources required for the task exceed the available processing resources at the central executive, it is expected that performance will deteriorate (Baddelely, 2012). However, research showed that the decline of performance during the concurrent tasks was less than expected (Baddelely, 2012). Perhaps the most important function of the central executive is the control of the attentional resources that are available for online cognitive processes. However, studies that examine the effect of the central executive on language learning are scarce. It has been speculated that the central executive can affect language learning through its impact on comprehension (Hambrick & Engle, 2002). Miyake and Friedman (1998) found that the central executive has beneficial effects for complex language processing. However, this area still needs further research.

The second component is the phonological loop which is the most studied and developed component of working memory (Baddeley, 2012). It is referred to as a slave system because the stimuli of the phonological loop are controlled by the central executive. It involves the storage of verbal items which can be maintained temporarily such as the telephone number. The stored items may decay in few seconds if they were not rehearsed (Baddeley, 2003). The phonological loop has the advantages of providing temporary storage and requiring minimal attentional resources (Baddeley, 2012). The phonological loop is supposed to bring information from a range of sources (verbal and non-verbal) to the episodic buffer (Baddeley, 2015). Research shows that the phonological loop is useful to language acquisition in terms of the storage of heard and spoken speech (Baddeley, 2012). It has been found that the phonological loop supports the acquisition of vocabulary in both L1 and L2 (Gathercole & Adams 1993; Papagno & Vallar 1992). Although it has been suggested that the phonological loop is helpful in learning vocabulary, it is not an essential component in such a process as there are cases of people who have impairment in Phonological Short-Term Memory (PSTM) and have been successful in learning vocabulary (Baddeley, 2015).

The visuo-spatial sketchpad is also controlled by the central executive. It constitutes the storage of visual and spatial objects. Like the phonological loop, it is limited in capacity; it can hold 3-4 items such as features relating to colour, location, and shape (Baddeley, 2003). It has been assumed that when the items held in this system belong to the same dimension, they are expected to compete for storage, whereas this is not expected to happen for items from different dimensions (Baddeley, 2003). The storage and manipulation of visuospatial features is considered a useful tool in assessing non-verbal intelligence which is required in some fields such as architecture and engineering (Baddeley, 2003). The sketchpad is assumed to be a useful language learning tool as it can facilitate the acquisition of the semantic features of new words by establishing association between visual stimuli and meanings of these words (Baddeley, 2003). A good example of such tasks is the mnemonic task. Mnemonic tasks support storage and retrieval of information by using images to represent information. That is, these tasks associate information with meaning. Episodic buffer was proposed to overcome the shortcoming of removing storage from the central executive (Baddeley, 2000). It is a storage system that is capable of combining information from the visuospatial and verbal subsystems and linking it with further information from perception and long-term memory (Baddeley, 2003). The phonological loop and the visuo-spatial sketchpad feed into the buffer, bringing information from a range of sources (Baddeley, 2015). The episodic buffer, assumed to be accessible to awareness (Baddeley, 2015), is regarded as a passive store because it can only hold episodes but cannot make any further operations on these episodes (Baddeley, 2015).

Baddeley's model was criticised by Cowan (2015), particularly because it states that each stimulus goes to a specific buffer (e.g., the phonological stimulus goes to the phonological loop). Instead, Cowan thinks that a stimulus can activate different features at the same time (Cowan, 2015). Cowan's model is an attempt to develop a unitary model of information processing. This model has a single memory repository and a main controller that provides general processing capacity (Cowan, 1999). In this model, features of incoming stimulus activate information in the long-term memory if this information has similar features to the incoming stimulus (Cowan, 1999). One of the differences between Cowan's model and Baddeley's model is that Cowan's model does not have special modules for different stimulus features such as phonological, orthographic, visual features. This means that it is a unitary model in which the phonological and visuospatial are not separated. Similar to Baddeley's model, the embedded-processes model of Cowan (1988, 1999) is capacity limited system. That is, it can hold a few items at a time.

Although Cowan's model (1988, 1999, 2001) has borrowed some of its underlying principles/concepts from the model proposed by Baddeley and Hitch (1974), more emphasis has been placed on the role of the focus of attention in creating new information. According to this model, new concepts are created as a result of the focus of attention where the linking of elements occur to form new information (Cowan, 2001, 2015). This suggests that the decay of information can be avoided if attention is paid to this information. This principle has an important implication for language learning: to learn a language, learners need to limit the focus of attention to a few items at a time (Cowan, 2001). However, Cowan (2015) argues that the main difference between the two models is the trade-off between verbal and visual items when they are processed at the same time. This is because the focus of attention can maintain only items of either type at a time. Despite the criticism against Baddeley's model, it is the most viable model and has received the most empirical support in WM research. As such, this model will be adopted as the theoretical framework for examining WM in the current study. It can be seen from reviewing Baddeley's model of working memory that two components of working memory are considered to relate to language learning and processing. The next section will discuss the relationship between working memory and L2 processing.

2.10.2. Working memory and L2 processing

The view of working memory as a limited-capacity cognitive system suggests that our consciousness can maintain a limited amount of information (Baddeley, 1996). This has an implication for language performance given that only a limited amount of data can be available for processing. The role of working memory is perhaps more important in L2 than in L1 processing due to the controlled nature of L2 processing (Fortkamp, 2000). L2 processing requires great cognitive resources because most speech components still lack automatization (Kormos, 2006). As a result, it is expected that working memory resources would be largely taxed in L2 processing, and the consumption of working memory resources is expected to have serious consequences in L2 processing (Skehan, 2015). L2 researchers have attempted to explore the effects of working memory on different aspects of L2 processing and production (e.g., Ahmadian, 2012; Georgiadou & Roehr-Brackin, 2017; Gilabert & Muoz, 2010; Guara-Tavares, 2008; Kormos & Sáfár, 2008; Mojavezi & Ahmadian, 2014; William, 2012). However, L2 research in this area yield inconsistent results in terms of the effects of working memory capacity (WMC) on different aspects of processing and production.

Researchers argued that the involvement of working memory in language processing relates to the degree of automatization of processes (Harrington, 1992; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Wright, 2015). In other words, controlled processes such as self-monitoring are likely to be more dependent on working memory resources during their functioning, whereas automatized processes are less dependent on working memory resources. Researchers also assume that in the case of L2 processing, WMC is likely associated with levels of proficiency (Harrington, 1992; Harrington & Sawyer, 1992; Miyake & Friedman, 1998). In other words, some researchers believe that WMC is likely to vary in the course of L2 acquisition due to the increased command of language and the gradual automatization of the linguistic system (Harrington, 1992; Harrington & Sawyer, 1992; Miyake & Friedman, 1998). Firandi and Weissheimer (2009) found that the relationship was stronger between WMC and L2 processes in the lower level proficiency than in advanced learners in language use and acquisition. They concluded that limitations in working memory are closely related to the consciously controlled processes rather than automatic processes (Firandi & Weissheimer, 2009). Furthermore, it was argued that working memory resources can be implicated in cognitively demanding tasks (Li, Ellis, & Zhu, 2019; Robinson, 2007). This suggests that in situations where cognitive resources are largely depleted, extra working memory resources are likely to be beneficial to L2 processing. However, this assumption has been contradicted by other researchers who argue that the cognitively demanding task would have a negative impact on working memory functioning (Michel, Kormos, Brunfaut, Ratajczak, 2019). Thus, it is still not clear in what ways working memory can influence language processing during the performance of the demanding tasks.

Although some of the assumptions mentioned above about the association between WMC and L2 processing are theoretically plausible, there is still not enough empirical evidence to validate such postulations. In fact, the relationship between working memory and L2 processing is controversial. Researchers from the fields of psychology and neuropsychology argue that certain components of working memory have their own resources (e.g., the phonological loop) which are separable from the resources of the central executive (Gathercole & Baddeley, 1993). Furthermore, it has been claimed that some speech processes (e.g., assigning syntactic structure to sentence) have their own resources that do not place demand on the general working memory resources usually measured by working memory tests (Caplan & Waters, 1999). This shows that there is some disagreement about the role of working memory in L2 processing, and therefore, any interpretation of data analysis in this context needs to be cautious. The next section will look at the association between working memory and L2 self-monitoring.

2.10.3. Working memory and L2 self-monitoring

Self-monitoring is a consciously controlled process that draws on the limited-capacity cognitive system of working memory (Levelt, 1989). However, it is still not clear in what ways or what components of working memory are involved in self-monitoring. What is known so far about the association between working memory and self-monitoring can be summarized as follows:

1- Error-detection and correction involves consciously controlled processes. According to the PLT, when errors or inappropriateness are detected, the monitor sends signal to the working memory, and then the working memory issues instructions for corrections (see Section 2.4). This means that the functions of self-monitoring draws on working memory resources.

2- Due to the limited capacity of working memory, self-monitoring can potentially deal with few items of internal or external speech at a time. This means that L2 speakers might prioritize certain aspects of their speech during the process of monitoring (items in the working memory). As such, self-monitoring is considered context sensitive (Levelt, 1989), which means that the foci of monitoring are likely to be influenced by contextual effects during monitoring (see Section 2.6.2).

3- At the Conceptualiser level, the messages usually ‘pass through’ the working memory, and they remain available for some time for comparison with the actual speech output (Levelt, 1983, p. 47). This suggests that larger working memory storage could assist monitoring during its matching function by holding some items for a short time (see Section 2.4).

Although it has been advocated that L2 self-monitoring studies need to investigate the association between monitoring aspects and working memory (Kormos, 2006), only few studies have attempted to do so (e.g., Georgiadou & Roehr-Brackin, 2016; Mojavezi & Ahmadian, 2014). Mojavezi and Ahmadian (2014) investigated the relationship between WMC and self-repair behaviour in L1 and L2 oral production. Participants were 40 intermediate learners of English; their L1 was Farsi. Silent video cartoons were used to elicit the speech. A working memory listening span test was used to measure WMC where L2 learners were given sets of sentences in L1 Farsi. They were also required to write down the last word in each sentence of the set. The results showed a positive significant correlation between WMC and A-repair in L1. There was also a negative significant correlation between WMC and E-repair in L1. These results were interpreted as signifying that L1 speakers are sensitive to social norms and appropriateness of their messages, that is, L1 speakers with larger WMC tend to allocate their working memory resources to attend to the appropriateness of their message, thus more A-repair was produced. For E-repair, it was assumed that as the processes of L1 speech proceed automatically, the need for making E-repair decreased. Concerning the relationship between WMC and L2, it was found that there was a negative significant correlation with D-repair and a positive significant correlation with E-repair. It was concluded that participants with larger WMC are more likely to allocate their working memory resources to focus on form so that more E-repair is produced (Mojavezi & Ahmadian, 2014). The significance of this study is that it was the first L2 study that examined the relationship between WMC and different types of self-repair in L1 and L2 speech.

Georgiadou and Roehr-Brackin (2017) examined the relationship between executive working memory and phonological short-term memory (PSTM) on the one hand, and fluency (speech rate, pauses, and repetition) and self-repair behaviour on the other. The participants were seventy-seven L1 Arabic speakers learning L2 English at two levels of proficiency (lower and intermediate). This study employed interviews to elicit self-repair. It used a word recall test in L2 English (WRT) to measure PSTM, and a backward digit span test (BDST) in L1 Arabic and a listening span test (LST) in L2 English to measure learners' executive working memory. The findings showed a significant negative correlation between working memory and the number of pauses in the intermediate participants as measured by BDST, suggesting that the intermediate participants with greater executive working memory capacity paused less. It has been assumed that larger WMC assists intermediate learners during the

Conceptualiser and Formulator levels (Georgiadou & Roehr-Brackin, 2017). Data also showed no significant relationships between any of working memory measures and repetition, speech rate, or the overall number of self-repairs made by the two proficiency groups. It seems that self-repair draws on factors other than working memory, particularly in lower proficient learners who have various reasons for making repair rather than merely correcting errors (Georgiadou & Roehr-Brackin, 2017). It should be noted, however, that the findings of these two studies are inconsistent. The influences of working memory vary according to task types, learners age, and the aspects of language ability examined (Juffs & Harrington, 2011). It can also be seen that these studies focus on certain aspects of monitoring (i.e. self-repair and pauses), thus future studies need to look at other aspects of monitoring such as temporal phases of repair, accuracy, and disfluency which might provide important insights into the association between working memory and the functioning of self-monitoring.

2.10.4. Measurement of working memory capacity

The capacity to hold and manipulate resources has been used to predict the way working memory, as a cognitive system, operates in relation to L2 processing and production. Temporary storage and processing are believed to relate to the two most important components of working memory which are phonological loop and the central executive (see Section 2.10.1). It has been argued that these two components need to be carefully distinguished to ensure reliable assessment of their involvement in different aspects of L2 speech (Wright, 2015). Within the fields of SLA and cognitive psychology, there is still an ongoing debate and lack of agreement in terms of the reliability and validity of the current tools employed to assess WMC (Mitchell, Jarvis, O'Malley, & Konstantinova, 2015). However, it needs to be asserted that when intending to examine the extent to which each component of working memory plays a different role in L2 processing, the test should be compatible with the theoretical underpinnings of the component under investigation in order to avoid serious consequences, especially during the interpretation of the data. For instance, when assessing the central executive of working memory, researchers tend to employ certain tools (e.g., Backward Digit Span Test) that can measure both storage and attentional control (e.g., Georgiadou & Roehr-Brackin, 2017). According to the update of Baddeley's model (see Section 2.10.1), it has been clarified that the central executive was freed from its function as responsible for storage, and this function was assigned to a new component which is the Episodic buffer. Thus, the claim that BDST can assess only the central executive is not theoretically supported.

Working memory tests currently employed in L2 research cannot be considered clear-cut measures of working memory or its different components. These tests will benefit from developments in other fields such as cognitive psychology. Although many L2 studies assess WMC as a unitary construct (e.g., Ahmadian, 2012; Awwad, 2017; Kormos & Sáfár, 2008), there is an ongoing debate which highlights the need to distinguish between the different components of WM such as phonological short-term memory (PSTM) and executive working memory (EWM), and how they relate to various features of language ability (e.g., Mitchell, Jarvis, O'Malley, & Konstantinova, 2015). It has also been argued that WM components may be implicated differently in various stages of language proficiency (Mitchell et al., 2015). For example, Juffs and Hanrrington (2011) suggest that there are great effects of the phonological memory in the lower proficiency learners. The executive memory, on the other hand, is suggested to affect higher level learners (Mitchel at al., 2015). This discussion provides a good introduction into measuring the different components of WM rather than treating it as a unitary construct. Currently, many L2 studies assess WM as a unitary construct (e.g., Ahmadian, 2012; Awwad, 2017; Kormos & Sáfár, 2008; Mojavezi & Ahmadian, 2014) which may result in limitations in the research design and generalisability of the findings (Mitchel et al., 2015). However, further research is still needed to produce valid and reliable WM tests that can tap into different components of WM (domain-specificity). As such, the current research will follow the bulk of L2 studies which employ WM tests that relate to domain-generality of the WM measurement.

A number of suggestions have been provided for researchers who intend to include working memory in their investigation. It has been recommended that working memory tests need to correspond to the modes of the construct being examined (Grabowski, 2010). That is, to examine WMC in speaking tasks, researchers need to use speaking or listening tests (Cho, 2017). Furthermore, to avoid assessing other constructs (e.g., L2 proficiency) along with working memory, some researchers have recommended using language-independent measures; or measures that are based on the L1 (Georgiadou & Roehr-Brackin, 2017). A range of working memory measurements have been recommended for use in SLA studies, e.g., the counting span, operation span, and reading span tasks, and these are among the most commonly used tools in cognitive psychology (Conway, Kane, Bunting, Hambrick, Wilhelm,

& Engle, 2005). The speaking span task and listening span task have also been suggested by researchers in cognitive psychology (e.g., Finardi & Weissheimer, 2009; Goo, 2010). Researchers used the digit span tests which include: Forward Digit Span and Backward Digit Span (e.g., Kormos & Sáfár, 2008). The backward digit span and the reading and listening span tasks are frequently regarded as instruments testing more than just phonological short-term memory: they are claimed to assess the capacity of complex verbal working memory, including the functioning of the central-executive, which is responsible for regulating attention (Gathercole, 1999; Hale, Hoeppner & Fiorello, 2002). BDST is believed to assess temporary storage as well as manipulation of stored information (Kormos & Sáfár, 2008).

In short, assessing WMC in L2 research is not an easy task. Besides the open debate about the relationship between working memory and L2 processing (Section 2.9.3), there is also ongoing controversy about assessing different components of working memory and how learners' age and speaking tasks can affect the assessment. All of these factors need to be taken into consideration when assessing and interpreting the findings of research about the relationship between working memory and L2 processing and production.

2.11. The existing research gaps

The purpose of this section is to highlight gaps that have been identified in previous literature. The section is divided to two parts: the first part discusses the theoretical issues that have been identified in the previous literature on L2 self-monitoring, and the second part presents the methodological issues in examining L2 self-monitoring.

2.11.1. Theoretical issues

Despite the fact that many theoretical assumptions of monitoring theories, such as the Perceptual Loop Theory (PLT), have been empirically validated in the field of psycholinguistics and L1 research, testing monitoring theories in the L2 context still needs further research. In fact, most of the current L2 studies have focused on examining the effects of different task characteristics or task conditions on self-repair, and only few studies have attempted to test the assumption of monitoring theories in the L2 context or have related the findings to these theories. Researchers identified proficiency level as an important area for investigating the extent to which self-monitoring can change as a result of the development of the linguistic system as discussed in Section 2.7 and Section 2.8. In fact, there is still a need for further studies to empirically test various assumptions of the PLT on L2 speakers at different levels of proficiency.

One of the main assumptions that needs to be examined in the L2 context is the extent to which resource limitations can influence L2 self-monitoring as L2 speech processes are already consumed due to lack of automatization as discussed in Section 2.7. Although it has been assumed that self-monitoring is dependent on cognitive resources, it is still not clear to what extent resource limitations can affect monitoring functions. It has been indicated that a complex experiment (such as auditory distractor) may make self-monitoring processes 'more intense' (Levelt et al., 1999, p. 6). It is not clear what is meant by 'intense' in this context; it probably means that the auditory loop of the monitor could be more active in complex conditions. In addition, despite the fact that monitoring is dependent on working memory resources, there are mixed results as well as gaps in the literature in terms of the relationship between different aspects of self-monitoring and working memory. As such, this section indicates the need for further investigation of self-monitoring in relation to proficiency level, task condition, and working memory. Also, L2 research mainly focus on self-repair when examining self-monitoring. However, other aspects of monitoring need to be examined to ensure studying this construct from different perspective (e.g., disfluency, different types of repair, temporal phases of repair, and ratio of error-correction). It is also useful to examine accuracy of speech production as a way of tracking the development of the sub-processes of the Formulator and the extent to which its development can affect L2 self-monitoring behaviour.

2.11.2. Methodological issues

Although dual task condition is a preferred technique in examining the impact of resource limitations on self-monitoring, it has rarely been implemented in L2 research. The only L2 study that employed dual-task condition operationalized this condition in a way that was not compatible with the theoretical considerations of concurrent tasks (Section 2.8.2). The network description task which is commonly used to investigate self-repair has been criticized for directing learners' attention towards one focus of monitoring, namely, lexical errors. Performing network description task might not resemble real life contexts in requiring speakers to correct every single error they make, and therefore this task might not clearly depict foci of monitoring in real life contexts.

Furthermore, dual task condition as used in previous research was not carefully operationalised. That is, the concurrent task in previous studies did not follow theoretical considerations of tasks similarities that can lead to significant consumption of cognitive resources. It has been recommended that the operationalisation of variables under investigation needs to be theoretically driven, and that shortcomings of previous studies need to be taken into account (Awwad, 2017). As such, this study intends to operationalise the dual task condition following the guidelines proposed by Wickens (2007) for assessing the degrees of similarities between concurrent tasks.

Despite the fact that proficiency is the most important factor when studying the changes in L2 processing, there are issues in assessing this construct. More comprehensive measurement needs to be used when examining this construct. Besides the importance of using reliable and valid tests, it is also important to employ tests that are compatible with the modality that is being assessed in the research. The Methodology chapter will provide a full discussion of the tests which are used in the current study, namely, the C-test and the Elicited Imitation Test.

2.12. Conclusion

This chapter has presented a review of the key L1 and L2 speech models, followed by discussion and evaluation of the existing self-monitoring theories. The chapter has also presented the construct of self-monitoring and the theoretical and methodological factors considered in its study. Automaticity has been discussed in relation to self-monitoring. A full discussion of three important factors which influence self-monitoring behaviour has been provided; these factors are proficiency level, task condition, and working memory capacity. The last section has presented the theoretical and methodological gaps that have been identified when reviewing the literature.

Chapter 3: THE PILOT STUDY

3.1. Introduction

This chapter presents the pilot study which was conducted mainly to design and test the materials of the study. The chapter starts by presenting the aims of the pilot study. This is followed by a detailed description of the procedures of designing picture prompts and dual task condition. Further explanation is provided about the recruitment of participants and the data collection. The results are discussed in relation to the theoretical principles and previous studies. This chapter concludes by discussing the limitations of the pilot study that need to be considered in the main study.

3.2. Aims of the pilot study

The pilot study is conducted to serve two important purposes. First, to develop two oral narrative picture prompts which are similar in terms of complexity, number of elements, propositions, vocabulary and structures. Piloting is considered an essential step for developing picture prompts which are not selected from previous literature (De Jong & Vercelloti, 2016). The second aim of the pilot study is to trial the operationalisation of the dual task condition (for further discussion, see Section 3.4.2). As the current study aims to examine the increased cognitive demand of task condition, manipulated along dual task condition, on L2 self-monitoring behaviour, it is important to test the operationalisation of dual task condition before using it in the main study. The current study introduces a novel operationalisation of the dual task condition (see Section 3.4.2). As such, it is important to pilot dual task condition to assess the extent to which it can impose cognitive demand on L2 self-monitoring behaviour.

3.3. Participants

Twelve native speakers of Arabic participated in the pilot study. They were L2 learners enrolled at the Department of Foreign Languages in Taif University, Saudi Arabia. Their age ranged from 19–23 years. The C-test was administered to assess learners' proficiency levels. As the pilot stage intended to focus on participants from the same proficiency level, twelve participants who scored between 61 and 80 points out of 100 in the C-test were selected. Aligning these scores with the Common European Framework of Reference (CEFR) (Council of Europe, 2001), their level of proficiency was at the B2 level, which was considered as an upper intermediate level of L2 proficiency.

3.4. **Tasks and Materials**

The first part of this section presents the procedure of designing two similar picture stories following the suggestions of De Jong and Vercelloti (2016). Using picture prompts that were not similar would result in different speech features (De Jong & Vercelloti, 2016). As such, it was decided to design oral narrative picture prompts comparable along several dimensions so that they would have similar effect on the elicited speech (Section 3.4.1). The second part relates to designing dual task condition following the guidelines of tasks similarity proposed by Wickens (2007) (see Section 3.4.2). This is particularly important because the degree of similarity between simultaneous tasks determines the extent to which dual task condition can consume available cognitive resources (Wickens, 2007). If dual task condition is not systematically designed, there would be negative consequences during data collection and the interpretation of findings.

3.4.1. **Picture prompts**

The current study attempts to avoid shortcomings identified in previous monitoring research in terms of the task used to elicit speech, namely, the network description task. This task was criticised for prompting learners to focus on a particular aspect of speech, namely, the lexical aspect (see Section 2.4.3). L2 research shows that oral narrative picture prompts 'are regarded as valid pedagogic tasks which are frequently used across different educational settings for teaching, learning, and assessment purposes' (Tavakoli, 2011, p. 73). This suggests that describing a picture story resembles real life conditions and probably involves cognitive demand similar to what is experienced in real life contexts. Picture prompts are also easily administered and practical in eliciting spontaneous speech (Van Hest, 1996). As such, it was decided to use oral narrative picture prompts to elicit L2 learners' oral performance in the current study. To avoid repetition effects, it has been suggested that researchers need to use a different picture prompt for each task condition (Ishikawa, 2007; De Jong & Vercelloti, 2016). However, it has been argued that a single prompt for each task condition could result in a confounding effect as the prompts might not elicit similar speech samples (De Jong & Vercelloti, 2016). Some researchers argue that seemingly similar tasks differed in the language they elicited (De Jong & Vercelloti, 2016). It has been suggested that picture prompts need to be comparable along different task features including structure, storyline complexity, and number of elements to control for task effects on elicited speech samples (De Jong & Vercelloti, 2016).

Following suggestions of previous studies, it was decided to employ two comparable oral narrative picture prompts to elicit speech samples in the current study. Efforts were taken to find two similar picture prompts in previous literature and comics websites. Unfortunately, picture prompts that are similar across several dimensions could not be found. As such, the decision was taken to create two original prompts. To do this, a certain procedure was conducted. First, two comparable stories were made up. Efforts were taken to ensure that these stories were similar in structure, characters, storyline complexity and the number of elements. Six-frame prompts were designed for the two picture prompts. A series of small pre-pilots were conducted to develop these picture stories as suggested in previous literature (De Jong & Vercelloti, 2016).

Scenarios of the two stories were created, taking into consideration comparability of sequence of events (i.e. beginning, climax, and happy end). The first story, the Ship, narrates a story of a group of friends who went on a journey by ship. Although the weather was nice at the beginning, a storm arose and got stronger. Later, people managed to reach an island and stayed there for a while. Darkness was everywhere so that a fire was lit. After that, a helicopter approached them. The last picture shows that they were safe and celebrating. The second story (the Fire) describes a scene where children came to school in the morning. They were in their class talking with their teacher in a computer lab. Everything seemed quiet but later a fire started in the classroom and then spread through the school. Luckily, the teachers helped the students to escape the class. The fire engine and ambulance arrived. Finally, they were safe and with their parents. Considerable efforts were put into ensuring that the stories would not bring up bad feelings for the participants. For example, it was decided that the stories should avoid any incident of people drowning or being affected by the fire.

To ensure that vocabulary of the two stories belonged to the same level of complexity, the written scripts of the stories were submitted to VocabProfilers (Cobb, 2017). This website divides the words of texts according to lexical frequency into first and second thousand levels (1K, 2K) which are more familiar than academic and off-list words. This website is based on the work of Laufer and Nations (1995). This step was conducted before piloting the tasks. VocabProfilers showed that 93.75% and 96.07% of vocabulary in the Fire and the Ship respectively belong to 1K and 2K levels (see Appendix 3). The results suggested that the two tasks elicited comparable vocabulary in terms of their frequency, and therefore the tasks were also comparable in terms of the vocabulary items they elicited.

After making sure that the structure of the two stories, and vocabulary were similar, the next step was to design the picture prompts. For this purpose, PowerPoint software and Paint X Premium were used to manipulate shapes and clipart to design pictures. Tens of clipart shapes were combined to create each picture story (see Appendix 4). One further step was taken to ensure comparability of the tasks which was the number of elements in the two picture prompts. That is, there is a need to control the number of items in each frame of the picture prompts, e.g., buildings, persons, and furniture items. (see Appendix 4). These picture stories were pre-piloted several times, and the participants' perceptions about similarity of stories were considered. Further modifications to the picture prompts were implemented accordingly. Therefore, it was concluded that the two picture stories were similar and would elicit similar speech samples. Finally, instructions were added in Arabic and English to the two picture prompts (see Appendix 5).

3.4.2. **Dual task condition**

Section 2.9 discussed the importance of the dual task condition being systematically operationalised, and that the operationalisation needs to draw on theoretical principles. As such, it was decided that the dual task condition in this study would be designed following the guidelines provided by Wickens (2007) for assessing the similarity of the concurrent tasks. The dual task condition involves primary and secondary tasks. The picture prompts (described above) is the primary task in the current study. It can be noted that performing the picture prompts involve the three dimensions described by Wickens (2007). First, the modality of the picture stories is visual. That is, these prompts present visual rather than auditory input to participants. Second, the processing code of these prompts is verbal which means that participants process linguistic rather than spatial data. Finally, the processing stages of the picture stories involve perception, cognition, and responding. That is, participants interact with the task by observing, thinking about it, and responding to it.

The secondary task needs to include the same three dimensions described above to ensure that cognitive resources are consumed by the dual task condition. That is, the secondary task needs to be visual, verbal, involving perception and cognition, and responding. According to Wickens (2007), if the two concurrent tasks are similar, they would draw on the same resource pools, and therefore they would consume available cognitive resources. Secondary tasks in previous research employing the dual task methodology have been criticised for not reflecting real life contexts (Révész, Michel, & Gilabert, 2016; Révész, Sachs, & Hama, 2014). Secondary tasks that were used in previous studies include a finger tapping task (e.g., Declerck & Kormos, 2012), a background colour changing task (e.g., Révész, Michel, & Gilabert, 2016), and a letter colour detection task (e.g., Sasayama, 2016). It has been recommended that future studies need to employ different forms of secondary tasks 'that are more likely to be encountered in real-life language use situations' (Révész, Michel, & Gilabert, 2016, p. 735). The current study attempts to avoid such a limitation by using two concurrent verbal tasks that are likely to be encountered in real-life situations. In daily life, it is common to find people performing two simultaneous verbal tasks such as speaking and typing text messages on their mobile phones or listening to a lecture and taking notes.

3.4.3. **E-Prime Psychological Software**

To design and run the dual task experiment, E-Prime Psychological Software (3.0) was used (Schneider, Eschman, & Zuccolotto, 2002). E-prime is suitable for computerised experimental design, because it handles milliseconds precision timing efficiently while a task is being administrated. The secondary task involves bubbles appearing respectively on a computer screen simultaneously along with the picture stories. It was not possible to have the pictures static in E-Prime while the bubbles are popping on, thus several copies of the same picture were made, and a bubble was pasted on different locations in each slide (see Appendix 6). The slides were set to be shown randomly. A bubble appears every 5 seconds on the screen. Each bubble stays only for 5 seconds and then disappears if no response is made by participants. Each bubble contains a word naming either an animate or inanimate object (see Appendix 6). Participants were instructed to press a certain button when they decided that the word described animate objects such as animals. A different button needed to be pressed when a word described inanimate objects such as classroom objects. Two keyboard buttons (Z and M) were marked with Arabic translations of 'animate' and 'inanimate' which are (حي) and (جماد) respectively to make it easy for participants to focus on the experiment at hand rather than trying to find the right button to press (Albarqi, 2018). This step would reduce the amount of attention given to finding the right keyboard buttons. The instructions were presented in Arabic, and participants were given one minute to read. Once they were ready, they could press the start button to begin the experiment. It can be noted that the secondary task in the current study involves most of the dimensions of the primary task; the only exception is that the dimension of responding in the primary task is verbal whereas it is manual in the secondary task.

The secondary task consisted of 24 trials: 4 practice trials which involve narrating a different picture story and simultaneously performing the secondary task. The aim of the practice trials is to help participants understand the experiment and reduce their anxiety. To avoid repetition effects, a different picture story was used in the practice trials (see Appendix 6-B). The main part of the secondary task includes 20 trials in which participants were required to narrate one of the picture stories (described in Section 3.4.1) and simultaneously press a button that corresponded to words showing on the screen (describing animate or inanimate objects). Single task condition entails narrating a picture story, Ship or Fire (Appendix 5), which is presented using Microsoft PowerPoint program on a laptop screen. To reduce any potential differences, the picture stories were counterbalanced in the two task conditions (single and dual conditions). These two tasks (primary and secondary tasks) were designed specifically for the purpose of the present study and they were pre-piloted several times before their use in this pilot study.

3.5. **Ethical Procedures**

The present project followed Reading University's Ethics Guidance. The ethical procedure was reviewed and approved by the School Ethics Committee. Participants signed a consent form and an information sheet which had a description of the study (see Appendices 1 and 2). Each participant was given a copy of the signed consent form and another copy was kept with the researcher. The researcher told the participants that their privacy and confidentiality will be protected and that they are free to withdraw from the study at any time.

3.6. **Data Collection**

The pilot study started by obtaining approval to conduct the experiment at the Department of Foreign Languages in Taif University during a summer term. The researcher met the students and explained the general aims of the project. After that, the C-test was administered to eighty students, and only those who completed all the passages of the test were selected to participate in the study (i.e. 40 participants). Aligning the C-test with the CEFR (the Common European Framework of Reference for Languages), three proficiency levels were identified which are A2, B1, B2. It was decided that participants at B2 level would be selected for the pilot study (n = 14) because this group had a greater number of participants than the other groups (A2, B1). However, only 12 participants came to the next session where the researcher met each student individually and explained the instructions of the two task conditions. Half of participants (six students) performed in the single task condition and the other half in the dual task condition. They were given certificates of appreciation for participating in this study. There were a number of issues encountered during the pilot study, among which was the limited time available for the researcher to meet students during the summer term courses; thus, only one proficiency level (B2) was selected.

3.7. Study Design

The current pilot study employed a 2x2 between-subject design which entailed half of the participants performing in the single task condition and the other half performing in the dual task condition (see Table 3.3 below).

Table 3.3. The study design of the Pilot Study

Study Design	Independent Variables	Dependent Variables
Between-subject N = 12	Task condition (Single vs Dual) Picture stories (Fire vs Ship)	Self-monitoring aspects: 1. Pauses 2. Repair types 3. Temporal phases of repair

In addition, half of the participants narrated the Ship story whereas the other half narrated the Fire story because the current study aimed to establish similarity between these two picture prompts. The two picture stories were counterbalanced in the two task conditions. This means that there are two independent variables with two levels for each, which are task conditions (single and dual), and picture stories (Ship and Fire). Dependent variables include three aspects of monitoring: pauses, repair types, and temporal phases of repair. Section 3.8 below presents further details about these aspects of monitoring.

3.8. Data Coding

Twenty-four speech samples were transcribed using Goldwave software. All speech samples were segmented into AS-units following Foster, Tonkyn, and Wigglesworth (2000). Then, data were coded for measures of repairs. PRAAT was used to measure pauses and temporal phases of repair in milliseconds (Boersma & Weenink, 2008). Repair types were calculated in terms of a number of repairs per minute (see Table 3.2 below). To check interrater reliability, 10% of the data was coded by an expert rater and a high percentage of consistency (92%) was obtained among raters. It was then concluded that a reasonable level of interrater reliability was achieved.

3.8.1. **Measures of pauses**

The decision to use the measures of pauses was based on one principle of the PLT which states that disfluency is regarded as a corrective measure to anticipated errors (Levelt, 1983). In addition, this study employed both mid-clause and end-clause pauses in an attempt to explore how different types of pauses are affected by task demand. L1 and L2 studies show that pausing in the middle or at the end of clause are closely linked to monitoring (Ahmadian & Tavakoli, 2014; Ahmadian et al., 2012; De Jong et al., 2012; Kircher, Brammar, Levelt, Bartles, & McGuire, 2004; Levelt, 1983; Michel, 2011; Tavakoli, 2011, 2016; Witton-Davies, 2014). The measures of pauses were calculated per 60 seconds and the threshold for a pause was set at > 0.25 seconds following De Jong & Bosker (2013) who suggest that 0.25 seconds is a good threshold for silent pauses in L2 research.

Table 3.4. Measures of self-monitoring aspects in the Pilot Study

Dimension	Measure	Definition
Pauses	Number of mid-clause pauses	The total number of mid-clause silent pauses 0.25 second divided by the total time of speech in seconds.
	Mean length of mid-clause pauses	The total duration of mid-clause pauses divided by the total time of speech in seconds multiplied by 60.
	Number of end-clause pauses	The total number of end-clause silent pauses 0.25 second divided by the total time of speech in seconds.
	Mean length of end-clause pauses	The total duration of end-clause pauses divided by the total time of speech in seconds multiplied by 60.
Repair types	Error-repair	The total number of repairs that are made to correct linguistic errors (lexical, grammatical, and phonological) divided by the total time of speech in seconds.
	Message-repair	The total number of repairs that are made to correct inappropriateness in the message divided by the total time of speech in seconds.
Temporal phases of repair	Error-to-Cut-off	The total length of Error-to-cut-off in milliseconds divided by the total number of repairs.
	Cut-off-to-repair	The total length of Cut-off-to-repair in milliseconds divided by the total number of repairs.
	Repair	The total length of Repair in milliseconds divided by the total number of repairs.

3.8.2. **Measures of repair types**

As this pilot study was mainly conducted to obtain a generic view of the influence of dual task condition on self-monitoring, and for a practical reason, different types of repairs were grouped into two main categories which include:

1- Message-repair which involves modifying an utterance due to inappropriateness. It also includes abandoning the message and producing a different one.

2- Error-repair involves lexical, grammatical, or phonological repair. It also includes tenses, preposition, determiner, and inflections.

3.8.3. **Measures of temporal phases of repair**

Temporal phases of repair entail the duration of different phases of L2 repairs as demonstrated in Figure 3.7. These phases are Error-to-cut-off, Cut-off-to-repair, and repair (see Section 2.4.2).

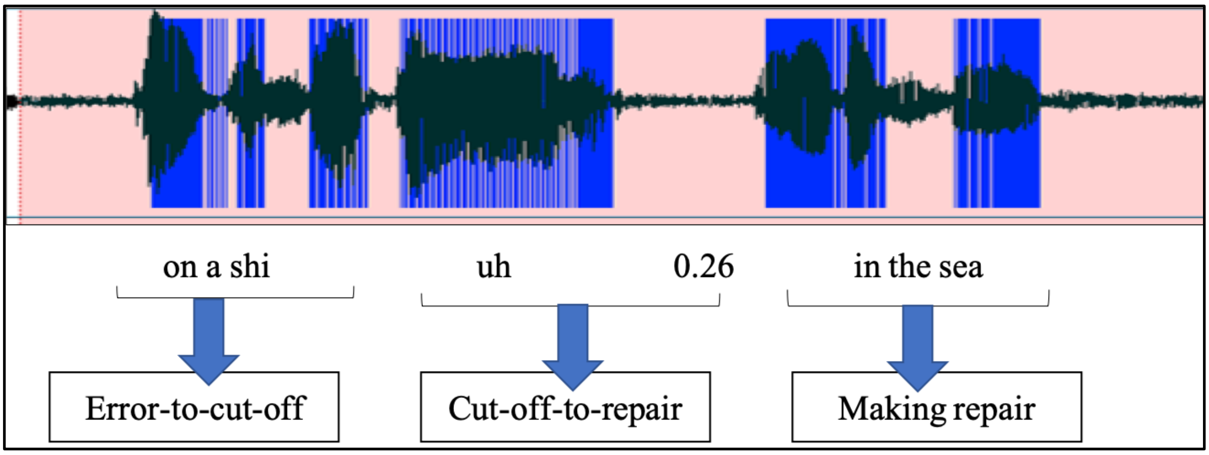


Figure 3.7. Temporal phases of repair

These measures involve the durations of different stages of repair:

1. Error-to-cut-off (first phase) entails the distance from the onset of error or inappropriate item to the offset of error or inappropriate item.
2. Cut-off-to-repair (second phase) shows the duration from the offset of error to the onset of repair.
3. Repair (third phase) involves the duration of correction or repair.

The rationale for including this category of monitoring draws on the assumption that timing of repair phases is helpful in making inferences about how L2 self-monitoring mechanisms operate in different task conditions (Kormos, 2006). Temporal repair phases have rarely been investigated in L2 research, and to the best of my knowledge only Declerck and Kormos (2012) examined the three main phases of repair under dual task condition. In this pilot study, the temporal phases of the two types of repair were combined. For example, Error-to-cut-off represents the first phase of repair in both Message-repair and Error-repair.

3.9. Data Analysis

Descriptive analysis (means and standard deviations) was first run for measures of pauses, repair types, and temporal phases of repair to provide first impressions about the differences in self-monitoring aspects in the two picture stories and the two task conditions. Then, a series of independent t-tests was run to examine the effect of picture prompts and task condition on measures of pauses and repair. It is worth mentioning that in this pilot study, there are many dependent variables which means that the statistically significant differences may occur due to error or chance. However, it is only a pilot study and this point is carefully considered for the main study. Data are presented in tables 3.5 and 3.6 below.

Table 3.5. Self-monitoring aspects in the two picture stories (the Pilot)

Dimension	Measures	Picture Stories				
		Ship		Fire		Sig. (two-tailed)
		M	SD	M	SD	p
Pauses	Mid-clause pauses number	14.11	3.27	12.60	4.75	.37
	Mid-clause pauses length	.24	.10	.16	.07	.048
	End-clause pauses number	8.51	2.79	7.52	2.16	.34
	End-clause pauses length	.15	.06	.15	.10	.89
Repair types	Global-repair	.19	.37	.03	.019	.15
	Message-repair	.021	.019	.018	.016	.56
	Error-repair	.021	.019	.01	.015	.12
Temporal phases of repair	Error-to-cut-off	3.13	2.46	1.54	.82	.05
	Cut-off-to-repair	1.96	1.95	1.44	1.22	.44
	Repair	2.62	2.76	1.89	1.01	.39

*p < 0.05

3.10. **Results**

3.10.1. **Picture story**

It was expected that the two picture stories (Ship and Fire) would elicit similar speech features (pauses, repair types, and temporal phases of repair). Descriptive data in Table 3.5 above indicates that all measures of pauses, repair and temporal phases of repair increased in the Ship prompt rather than the Fire prompt. The increased number of repairs and pauses suggest that the task may have been demanding. The results showed that there were some significant differences between the two picture stories. Length of mid-clause pauses increased more significantly in the Ship (M = .24, SD = .10) than the Fire (M = .16, SD = .07); t(22) = 2.09, p = 0.048. Likewise, Error-to-cut-off duration increased more significantly in the Ship (M = 3.13, SD = 2.46) than the Fire (M = 1.54, SD = .82); t(13.45) = 2.12, p = 0.05. The data suggests that the Ship was more demanding than the Fire so that L2 learners made longer pauses in the Ship than the Fire. They also took longer time before they interrupted their speech when detecting an error or inappropriateness in the Ship story which suggests that they encountered certain difficulty in the Ship as compared to the Fire story. Therefore, it can be concluded that these two picture prompts are not similar.

3.10.2. **Task condition**

It was expected that the dual task condition would impose greater demand on the cognitive resources than the single task condition. Thus, a greater number and length of pauses would be produced in the dual task than the single task conditions. It was also predicted that the dual task condition would elicit less repair and longer intervals of repair phases than the single task condition. Descriptive statistics in Table 3.6 below indicate that all measures of self-monitoring (pauses, repair, temporal phases of repair) increased in the dual task condition as compared to the single task condition, with the exception of silent pauses at clause boundaries. This suggests that L2 learners might find the dual task condition more demanding than the single task condition. It is, however, surprising that repair types (Message-repair and Error-repair) increased rather than decreased in the dual task condition rather than the single task condition. The data presented in Table 3.6 below shows that there were no significant differences between the two task conditions in terms of pausing measures. As for repair types, although a small number of Error-repair was found in the two task conditions, it should be noted that the dual task condition elicited significantly more

Error-repair (M = .02, SD = .02) than the single task condition (M = .008, SD = .01); t(15.49) = -2.11, p = .05.

Table 3.6. Self-monitoring aspects in the two task conditions (the Pilot)

Dimension	Measures	Task condition				
		Single		Dual		Sig. (two-tailed)
		M	SD	M	SD	*P
Pauses	Mid-clause pauses number	12.27	1.18	14.43	1.12	.20
	Mid-clause pauses length	.19	.03	.20	.02	.92
	End-clause pauses number	8.23	.79	7.80	.66	.68
	End-clause pauses length	.14	.02	.16	.02	.52
Repair types	Global-repair	.02	.01	.19	.37	.14
	Message-repair	0.01	0.01	0.02	0.02	.37
	Error-repair	0.008	0.01	0.02	0.02	.05
	Error-to-cut-off	1.60	.38	3.07	.65	.06
	Cut-off-to-repair	1.06	.32	2.35	.52	.04
	Repair	1.37	.28	3.14	.72	.03

*p < 0.05

Message-repair was similar in the single task condition (M = .01, SD = .01) and the dual task condition (M = .02, SD = .02); t(20.08) = -.90, p = .37. Global repair was also similar in the two task conditions. This shows that dual task condition had a considerable influence on the Error-repair which suggests that L2 learners made more corrections in terms of linguistic errors in the dual task rather than in the single task conditions.

Concerning the temporal phases of repairs, the data indicates that the dual task condition was more demanding than the single task condition. The first phase of repair (Error-to-cut-off) was approaching more significance (p = .06) in the dual task condition (M = 3.07, SD = 0.65) than in the single task condition (M = 1.60, SD = .38). This means that L2 learners spent slightly longer time on the first phase of repair in the dual task rather than in the single task conditions. Cut-off-to-repair interval, on the other hand, increased significantly (p = .04) in the dual task condition (M = 2.35, SD = 0.52) than the single task condition (M = 1.06, SD = 0.32). This suggests that the increased demand of the task condition led L2 learners to spend longer time on the second phase of repair (Cut-off-to-repair) than in the single task condition. Also, significantly longer repair phases, (p = .03) were produced in the dual task condition (M = 3.14, SD = 0.72) than in the single task condition (M = 1.37, SD = .28). This suggests that L2 learners needed longer time to make repair in the dual task condition than in the single task condition. Thus overall, the data in this section suggests that the dual task condition imposed greater cognitive demand on some aspects of L2 self-monitoring, particularly Error-repair and temporal phases of repair. Therefore, it can be assumed that the dual task condition might properly be operationalised in this pilot study.

3.11. **Discussion**

The pilot study aimed to develop two similar oral narrative picture prompts. The two picture stories were designed specifically for the purpose of the current study following the guidelines of comparable picture prompts proposed by De Jong and Vercelloti (2016). The pilot study also aimed to test the operationalisation of the dual task condition which was designed following the theoretical principles of tasks similarity proposed by Wickens (2007). The data will be discussed in relation to previous studies. The findings will also be interpreted in light of the principles of the PLT. These are discussed in Sections 3.12 and 3.13.

1.1.1. Picture Prompts

As several steps were taken to ensure the similarity of the two picture stories (see Section 3.4.1), it was expected that the picture stories (Ship and Fire) would elicit similar aspects of monitoring (pauses, repair types, and temporal phases of repair). The data shows that performing the two picture stories resulted in differences in some aspects of monitoring, namely, the length of within-clause pauses and the first phase of repair (Error-to-cut-off). Previous research suggests that within-clause pauses relate to difficulties in some sub-processes of the Formulator (e.g., lexical access and syntactic encoding) (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016; Skehan, Xiaoyue, Quian, & Wang, 2012). This means that the Ship prompt might be more demanding than the Fire especially at the Formulator level. In terms of the vocabulary frequency, the Ship has more frequent lexical items (96.07%) than the Fire (93.75%), thus it is unlikely that the Ship is more demanding in terms of the vocabulary. The data also showed that significantly longer intervals at the first phase of repair (Error-to-cut-off) was produced in the Ship than the Fire prompt. Previous research on self-monitoring found that conceptual difficulty of tasks slowed down the first phase of repair (Error-to-cut-off) (Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008). This suggests that the Ship imposed greater conceptual demand on L2 learners. Thus, it can be concluded that although some procedures were used to make the two picture stories similar, they are still different.

The pilot study attempted to establish similarity between the two picture prompts along certain dimensions, namely, structure, storyline complexity, number of elements, and vocabulary frequency. However, it seems likely that other aspects of the Ship generated more cognitive demand during performance such as the familiarity of the story itself. That is, the classroom setting (in the Fire) might be much more familiar to Saudi students than going on a journey by the sea, given that they experience classroom context in their daily life and that they are often provided with instructions of how to act in the case of fire. Other potential cause of variance in the two picture prompts could be the places and times of different events in the stories. That is, the Ship story took place in several locations including the seaport, in a ship during calm weather and a storm, on an island during day and night, and in a house. In the Fire, almost all the events occurred either inside or outside the school at a specific time of a school day. Therefore, it was evident that further action needed to be taken to ensure the comparability of the two picture prompts before using them in the main study.

1.1.2. Task condition

The pilot study also aimed to test the operationalisation of the task condition. According to the PLT, self-monitoring is dependent on cognitive resources (Levelt, 1983). It has been argued that the dual task condition would negatively affect self-monitoring (Oomen & Postma, 2001, 2002). The operationalisation of the dual task condition in the current pilot study was based on the theoretical underpinnings of the concurrent tasks proposed by Wickens (2007). It was expected that greater number and length of pauses would be produced in the dual task condition as compared to the single task condition. It was also predicted that dual task condition would elicit less repair and longer intervals of repair phases. The following sub-sections will discuss the effects of task condition on the three aspects of self-monitoring employed in the current pilot study (pauses, repair types, and the temporal phases of repair).

1. Pauses

According to the PLT, pauses are hypothetically produced as corrective measures to deal with expected errors (Levelt, 1983). That is, pauses are produced by speakers to buy time for correcting errors before the speech is produced. As such, it was expected that a greater number and longer duration of pauses would be produced in the dual task condition rather than in the single task condition. The data showed that there were no significant differences between pauses in the two task conditions. These findings are in line with Declerck and Kormos's (2012) study which found that there were no differences between disfluency in the two task conditions. A number of researchers argue that pausing might relate to personal speaking style (Declerck & Kormos, 2012; De Jong, Groenhout, Shoonen, & Hulstijn, 2015; Derwing, Munro, Thomson, & Rossiter, 2009; Krashen, 1981; Lennon, 1990; Skehan & Foster, 2005). However, it could be assumed that disfluency as an aspect of monitoring would be affected by other factors such as the development of proficiency. As such, future studies need to examine different aspects of disfluency (such as repetition and filled pauses) at different proficiency levels.

2. Repair types

It was hypothesised that dual task condition would consume available cognitive resources during L2 processing thus less repair would be produced. The results indicated that participants made significantly more Error-repair in the dual task condition than in the single task condition. Means and standard error of means (SEMs) of repair types in the two task conditions are presented in Figure 3.8 below.

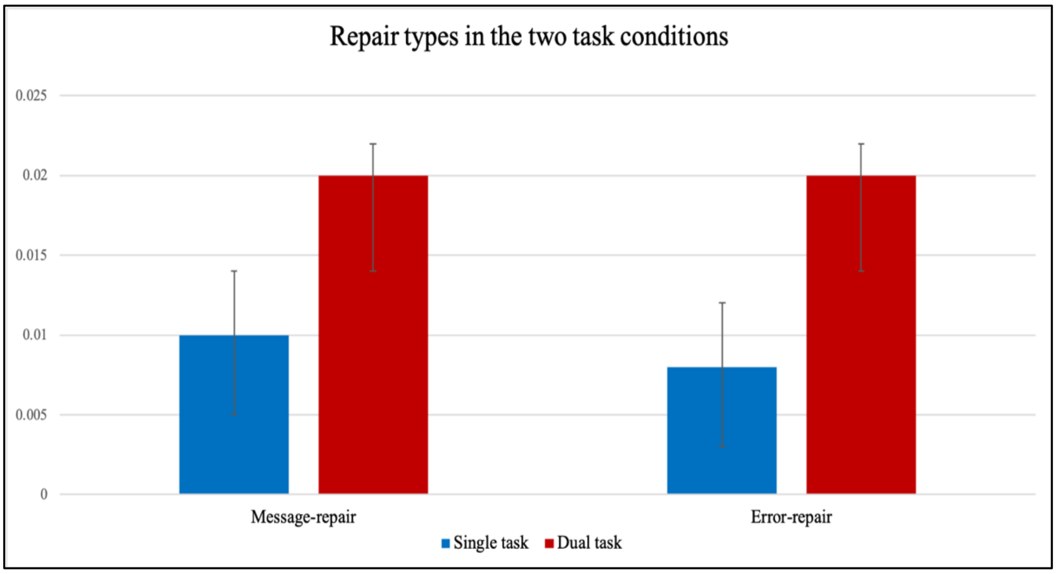


Figure 3.8. Repair types in the two task conditions (the Pilot)

On the other hand, there was no significant differences between Message-repair in the two conditions. Figure 3.8 above shows that both types of repair (i.e. Message-repair and Error-repair) increased during the dual task condition, yet, only Error-repair reached significance in the dual task condition. The data contradicted the hypothesis stated above. This suggests that due to the limited cognitive resources during the dual task condition, the monitor might direct the available resources to focus on a certain aspect of speech which is, in this case, Error-repair. In other words, with the increased demand of the dual task condition, L2 learners might focus mainly on correcting linguistic errors to ensure production of correct utterances. This is in line with L2 research which found that with the increase of task demand, students noticed more errors and corrected a higher proportion of their errors (e.g., Gilabert, 2007).

Importantly, the increase of Error-repair during the dual task condition may suggest that L2 self-monitoring became more active in the dual task condition to enable the speakers to cope with the increased demand, as can be seen in Figure 3.8 above. It appears that consuming cognitive resources along the dual task condition may not have a negative impact on repair types. In other words, both types of repair (Message-repair and Error-repair) increased rather than decreased in the dual task condition, although only E-repair reached more statistical significance in the dual task as compared to single task condition. This finding supports Levelt's (1989) assumption that self-monitoring is sensitive to contextual factors, and that the monitor becomes intense in the cognitively demanding condition (Levelt et al., 1999). This finding is also in line with the assumption that certain task characteristics may draw attention towards linguistic form whereas other tasks may deviate attention from focusing on form (Robinson, 2011). However, such an assumption needs to be tested in future studies with a larger number of participants for better validation.

3. Temporal phases of repair

It was hypothesised that the durations of repair phases would be longer in the dual task condition than in the single task condition. This hypothesis was partly supported as only two repair phases were significantly longer in the dual task than the single task conditions. Although it was expected that the dual task condition would result in longer Error-to-cut-off phase (the first phase of repair) as compared to the single task condition, the data indicated that the duration of this phase increased in the dual task condition but it did not reach significance. Means of repair phases in the two conditions are illustrated in Figure 3.9 below.

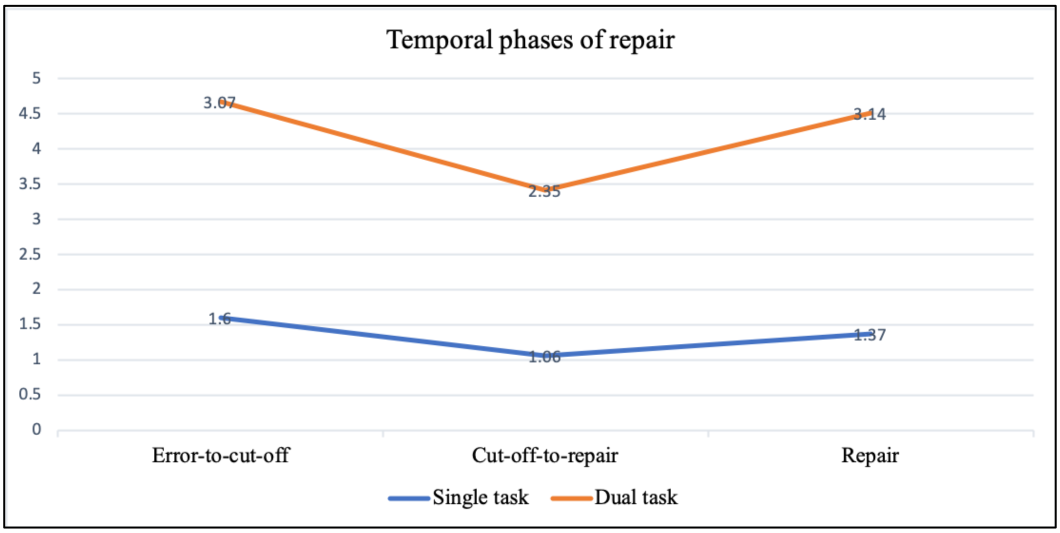


Figure 3.9. Repair phases in the two task conditions (the Pilot)

The data suggests that L2 learners were able to detect their errors or inappropriateness in a similar duration in the two task conditions. There is still ongoing debate about whether this phase is used for error-detection or involves both error-detection and planning (see Section 2.4.2 for further discussion).

The Cut-off-to-repair phase (the second phase) was expected to be longer in the dual task condition as compared to the single task condition. This phase is used to hold the floor while planning repair and to give a signal to listeners that there is a problem in the speech (Levelt, 1989). The data shows that this phase was significantly longer in the dual task condition, which suggests that L2 learners spent a significantly longer interval on re-planning their utterances when they performed the dual task as compared to the single task conditions. According to the PLT, Error-repair involves a longer duration of Cut-off-to-repair phase than other repair (A-repair) as Error-repair entails correcting erroneous information. In the current data, temporal phases of repair were globally analysed, so that it is not clear to what extent this phase was affected by dual task condition in different repair types. As for the third temporal phase of repair (repair execution), it was expected that this phase would be longer in the dual task condition as compared to the single task condition. The findings confirm that this phase was significantly longer in the dual task as compared to the single task condition. This suggests that the exhausted cognitive resources in the dual task condition slowed down the execution of repair. The data of the temporal phases of repair supports the PLT which assumes that consuming cognitive resources affect monitoring functions (Levelt, 1983). In this case, it slowed down temporal phases of L2 repair.

To conclude, this pilot study provides evidence that dual task condition, as operationalised in this study, seems to have consumed available cognitive resources. That is to say, the increased cognitive demand of the dual task condition probably led L2 learners to focus solely on one type of repair during speech processing, namely, Error-repair. In addition, producing prolonged temporal phases of repair in the dual task condition as compared to the single task condition could be used as further evidence to confirm that dual task condition imposed greater cognitive demand during L2 processing. However, in this pilot study, the temporal phases of repair were globally examined, and this did not indicate how phases of each repair type function in different task conditions. Future research needs to examine the temporal phases of different repair types separately to find out how these phases of repair are affected by the increased demand of task condition.

3.12. **The closing remarks of the Pilot Study**

The pilot study was conducted to develop two similar oral narrative picture prompts and to test the operationalisation of the dual task condition. Concerning the picture stories, it was expected that the Ship and Fire would elicit similar speech samples provided that several steps were taken to ensure the similarities of the picture prompts. The results showed that there were differences between the two picture stories in terms of mid-clause silent pauses and one of the temporal phases of repair (Error-to-cut-off). Previous studies found that pausing within clause boundaries relates to difficulties at the Formulator level (Lambert et al., 2017; Skehan et al., 2016; Skehan et al., 2012). Likewise, it has been found that the Error-to-cut-off phase slowed down as a result of conceptual demand (Broos et al., 2018; Hartsuiker et al., 2008). It was then concluded that the picture prompts are not similar, and that further action needs to be taken to ensure the similarity of the picture prompts.

Concerning task condition, it was predicted that dual task condition would impose greater cognitive demand on L2 learners; thus, more pauses and longer duration of temporal phases would be produced in the dual task rather than the single task conditions. It was also expected that less repair would be produced in the dual task as compared to the single task conditions. The results indicate that participants made significantly more Error-repair in the dual task condition rather than the single task condition. This data suggests that due to resource limitations, L2 learners might focus on one aspect of their speech during monitoring which is Error-repair, since erroneous output is likely to affect communication. With respect to the temporal phases of repair, the duration of the three phases of repair increased in the dual task condition although only two phases reached statistical significance. The significant increase in the duration of Cut-off-to-repair and repair phases indicate that these two phases were influenced by the dual task condition. These findings suggest that dual task condition as operationalised in the current pilot study imposed great cognitive demand on some aspects of L2 self-monitoring. Thus, the current operationalisation of the dual task condition will be employed in the main study.

One of the issues that needs to be addressed in the main study is the need to employ more measures of self-monitoring. The pilot study used pauses, repair types, and temporal phases of global repair. The main study needs to include more disfluency measures provided that disfluency measures are useful in examining self-monitoring as suggested in the PLT (Levelt, 1983). There is also a need to investigate different types of repair as classified in Kormos's taxonomy (1999), given that this classification of repair can highlight different foci of monitoring during spontaneous speech production. Temporal phases of different types of repair need to be examined separately as these phases could provide important information about the different processes of self-repair such as error-detection and planning. Another important aspect of monitoring that needs to be examined in future studies is the ratio of error-correction (number of errors divided by number of repair). It was suggested that ratio of error-correction is an accurate predictor of self-monitoring (Oomen & Postma, 2001, 2002).

Other limitations of the current pilot study include the small sample size (n = 12), and recruiting participants from one proficiency level, namely, the B2 level. The main study needs to recruit a larger number of participants from different proficiency levels. Examining self-monitoring at different proficiency levels could inform L2 research about the development of different aspects of self-monitoring.

Chapter 4: METHODOLOGY

4.1 Introduction

The purpose of this chapter is to present the objectives and the experimental materials of the current study. In Section 4.3 the research questions, which were formulated on the basis of the literature review in Chapter 2, are introduced. The design of the study is discussed in Section 4.4. This is followed by a description of the recruitment of participants and the experimental materials. This chapter also introduces tests of proficiency and working memory capacity (Sections 4.6.3 and 4.6.4). Procedures and methods of data collection are presented in Section 4.8, and a detailed explanation of coding procedures and measures of monitoring aspects is provided in Section 4.9. This chapter concludes with describing the procedures of checking inter-rater reliability and data analysis.

4.2 Aims of study

This study aims to investigate the extent to which L2 self-monitoring aspects are affected by proficiency levels, task conditions and working memory capacity. In terms of proficiency level, researchers from the neuroimaging field argued that language proficiency is likely the most important factor that influences the bilingual language system (Abutalebi, Cappa & Perani, 2001) (see Section 2.8). Previous research on self-monitoring has suggested that proficiency influences some aspects of L2 self-monitoring such as self-repair foci (see Section 2.6.1). As such, the current study aims to explore the effects of proficiency development on different aspects of self-monitoring including disfluency, different types of self-repair, temporal phases of repair, and accuracy.

This study also aims to explore the impact of dual task condition on the functioning of self-monitoring, since self-monitoring is sensitive to contextual factors (Levelt, 1983) (see Section 2.9). The choice of dual task condition was based on previous studies which regarded the dual task paradigm as a preferred psycholinguistic technique for examining the influence of resource limitations on self-monitoring (see Section 2.9.1). As such, this study uses dual task condition to examine the effects of resource limitations on L2 self-monitoring aspects.

Section 2.10.3 demonstrated that self-monitoring draws on working memory resources during its functioning (Levelt, 1983). Thus, it is important for self-monitoring research to examine the relationship between working memory capacity and L2 self-monitoring aspects. Despite the importance of the relationship between working memory and self-monitoring, few studies have so far examined this relationship (e.g., Mojavezi & Ahmadian, 2014) (see Section 2.10.3). Thus, this study seeks to understand the extent to which working memory capacity is implicated in L2 self-monitoring behaviour. The next section presents research questions (RQs) and hypotheses (Hs) of the study.

4.3 Research Questions (RQs) and Hypotheses (Hs)

This section presents the research questions (RQs) that guide the current investigation. Some research questions have hypotheses (Hs) which were formulated following the premises of the PLT (Perceptual Loop Theory) and previous studies (see Section 2.11). It should be noted that the hypotheses relate to the functioning of two loops of the monitor (the inner loop and the auditory loop) which were discussed in Section 2.3.3. The differences in proficiency and WMC are expected to support the functioning of the inner loop of monitoring and therefore lead to the decrease of some aspects of monitoring in overt speech such as disfluency, certain repair types (E-repair sub-categories), and temporal phases of repair. A higher proficiency level and a larger WMC are also expected to promote accuracy in L2 speech. The increased demand of task condition, on the other hand, is expected to tax the inner loop by consuming the available cognitive resources which would subsequently lead to increasing the number of monitoring aspects that the auditory loop detects. That is, more aspects of monitoring are expected in overt speech (e.g., disfluency, E-repair sub-categories, and temporal phases of repair). A detailed account of this assumption will be provided under each RQ below.

RQ1. To what extent does proficiency level affect L2 self-monitoring behaviour in terms of disfluency, repair types, repair temporal phases, and accuracy?

H1. Disfluency and accuracy: According to the PLT, disfluency (e.g., filled pauses, mid-clause silent pauses, and repetition) is produced as corrective action to anticipated errors (Levelt, 1983). It could be predicted that if disfluency is corrective action to anticipated errors, they are expected to decrease with the development of proficiency. This is because accuracy increases as a result of proficiency development. Thus, it is expected that fewer disfluency features and a higher rate of accuracy would be produced by the Intermediate group as compared to the Elementary group. Silent pauses at clause boundaries are expected to behave differently from other disfluency measures. The Intermediate learners are expected to produce more silent pauses at clause boundaries compared to the Elementary learners (see Section 4.9.1).

H1b. Repair types: Based on previous L2 studies (e.g., Kormos, 2000a; Van Hest, 1996), it is expected that the Intermediate group would produce more A-repair, and fewer E-repair than the Elementary learners.

H1c. Temporal phases of repair: These features have rarely been examined in L2 research (see Section 4.9.3). Kormos (2000b) found that the first phase (Error-to-cut off) of A-repair was shorter in the speech of the higher proficiency learners than intermediate learners. Thus, it is expected that duration of A-repair first phase (Error-to-cut-off) would decrease in the Intermediate learners as compared to the Elementary learners. It is also expected that other phases of repair would be shorter in the Intermediate group as compared to the Elementary group.

RQ2. To what extent does task condition affect L2 self-monitoring behaviour in terms of disfluency, repair types, repair temporal phases, and accuracy?

H2. Based on the PLT, it is hypothesised that the dual task condition would consume available cognitive resources, which support the functioning of the inner loop, and thus it would have detrimental effects on pre-articulatory monitoring (see Section 2.4.3). This would result in more disfluency, longer intervals of repair, greater number of E-repair, fewer number of A-repair, and error-corrections in the dual task as compared to the single task conditions. Rate of accuracy is expected to decrease in the dual task as compared to the single task conditions.

RQ3. Is there an interaction between proficiency levels and task conditions in influencing L2 self-monitoring behaviour in terms of disfluency, repair types, repair temporal phases, and accuracy?

H3. An interaction effect is predicted between proficiency level and task condition in relation to monitoring behaviour. That is to say, it is expected that the higher proficiency participants would perform better than the lower proficiency participants under the dual task condition.

RQ4. Is there a relationship between working memory capacity and L2 self-monitoring aspects (disfluency, repair types, repair temporal phases, and accuracy) in the single and dual task conditions?

H4. A larger WMC is hypothesised to promote the functioning of the inner loop during the pre-articulatory monitoring. However, the more demanding task condition is expected to have a negative impact on the working memory functioning (Michel, Kormos, Brunfaut, Ratajczak, 2019). This means that in the current investigation, the dual task condition might have a negative influence on the functioning of working memory because it is expected to tax available cognitive resources. As such, it is hypothesised that WMC would enhance L2 monitoring in the single task rather than dual task conditions. The next section (4.4) presents a description of the current study design.

4.4 Study Design

The current study is considered a quasi-experimental study since that the recruitment of participants was not completely random. Quasi-experiment is similar to true experiment in all aspects with the exception of the random recruitment of participants (Dörnyei, 2007). The random recruitment and assignment of participants to one group or another is an essential feature of experimental design (Dörneyei, 2007; Paltridge & Phakiti, 2015). In language classrooms, researchers cannot always have access to a random selection of participants because they depend on intact classes or cohorts of students in a particular setting. As such, this is called quasi-experimental design because not all variables can completely be controlled (Paltridge & Phakiti, 2015). However, this practical limitation has become acceptable in contexts where random selection of participants is not possible (Dörnyei, 2007). To improve the design of the quasi-experimental study, it has been recommended that researchers should avoid allowing participants to select themselves to be in a particular group (Dörnyei, 2007). In other words, there should be random assignment of participants to different groups (Paltridge & Phakiti, 2015). As such, in the current study, the participants were randomly assigned to the two task conditions.

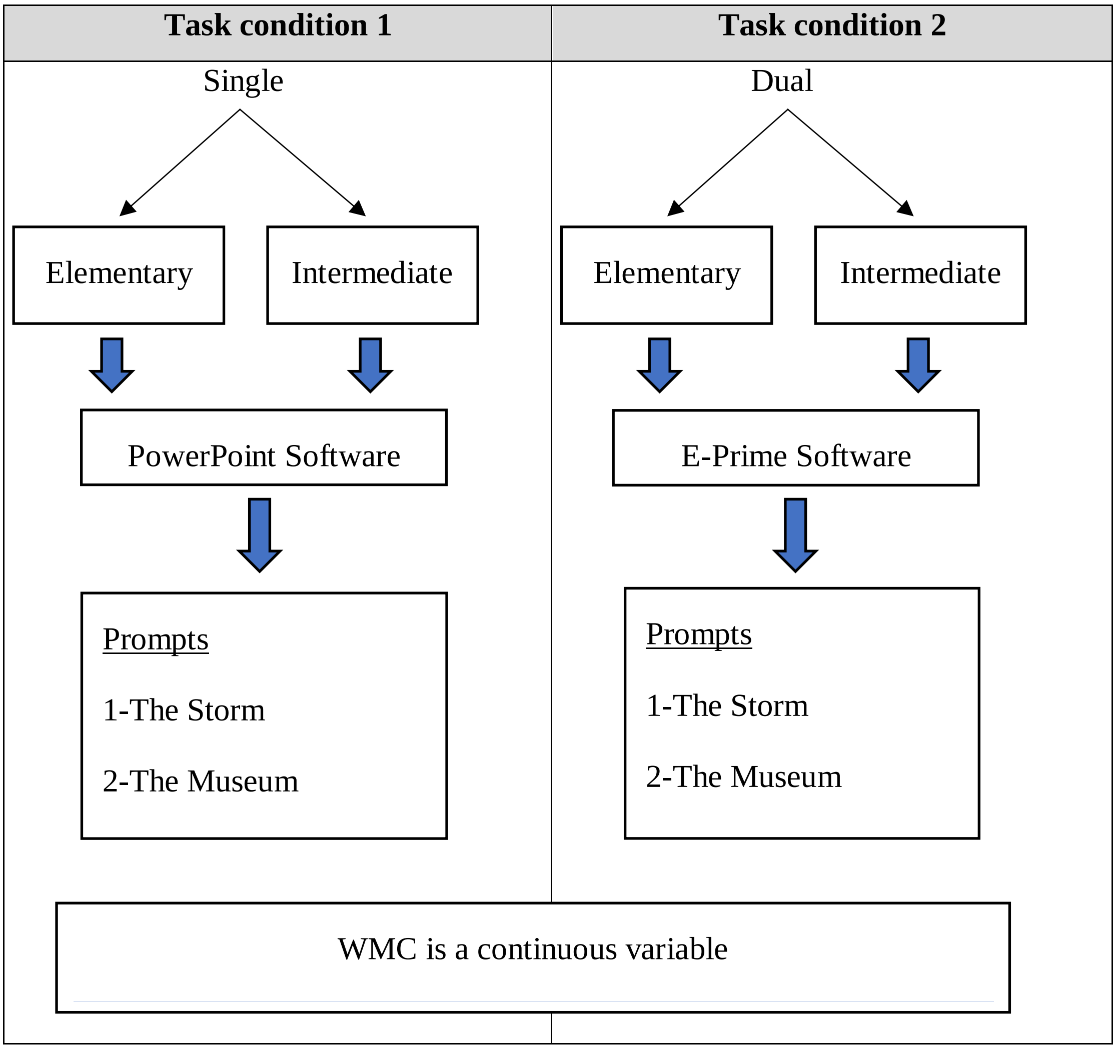
On the other hand, part of the current investigation is considered exploratory in nature. That is, as the temporal phases of L2 self-repair have rarely been examined in L2 research, the current study attempts to explore this aspect of monitoring. It is not clear to what extent these phases of self-repair are influenced by proficiency development. Likewise, L2 research lacks information about the way these phases of repair behave under different task condition (see Section 2.4.2 for further details). Exploratory research is preferred when a ‘process, activity, or situation has received little or no systematic empirical scrutiny’ (Stebbins, 2001, p.7). Several definitions of the word ‘explore’ have been discussed by Robert Stebbins (2001). One of these definition states that ‘to explore something is to examine it systematically’ (Stebbins, 2001, p.7). This type of exploration is regarded as limited because the researcher already knows what to look for (Stebbins, 2001, p. 3). In other words, such exploratory study does not involve innovation or broad discovery, it only aims to systematically examine a certain feature (Stebbins, 2001). To the best of my knowledge, this is the first study that examines the three temporal phases (Error-to-cut-off, Cut-off-to-repair, Repair) of different repair types (A-repair, D-repair, E-repair sub-categories) in two proficiency levels (elementary, intermediate) under two task conditions (single and dual), as demonstrated in Table 4.7 below. Thus, this specific part of the current investigation is considered exploratory.

Table 4.7. The variables of the study

Study Design	Independent Variables	Dependent Variables
A 2x2 between-subject Design **N = 66**	**1- Proficiency Level (PL)** Elementary group (Group 1) Intermediate group (Group 2) **2- Task Condition (TC)** Single task condition (Task condition 1) Dual task condition (Task condition 2) **3- Working Memory Capacity** Continuous variable	**L2 self-monitoring behaviour** 1- Disfluency 2- Repair types 3- Temporal phases of repair 4- Accuracy

In the current study, a 2x2 between-subject factorial design was employed to examine the effects of proficiency level and task condition and their interaction effects on L2 self-monitoring behaviour. There are three independent variables in the current study which are proficiency level, task condition, and working memory capacity. Proficiency level and task condition are between-participant variables, whereas working memory capacity is a continuous variable, as indicated in Table 4.7 above. Between-subject design was used in the current study for a practical reason. In other words, it was more convenient to run one software at a time for each participant, so that each participant could perform in either single or dual task conditions, rather than running two programs at the same time for a participant to perform two task conditions, as indicated in Table 4.8 below. Participants were divided into two groups: one group performed in the single task condition and the second group performed in the dual task condition, as illustrated in Table 4.8 below. Each task condition involved two proficiency groups: Elementary and Intermediate groups (see Table 4.8 below).

Table 4.8. Study design and software



Oral narrative picture prompts were counterbalanced in the two task conditions to control any effect of order and practice. WMC was measured before learners started their performance under the task conditions. Dependent variables include four aspects of L2 self-monitoring which are disfluency, repair types, temporal phases of repair, and accuracy (see Section 4.9).

4.5 Participants

This study recruited 66 learners of L2 English from the Department of Foreign Languages in Taif University, Saudi Arabia. The participants were undergraduate students whose major was English. They were female, aged 18-23, and spoke Arabic as their first language. Prior to their participation in the study, these students had received 8-9 years of formal English instruction, 7 years at primary and secondary schools, and 1-2 years at the university. Most of the participants were in their first year at the university, some of them in second or third year. The participants on the whole had never studied or lived in English-speaking countries. Only one student lived in Australia until the age of four, then returned to Saudi Arabia, another lived in the UK for three months; these two students were considered relevant in the current study based upon their level of proficiency. Proficiency level was assessed using the C-test and Elicited Imitation Test (for further details, see Section 4.6.3).

4.6 **Tasks and instruments**

This section describes the materials and tools used to collect data and assess proficiency and working memory capacity in the present study. Section 4.6.1 presents the procedure employed to ensure the similarity of the two oral narrative picture prompts. Section 4.6.2 provides a description of the two task conditions (single and dual). This is followed by detailed descriptions of assessing language proficiency and working memory capacity in Section 4.6.3 and Section 4.6.4 respectively. Moreover, this section (4.6) summarises the result of piloting the picture prompts. It also introduces preliminary impression about participants' behaviour in terms of the dual task condition (Section 4.6.2), proficiency level (Section 4.6.3) and working memory capacity tests (Section 4.6.4).

4.6.1 Picture prompts

In the current study, two oral narrative picture prompts were used to elicit L2 learners' oral performances. To avoid repetition effects, it has been suggested that a different picture prompt needs to be used for each task condition (De Jong & Vercelloti, 2016; Ishikawa, 2007). On the other hand, others have argued that a single prompt for each task condition would result in a confounding effect since the prompts may not elicit similar speech samples (De Jong & Vercelloti, 2016). Moreover, seemingly similar tasks have been found to differ in the language they elicited (De Jong & Vercelloti, 2016). Therefore, it has been suggested that picture prompts need to be comparable along different task features including structure, storyline complexity, and number of elements to control for task effects on elicited speech samples (De Jong & Vercelloti, 2016) (see Section 3.4.1). These suggestions were considered when designing the oral narrative picture prompts of the current study (see Section 3.4.1). The picture prompts were piloted, and it was expected that the two picture prompts (Ship and Fire) would elicit similar speech features (disfluency, repair types, and temporal phases of repair). The data of the pilot study showed that performing the two picture stories resulted in differences in some aspects of monitoring, namely, the length of within-clause pauses and the first phase of repair (Error-to-cut-off) (see Section 3.10.1). As the data showed that the Ship was more demanding than the Fire, it was concluded that the two picture stories were not similar.

To identify potential differences between the two picture prompts, they were presented to a group of English teachers in the same institution where the participants were recruited. Some of the teachers pointed out that there were more details in the Ship picture prompt than the Fire. Others added that the story of the Ship might not be familiar to participants. In fact, Section 3.12 suggests that the greater variety of settings and timeframes in the Ship than the Fire might account for the differences between the two picture prompts. As such, it was decided that these picture prompts need to be revised thoroughly and re-piloted. This process started by changing the setting in the Fire, so that a museum was chosen instead of a school. Then, a new scenario was written and submitted to VocabProfilers (Cobb, 2017). The results displayed that the new stories in the Ship and the Museum have respectively 96.07% and 92.15% of vocabulary relate to 1K+ 2K levels. Off-list words in Museum was 6.86% and 3.92% in Ship. It should be noted that most words which were classified as off-list in the Museum are relatively common words such as museum, exit, ambulance. The VocabProfilers showed that off-list words in the Ship were seaport, dolphins, and helicopter. A smaller pilot study was conducted on the new tasks. During the piloting stage, participants suggested that it would be useful to reveal some words to participants before performance. As such, three words were given in each task. In the Ship picture prompt, participants will be given three words prior to performance: thunder storm, island, sharks; in the Museum picture prompt, they will be provided with statues, socket and fire engine.

To ensure similarity of the two picture prompts, a short survey was sent to a number of researchers and EFL teachers (n = 12) using the Survey Monkey website (see Appendix 7). Among the items of the survey were questions about vocabulary and storyline complexity. With respect to vocabulary, the results indicated that the majority of respondents agreed that the two narrative picture prompts are either similar or very similar in terms of vocabulary. Concerning complexity of stories, the survey showed that the majority of respondents felt that the stories are either similar or very similar in terms of storyline complexity. It was therefore assumed that the two picture prompts are likely similar and would elicit comparable speech samples. The updated versions of the oral narrative picture prompts were piloted. Eight native speakers of Arabic, who were learners of L2 English, were recruited to perform the two picture prompts (see Appendix 8). Speech samples were transcribed and coded for measures of breakdown (i.e. mid-clause and end-clause pauses frequency and duration). The rationale for using breakdown measures is due to the assumption that these measures can show the effects of task complexity on underlying speech processing (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016). Results showed that the two picture prompts elicited similar speech features, which suggests that the two picture prompts involved equal cognitive demand. Participants were also asked about their perception of task complexity (difficulty of stories) and most of their comments showed that they found the two tasks almost identical. Based on these steps, it was concluded that the two narrative picture prompts are similar, and therefore they were ready to be used in the data collection.

4.6.2 Task condition

Task condition in the present study was operationalised based on Wickens' (2007) proposal for degrees of similarity between simultaneous tasks (see Section 3.4.2). According to Wickens (2007), the degree of similarity between tasks can decide the extent to which they draw on the same resource pool (see Section 2.9.2). In the current study, task condition includes single and dual task conditions. The single task condition entails describing oral narrative picture prompts, whereas the dual task condition involves describing picture prompts as well as performing a simultaneous secondary task (see Section 3.4.2). The secondary task entails having words appear on the computer screen randomly. These words either describe animate or inanimate objects (see Appendix 5). The participants were required to hit a certain button in each case. Two selected keyboard buttons were marked with Arabic translation for the words 'animate' and 'inanimate' which are (حي) and (جماد) respectively (see Section 3.4.2 for more details). Word lists that were used in the secondary tasks include names of familiar animals (e.g., dog, cat, tiger), and items of furniture or familiar objects (e.g., door, pen, car) (see Appendix 9-C). An expert researcher and some English teachers were consulted about the familiarity of these word lists to L2 learners. The pilot study indicated that the operationalisation of task condition was effective in generating cognitive demand on L2 self-monitoring aspects as compared to the single task condition (see Sections 3.11 and 3.13). However, only few aspects of L2 self-monitoring were employed in the pilot study. To understand the effects of dual task condition on a wider range of monitoring aspects, further aspects of self-monitoring will be employed in the current study (see Section 4.9 below).

Task condition was introduced to two groups of participants as presented in Section 4.4 above. Half of participants performed in the single task condition (n = 33) and the other half performed in the dual task condition (n = 33). As each participant described two picture stories, the total number of speech samples was 132, with 66 speech samples in each task condition as summarised in Table 4.9 below.

Table 4.9. Single and dual task conditions

Task condition	Software	Picture prompts	Participants	Performances
Single	PowerPoint	Storm & Museum	33	66
Dual	E-Prime	Storm & Museum	33	66

E-Prime software (3.0) was used to administer the dual task condition. This software is suitable for computerised experimental design. It is mostly employed in the field of cognitive psychology due to its efficiency in handling milliseconds precision timing while a task is being administrated. That is, it calculates the reaction times when participants respond to stimuli. It also calculates the accuracy of hitting the appropriate buttons in the secondary task. Five responses in the secondary task were damaged during data collection. It seems that these responses were overwritten by other data which were calculated directly afterwards. Fortunately, all speech samples were recorded on Voice Record App and directly saved to OneDrive; thus, speech samples were all intact and properly stored during data collection. The data of the secondary task is not to be used to answer any research question; it is only used to provide a general impression about participants' behaviour during performing dual task condition, as illustrated in Table 4.10 below.

Table 4.10. Accuracy of responses and reaction times in the secondary task

Secondary task data	Mean	Min.	Max.	Std. Dev	Std. Err
Accuracy of responses Max = 1	0.72	.45	1.00	.136	.017
Reaction times Max = 5000 ms	1895.1	634.5	2888.5	461.6	58.15

Table 4.10 above summarizes the accuracy rate of keyboard responses and the reaction times during the secondary task, that is, the time that participants spent when responding to stimuli

in the secondary task. The data above demonstrated that the average of accuracy of keyboard responses was high (72%) which means that the majority of participants were engaged with the secondary task while they were describing the oral narrative picture prompts. This suggests that the aim of introducing the secondary task was probably met in this study. Reaction times data demonstrated that the average time of responding to stimuli was about 1.81 seconds (1895.1 ms). This suggests that most participants managed to respond to stimuli within a short time which means that the participants managed to decide within a short time whether a word on the screen describes an animate or inanimate object.

4.6.3 Language proficiency tests

As proficiency level plays an important role in understanding language processing in L2 learners (Dunn & Fox Tree, 2009), it is important to ensure the use of reliable tests to assess learners' proficiency level. In fact, the measurement of language proficiency has been criticised for being inadequately handled in the L2 field (Hulstijn, 2012) (see Section 2.8.2). Assessment of language proficiency is a complicated task since proficiency is a concept that is not easily quantifiable and not directly measurable (Paltridge & Phakiti, 2015). To ensure construct validity in the assessment, L2 researchers need to employ measures and tests that adequately capture the construct under investigation (Paltridge & Phakiti, 2015). Construct validity 'refers to the extent to which the research adequately captures the concept in question' (Paltridge & Phakiti, 2015, p. 13).

Researchers have proposed several definitions for language proficiency, as discussed in Section 2.8. Among these definitions is the one which states that proficiency is 'knowledge of language and the ability to access, retrieve and use that knowledge in listening, speaking, reading and writing' (Hulstijn, 2015, p. 21). Although a consensus on L2 proficiency definition has not yet been reached, Hulstijn's (2015) definition is comprehensive as it highlights several aspects of language ability including linguistic knowledge as well as the processing and production of a language. Thus, this definition is used as a working definition of proficiency in the present study. In other words, the current study will attempt to assess linguistic knowledge, processing and production of second language by employing two tests, namely, the C-test and the Elicited Imitation Test. It has been argued that using multiple measures can enhance construct validity (Paltridge & Phakiti, 2015). The next section will present the two tests used in the current study (i.e. the C-test and the Elicited Imitation Test).

The C-test

As the current study is interested in examining L2 self-monitoring via the analysis of learners' speech samples, it was decided to employ tests that can assess linguistic ability and processing in written and oral modalities. The C-test was selected because it is reliable and practical in assessing language ability (for further discussion, see Section 2.8.2). A bulk of literature suggests that the C-test is a reliable measure of L2 proficiency (Babii & Ansari, 2001; Dörnyei & Katona, 1992; Richard, Daller, Malvern, Meara, Milton, & Treffers-Daller, 2009; Sigott, 2004). As discussed in Section 2.8, C-test is described as practical and economical in use and administration, and objective in scoring (Dörnyei & Katona, 1992; Feldmann & Stememr, 1987; Klein-Braley, 1997; Klein-Braley & Raatz, 1984). More importantly, there is strong research evidence to suggest C-test ia a valid measure of general language proficiency. (Eckes & Baghaei, 2015; Eckes & Grotjahn, 2006; Raatz, 1984, 1985). Finally, C-test can be administered to a large number of participants at the same time. Based on this rationale, C-test was used for the purpose of screening in the current study. However, this test has been criticised for not including the assessment of oral language skills. This is considered a serious limitation for studies that investigate oral modality (Tremblay, 2011). Thus, in the current study, the C-test was used alongside the elicited imitation technique which has been accepted as valid in assessing oral language proficiency (Housen & Pierrard, 2005). It has been assumed that using two tests (written and oral) can give a comprehensive assessment of different aspects of learner's linguistic knowledge (Awwad, 2017). Furthermore, the C-test assesses declarative knowledge whereas EIT can measure procedural knowledge (Ellis, 2009; Erlam, 2006) (for further discussion, see Section 2.8.2). The C-test in the current study was adopted from (Daller & Yixin, 2017) (see Appendix 10). This test consists of five texts in the target language (English) to which the rule of two was applied: the second half of every second word of the second sentence; the first and last sentences are left intact; if a word has an odd number of letters, the bigger part is omitted; and one letter word such as I is ignored in the counting. There were 20 missing parts in each text and the students need to provide the missing parts. The maximum score that a student can obtain is 100. The Cronbach alpha coefficient for the C-test was .65. Although the ideal Cronbach alpha coefficient should be above .7 (Pallant, 2016), in the present situation this value should not be a problem. It is common to have low Cronbach value with scales that have items fewer than ten (Pallant, 2016). The C-test has only 5 items (5 passages). Thus, it can be assumed that the C-test has internal consistency and is regarded as reliable in the current context.

The C-test was administered to about 150 participants. Participants were familiarised with the C-test format by giving them some examples with answers before they started. Those who did not complete all the five passages were excluded from the study. It was decided to exclude participants who scored below 40 from participating in the current study. The rationale for this decision was based on the findings of the pre-pilot study as the learners who scored below 40 in the C-test found it difficult to perform the oral narrative picture prompts. In addition, it was difficult for those participants to properly perform the dual task condition experiment. This means that performing these tasks is challenging for them. Thus, they were excluded from the current study.

Few students obtained scores above 60, thus they were also excluded. The total number of participants who were selected to participate in this study were 70. Later, four students did not complete all requirements of the data collection, so they were excluded. The total number of participants are 66 and their scores are within the limits of 40 and 60 in the C-test as indicated in Table 4.11 below.

Table 4.11. Participants' scores in the C-test

C-test	N	Min.	Max.	Mean	Std. Dev.	Std. Err
Scores Out of 100	66	40.00	60.00	48.72	5.34	.46

Several practical issues such as the time limit for data collection and the approaching exams period did not allow for the inclusion of further proficiency levels during data collection.

The Elicited Imitation Test

As the current study focuses on assessing linguistic knowledge, processing and production of L2, the EIT is regarded as a good candidate for assessing language ability and processing in oral modality. It has been argued that the EIT can tap into the procedural knowledge of L2 learners (Ellis, 2005). The EIT has been used recently in a number of studies to assess linguistic proficiency in oral modality (e.g., Ellis, 2005; Erlam, 2006; Gaillard &Tremblay, 2016; Ortega, Iwashita, Norris, & Rabie, 2002; Wu & Ortega, 2013) (see Section 2.8.2). However, limitations have been identified in this test. First, it has been argued that if incorrect repetition takes place, it cannot be decided whether breakdown occurs at the level of comprehension or production (Hood & Lightbown, 1987). However, researchers believe that the location of the breakdown is not important, because a proficiency test is supposed to assess the efficiency of both comprehension and production systems in L2 learners (Gaillard &Tremblay, 2016). Another shortcoming which was presented is the possibility that the EIT may only elicit rote repetition (McDade, Simpson, & Lamb, 1982).

The above points have been addressed by Erlam (2006) in his review of a number of previous studies on the EIT which reveals three important points that prove its validity. First, studies found that development of learners' linguistic knowledge affects their performance on the tests. That is, if this test entailed only rote repetition, the participants' scores would not be affected by development in their linguistic knowledge. Second, it was assumed that L2 learners perform better if they can establish semantic representation of the sentence which suggests that a certain level of processing occurs (Erlam, 2006). Third, it was also argued that L2 learners tend to correct sentences that have grammatical errors which supports the view that the EIT involves processing of sentences rather than merely parroting them (Erlam, 2006). Furthermore, it has been argued that the EIT relates to the efficiency of processing and short-term memory (Bley-Vroman & Chaudron, 1994). The researchers assumed that the EIT involves a systematic procedure that starts with receiving the sentences through the auditory system, then establishing representation of it through the comprehension system, and this representation is kept in the short-term memory while the sentence is prepared for production (Bley-Vroman & Chaudron, 1994). Finally, the sentence is produced through the production system (Bley-Vroman & Chaudron, 1994). This shows that the EIT seems to involve reconstructing rather than merely rote repetition. However, it has been argued that the EIT assesses automatized explicit knowledge rather than implicit knowledge (Suzuki & DeKeyser, 2015). That is, the linguistic knowledge that the EIT assesses is the speeded-up explicit knowledge rather than the implicit knowledge that does not involve awareness. This finding, nevertheless, is not conclusive as it also demonstrated that the EIT involves reconstructive processing and does not mainly draw on working memory.

The current study adopted the EIT of Wu & Ortega (2013) (see Appendix 11-A). The EIT was administered to participants individually. Nineteen sentences were selected with an increasing number of syllables, that is, 7-19 syllables. The first sentence includes seven syllables, whereas the last sentence consisted of 19 syllables. The sentences were spoken by a native speaker of English and recorded for the purpose of the current study. Each participant listened to the auditory sentences through speakers and their productions were audio-recorded on a Voice Record App on iPhone 6s plus. Participants were asked to repeat as much of the sentence as they could, and they were given only one chance to do so. Sentences were given scores ranged from 0-4 points, with 76 points as a maximum total score for the whole task (see Appendix 11-B). Each participant was given a maximum of four points for a perfect repetition (repeat the whole sentence correctly), three points for accurate content repetition, two points for changes that affected meaning (in content or form), one point for repetition of half of the sentence or less, and zero for a single-word repetition or failure to repeat anything (see Appendix 11-B). The Cronbach alpha coefficient for the EIT, which has nineteen items (nineteen sentences), was .92 which suggests that it is reliable as the ideal Cronbach alpha coefficient should be above .7 (Pallant, 2016).

Table 4.12. Participants' scores in the EIT

EIT	N	Min.	Max.	Mean	Std. Dev.	Std. Err
Scores Out of 76	66	12	66	32.44	13.05	1.14

The participants' scores in the EIT are summarised in Table 4.12 above. To establish comparability between the C-test and the EIT, both tests were calculated per 100. Then, the averages of the two tests were combined as demonstrated in Table 4.13 below. Running Pearson correlation coefficient showed a positive medium correlation between the scores of the C-test and EIT, $r = .36$, $p = .01$. The scores of the two tests were thus judged suitable to be combined to represent the overall language ability. This decision can be justified by the purpose of the current study which needs to include oral and written tests because the construct under investigation (self-monitoring) draws on explicit as well as implicit knowledge. I believe that combining these two tests would provide a more comprehensive measure of language ability.

Table 4. 13. Average of participants' scores in the C-test and the EIT

Combined averages of C-test and EIT	N	Min.	Max.	Mean	Std. Dev.	Std. Err
Scores' averages	66	29.38	71.42	45.44	9.89	.86

After recruiting the participants their proficiency was tested through the C-test the result of the test suggested that the participants came from a range of proficiency levels. However, it was decided to have only two proficiency groups in the current study. The reason for this decision was to have enough participants in each group. The low-level proficiency learners were excluded from the study since they seemed not able to produce enough output, and the advanced learners were excluded because there were only very few of them. The cut-off point between the two groups was decided based on the average of the combined tests (45) which was considered a reference point for classifying the two proficiency groups. That is, those who scored below 45 were classified as the Elementary group, and above this score was regarded as the Intermediate group. In other words, those who obtained 29 – 45 on the combined tests belonged to the Elementary group, whereas those who got 45 - 71 belonged to the Intermediate group as demonstrated in Table 4.14 below.

Table 4. 14. Proficiency groups

Proficiency groups	Combined Tests' Scores	N
Elementary	29 - 45	35
Intermediate	45 - 71	31

To make sure that these two groups were significantly different, a normality test (Kolmogorov-Smirnov) was carried out. The data showed that the two groups were significantly different, $D (132) = .096, p < .004$. Therefore, it is concluded that the two proficiency groups in the current study are significantly different from each other.

4.6.4 Working memory capacity tests

Measurement of individual differences in working memory capacity (WMC) is very important when investigating self-monitoring behaviour. According to the Perceptual Loop Theory (PLT), monitoring is supposed to be influenced by the processing constraints of working memory (Levelt, 1983) (see Section 2.10.3). The capacity to hold and manipulate resources can predict the way working memory, as a cognitive system, operates in relation to L2 processing and production (see Section 2.10.2). Temporary storage and processing are believed to relate to the two most important components of working memory which are the phonological loop and the central executive (see Section 2.10.1). The backward digit span test (BDST) has been used in a number of studies to investigate WMC (e.g., Kormos & Sáfár, 2008). This test can measure temporary storage and processing of information (William, 2012). The backward digit span task, which is used to measure students' working memory capacity in the current study, was adapted from Awwad's (2017) study (see Appendix 12).

This test is supposed to test not only the phonological short-term memory but also the capacity of complex verbal working memory including the functioning of the central executive, which is responsible for regulating attention (Gathercole, 1999; Hale, Hoeppner & Fiorello, 2002). Kormos and Sáfár (2008) argued that the BDST assesses temporary storage as well as manipulation of stored information (see Section 2.10.4).

As BDST is primarily a non-verbal test, it has been indicated that this test is language independent (Awwad, 2017). This suggests that performing the BDST reduces the possibility of learners' linguistic proficiency influencing their scores on WMC tests (Wright 2010). The BDST in this study were presented in L1 (Arabic) and L2 (English). Seven sequences of increasing numbers in Arabic and English were delivered to participants at the same pace. The first sequence included 3 digits and the last sequence consisted of 9 digits. The two versions consisted of different numbers that did not follow any patterns. The participants were required to repeat these sequences of increased digits backwards. The English and Arabic versions were counterbalanced between participants. Each participant was given three attempts for each sequence. Participants' WMC spans were determined based on the last sequence of digits they repeated successfully twice (Wright, 2010). In other words, if participants could not repeat the same sequence twice out of three attempts, their WMC would be the last sequence they managed to repeat twice.

Table 4.15. Participants' scores in the working memory capacity tests

Scores on WMC tests	N	Min.	Max.	Mean	Std. Dev.
BDST in L1 (Arabic)	66	3	7	4.80	1.05
BDST in L2 (English)	66	3	7	4.55	.94

The participants' results on the two WMC tests are summarised in Table 4.15 above. The data in Table 4.15 above shows that participants' scores in WMC in both English and Arabic ranged from three to seven-digit span which is almost similar. A moderate and positive correlation was found between L1 and L2 BDST scores, $r = .62$, $p = .000$. Considering the moderate correlation between L1 and L2 WMC tests, the scores of the L2 BDST were adopted. Osaka and Osaka (1992) argued that WMC is essentially language independent, and therefore it could be assumed that no difference existed between BDST in L1 or L2.

4.7 **Ethical procedure**

To ensure the ethical practice of research was followed, the present project applied the Reading University's Ethics Guidance. This step should be done prior to conducting any research study involving participants. The ethical procedure was reviewed and approved by the School Ethics Committee in the University of Reading. The researcher articulated the purpose and significance of the study, the voluntary nature of participation, and the data management practices. Participants signed a consent form and an information sheet which included description of the study. Each participant was given a copy of the signed consent form and another copy was kept with the researcher. The researcher told the participants that their privacy and confidentiality would be protected and that they were free to withdraw from the study at any time.

4.8 **Data collection procedure**

Data was collected on the campus of Taif University, in Saudi Arabia, over a period of four weeks during the first semester of the academic year 2017-2018. On the first week, meetings with several classes were arranged by the secretary of the department. The researcher attended a number of classes and briefed the students about the study and explained the procedures of data collection. The students were told that participation was voluntary and would not affect their marks in any module. The students who chose to take part in the study were asked to read and sign the information sheet and the consent form (see Section 4.7). The C-test was administered to about 150 students in a pen-paper format and they were given 30 minutes to complete the test. Students were required to write their mobile numbers so that the researcher can contact them and schedule an appropriate time for each participant to complete the second part of the data collection. All students who completed the C-test were given certificates of thanks signed by the researcher and the first supervisor. A WhatsApp group was then established to schedule suitable time for participants according to their availability. The researcher ensured that participants chose the appropriate time slots that would not clash with their lectures timetable. A reminder was sent to participants a night before their scheduled meeting to ensure that participants would come on time. A quiet room at the Department of Foreign Languages was booked for the researcher for a month to complete the data collection procedures and meetings with participants. Each participant was given a slot of 30 minutes and extra time was provided if needed. Tools of data collection included a laptop, a microphone, mobile phone with Voice Record Pro App (see Table 4.16 below).

Table 4.16. Summary of data collection tools

	Tools	Purpose
1	C-test	Assessing proficiency level
2	Elicited imitation test	Assessing proficiency level
3	BDST	Assessing working memory capacity
4	Narrative picture prompts	Eliciting speech in the two task conditions
4	PowerPoint software	Performing the single task condition
5	E-prime software	Performing the dual task condition

Each session started with a short chat with a participant. The working memory capacity tests (BDST) were conducted in Arabic and English. Each working memory test started with a trial to make sure that participants understood the instructions. Then, digits with increasing number sets (3-9) were recited to participants (see Section 4.6.4 above). Each participant was asked to repeat each set of digits backwards. While listening, the researcher did the scoring on the participant's task sheet. Each participant was assigned a sheet in Arabic and English. If participants failed to repeat any set of digits twice, they were given the score of the previous digits. Arabic and English working memory tests were counterbalanced to reduce the effects of order. After that, the Elicited Imitation test was played using a microphone and each participant was asked to repeat the sentence when the recording stopped. Each participant listened to the sentences with an increasing number of syllables and was instructed to repeat sentences as accurately as possible. If a participant could not repeat any more sentences, the test stopped at this point. Their performances were recorded and saved to OneDrive to score later. The maximum score that participants could reach in the EIT was 76 (see section 4.6.3).

Each participant was given a short break before completing the next sets of tasks (picture prompts description). Participant were given a minute to look at the pictures and plan their speech. The instructions were written in Arabic (see Appendix 8). For the single task condition, oral narrative picture prompts were shown using PowerPoint software. Each participant described two picture stories (The Storm and The Museum). Picture prompts were counterbalanced to reduce practice and order effects. In the dual task condition, each participant started by reading instructions of the experiment in Arabic and performed a practice trial which involved describing a different picture prompt and simultaneously performing the secondary task. Once participants completed the practice trial, they were asked to press the start button when they were ready to start the experiment. They were given one minute to read the instruction and plan their speech. After that, each participant was required to describe the picture prompts and at the same time perform the secondary task. Participants were asked to pay equal attention to the two tasks (the primary and the secondary tasks). Each secondary task was given a specific name and directly saved to the computer. Participants' oral performances were recorded on Voice Record Pro App.

4.9 Coding the measures of self-monitoring aspects

Once the data were transcribed using GoldWave software which allows for fast-forwarding, rewinding, and changing the playback speed. The transcribed speech samples were coded for measures of monitoring aspects. Twenty-seven measures were employed to tap into L2 self-monitoring behaviour including disfluency, repair types, temporal phases of repair, and accuracy. Some studies on L2 self-monitoring mainly focused on self-repair as a measure of self-monitoring behaviour (e.g., Gilabert, 2007). However, L1 research has examined further aspects of self-monitoring such as disfluency, temporal phases of repair, and ratio of error-correction. In an attempt to look at L2 self-monitoring from several perspectives, the current study examined different aspects of L2 self-monitoring, namely, disfluency, repair types, temporal phases of repair, and accuracy. The next sub-sections introduce a detailed account of the measures used in the current study.

To segment speech samples into units, the AS-unit was employed (Foster, Tonkyn, & Wigglesworth, 2000). An AS-unit has been defined as a single utterance that consists of an independent clause or sub-clause with any subordinate attached to it (Foster et al., 2000). In this study, an AS-unit boundary is marked by upright slash (|) and a clause boundary is marked by a double colon (::) (see Appendix 13). There were 132 speech samples collected from participants in the two task conditions. Pauses and temporal phases of repair were calculated using PRAAT software (Boersmal & Weenink, 2005). About 2 hours of oral performances were transcribed and coded as only the first 60 seconds of each speech sample was included. However, to ensure the reliability of transcription, about 10% of the data was rated by an expert researcher, and then the researchers discussed the differences until they reached agreement.

4.9.1 Measures of disfluency

A number of measures was used to examine disfluency such as filled pauses, silent pauses, repetition. In terms of filled pauses, all filled pauses (uh, err, umm) were calculated and divided by the total speech time (60) and multiplied by 60, following Kormos and Dénes (2004) as demonstrated in Table 4.17 below.

Table 4.17. Disfluency in the current study

Dimension	Measure	Definition
Disfluency	Filled pauses	The total number of filled pauses (uh, umm, err) divided by the total time of speech in seconds and multiplied by 60.
	Mid-clause silent pauses	The total number of mid-clause silent pauses over 0.25 second divided by the total time of speech in seconds and multiplied by 60.
	End-clause silent pauses	The total number of end-clause silent pauses over 0.25 second divided by the total time of speech in seconds and multiplied by 60.
	Repetitions	The total number of repetitions (i.e. words, phrases) was divided by the total time of speech in seconds and multiplied by 60.

L2 studies differentiated between silent pauses according to their locations (e.g., Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016; Tavakoli, 2011). That is, there are two types of silent pauses: silent pauses that occur within clause and at clause boundaries (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016; Tavakoli, 2011). It has been assumed that silent pauses within clause boundaries relate to difficulties at the Formulator level whereas silent pauses at clause boundaries relate to the Conceptualiser (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016). It has been recommended that researchers need to adopt a reliable threshold for calculating silent pauses (De Jong & Bosker, 2013; Kormos & Dénes, 2004). This study followed Kormos and Dénes's (2004) proposal of using the threshold of 0.25 seconds for measuring silent pauses. That is, the number of pauses over 0.25 was divided by the total time of speech, and multiplied by 60 as shown in Table 4.17 above. Repetition was also calculated by dividing the total number of repetitions by the total time of speech, and was multiplied by 60.

4.9.2 Measures of repair types

The coding scheme of repair types was based on Levelt's (1983) taxonomy later adopted by Kormos (1999). According to this classification, A-repair entails modifying the way in which an utterance is produced to become more appropriate or accurate in a particular context (Kormos, 1999). In D-repair, the message is abandoned and replaced with a different message (Kormos, 1999). Error-repair entails correcting a lexical, grammatical, or phonological error (Kormos, 1999). Lexical-repair involves correcting errors in phrases, idioms, preposition and derivational morphology (see Section 2.5.1). Grammatical-repair involves correcting grammatical error such as inflectional morphologies, auxiliaries and some prepositions as explained in Section 2.5.1. Phonological repair entails correcting an error in intonation, stress, phoneme, allophone, or allomorph aspects (Kormos, 1998) (for further details, see Section, 2.5.1). Rephrasing-repair was not employed in the current study for practical reasons. During the coding process, it was noted that coding lexical items in some utterances might be confusing. That is, some lexical corrections could fit into different categories, namely, E-repair (lexical-repair) or A-repair.

The literature shows two different approaches that have been used in coding lexical-repair. First, using retrospective interviews has been regarded as an accurate and reliable method of classifying lexical-repair (e.g., Kormos, 1998). Employing self-reporting protocol can reveal a speaker's intention when s/he produces a repair. Although this method appears to be effective in classifying lexical-repair, it might be difficult in practical terms to implement in complex experimental studies with large numbers of participants. The second approach was proposed by Levelt and Cutler (1983), to classify lexical-repair to either A-repair or E-repair. Their guidelines entailed checking the semantic relationship between error and repair. That is, this approach entails examining whether there is a clear error in the utterance. For instance, if the original utterance (the erroneous word) includes an error that needs to be corrected, and the repair provides a correction to this error, then it is classified as lexical-repair (Levelt & Cutler, 1983). In contrast, when the original utterance does not involve a real error, yet the use of the word is not appropriate or accurate in the given context, it is considered as an A-repair (Levelt & Cutler, 1983). Researchers agree that the central characteristic of A-repair is that the original utterance does not contain an error (Ooman & Postma, 2001). That is, the purpose of A-repair is to enhance the way messages are communicated rather than to correct an error. In terms of lexical-repair, it was argued that lexical-repair frequently entails contrast between error and repair in the utterance (e.g., the use of left versus right; vertical versus horizontal) (Levelt & Cutler, 1983), as demonstrated in example 1 below.

Example 1. Left to green - er, right to green (Levelt & Cutler, 1983, p. 207)

This example shows two contrasting words, 'left' and 'right' and indicates that the use of 'left' is erroneous. According to the guidelines of Levelt & Cutler (1983), the correction of this error needs to include the antonym of the erroneous word which is in this case 'right'. It has been assumed that there might be a mismatch between the original intention and the overt utterance (Levelt, 1983, p. 50). That is, a speaker may produce 'left' but the input that was generated in the preverbal plan is 'right', thus upon hearing the error in overt speech, it was corrected. It was suggested that repair of this type need to be classified as E-repair as it entails correcting an apparent error (Levelt, 1983; Levelt & Cutler, 1983). Furthermore, it has been explained that E-repair occasionally involves a rejection of the original utterance by producing "no" before the repair (Levelt & Cutler, 1983). That is, the erroneous utterance includes an error that needs to be corrected. Thus, the speaker usually rejects errors in utterances by using 'no.' It is worth mentioning that the pilot data of the present study included some examples where participants apologised for making an error using 'sorry' which was produced either before or after repair. There were also few examples where participants said (لا) in Arabic, which means 'no,' after they produced an error. As such, it was decided that a repair can be classified as lexical-repair when it entails correcting an erroneous element, whereas a repair can be considered as an A-repair when it adds more specification or elaboration to the speech (Levelt & Cutler, 1983).

This approach is useful in categorising lexical-repair especially for the category of E-repair (see Example 2). However, sometimes it becomes difficult to determine the type of repair based on this approach as in Example 3 below. The following examples were taken from the pilot of the present study.

Example 2. and they feel <u>exciting</u> uh <u>excited</u>

Example 3. Uh <u>the studen'</u> uh <u>the school</u> go to visit a museum

In Example 2, the speaker produced 'exciting,' which is wrong in this context, thus, it was replaced later with 'excited'. This repair is considered lexical-repair because it involves correcting an erroneous item. Example 3 is challenging since we are not certain why the speaker chose to replace 'student' with 'school'. The approach proposed by Levelt and Cutler

(1983) might not be of much help in this case. This repair could be an A-repair or a D-repair since a speaker can abandon one utterance to produce another. When raters disagreed on a certain repair, a third rater was consulted. If no agreement was reached, repair was categorised as Rest repair. Rest repair involves repair that did not fit into the above described categories. To have systematic coding of different types of repair, the total number of each type of repair was divided by the total time of speech and multiplied by 60 as indicated in Table 4.18 below.

Table 4.18. Measures of repair types

Dimension	Measure	Definition
Repair types	D-repair	The total number of D-repair divided by the total time of speech in seconds and multiplied by 60.
	A-repair	The total number of A-repair was divided by the total time of speech in seconds and multiplied by 60.
	E-repair	The total number of E-repair (lexical, grammatical, and phonological) divided by the total time of speech in seconds and multiplied by 60.
	Lexical-repair	The total number of lexical-repair divided by the total time of speech in seconds and multiplied by 60.
	Grammatical-repair	The total number of Grammatical-repair divided by the total time of speech in seconds and multiplied by 60.
	Phonological-repair	The total number of Phonological-repair divided by the total time of speech in seconds and multiplied by 60.

Raw numbers of repair were calculated in each speech sample and entered into an Excel sheet before entering the data into an SPSS file. Table 4.19 below shows the total number of repairs and the number of each type of repair in the current data. Rest repair includes unclassifiable repairs of which there were few in the current data. Table 4.19 below shows that there were 354 repairs in the 132 samples, which is about 2-3 repairs in each speech sample. It can be noted that the greatest number of repairs is Grammatical-repair, followed by A-repair, lexical-repair, D-repair, and the last category is Phonological-repair.

Table 4.19. Repair types in the current study

Number of repairs	A-repair	D-repair	E-repair	Lexical-repair	Grammatical-repair	Phonological-repair	Rest repair
354	75	59	220	70	112	38	5

This data only gives a general impression about repair occurrences in the current study. Chapter 5 presents statistical analyses of the data in relation to the research questions.

4.9.3 Measures of temporal phases of repair

Although measuring temporal phases of repair is useful in investigating self-monitoring behaviour, this aspect of monitoring has rarely been investigated in L2 studies. In Levelt (1983), phases of repair were measured by calculating the number of syllables in utterances.

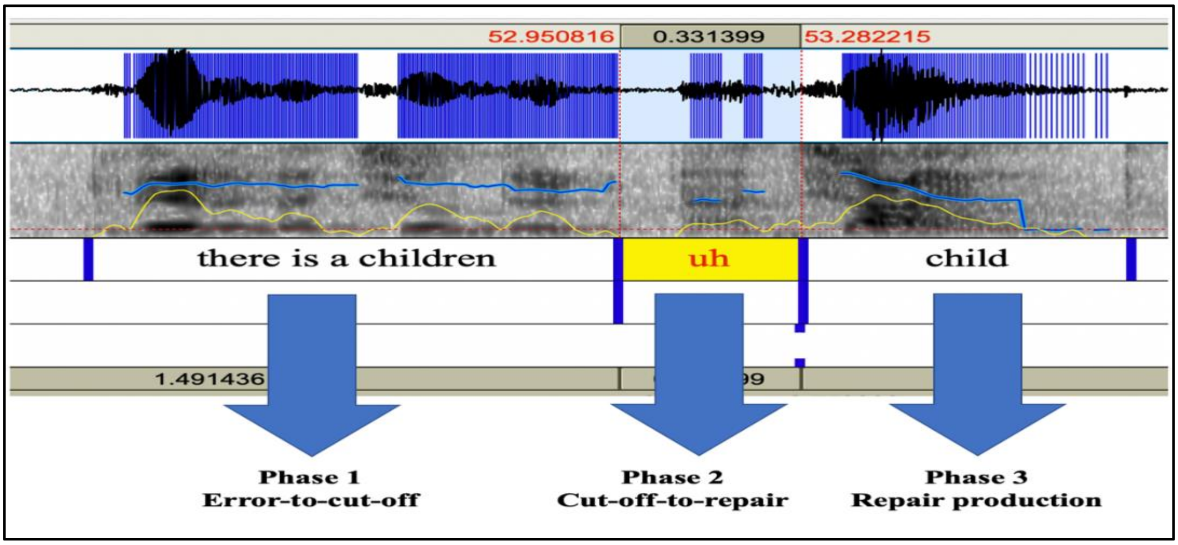


Figure 4.10. Calculating temporal phases of repair using PRAAT

However, for more accuracy, these phases have been measured in seconds in a number of recent studies (e.g., Kormos, 2000, Pillai, 2006). Temporal phases in the current study were measured is seconds using PRAAT (Boersmal & Weenink, 2005) as shown in Figure 4.10 above. According to Levelt (1983, 1989), there are three phases of repair: Error-to-cut-off (first phase), Cut-off-to-repair (second phase), and repair production (third phase) (see Section 2.4.2 for further details). Levelt's (1983) coding procedures were followed in dealing with these phases of repair as presented in the following sections.

a) **Error-to-cut-off (Phase 1)**

This phase includes producing an erroneous or inappropriate word or phrase. It is called reparandum or original utterance in Levelt's (1983) data. This interval shows the time needed to detect an error in one's speech (Levelt, 1983). However, this assumption is disputed, as discussed in Sections 2.3.3 and 2.4.3. The duration of this phase is calculated in seconds from the onset of erroneous or inappropriate word or phrase to the moment of interruption. The duration of each phase is put inside square brackets as can be seen in Figure 4.11 below.

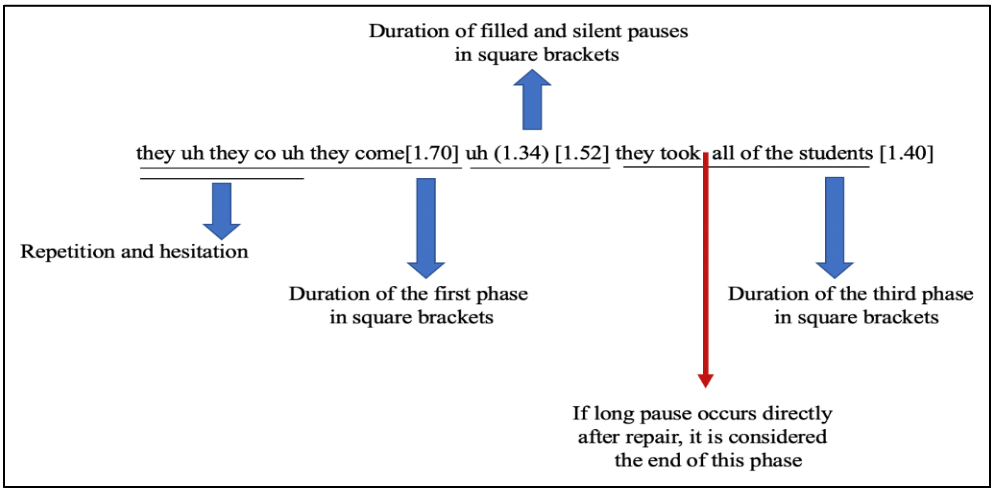


Figure 4.11. Procedure for coding repair phases

It was noticed during the coding procedure that the first phase of repair may involve hesitation or repetition, thus it was decided that any hesitation or repetition in the first phase of repair would be calculated as part of this phase since speakers might inspect their speech for potential erroneous or inappropriate element. In previous literature, there were no detailed guidelines in coding these phases of repair. One possible reason for lack of explanation of coding process could be that most monitoring studies used network description where majority of repair included replacing a colour or direction by another (e.g., Declerck & Kormos, 2012; Levelt, 1983). In such cases, repair might follow regular patterns which makes the coding of repair phases predictable. In the current study, it was difficult to decide in some repair cases the exact start and end of a particular phase such as the example in Figure 4.11 above. Thus, some specific rules were developed to make the coding procedure systematic as explained in this section.

b) **Cut-off-to-repair (Phase 2)**

This phase entails producing silent and/or filled pauses before executing the repair. This phase is calculated in seconds from the moment where speech flow stops to the moment of speech resumption. According to Levelt (1983), this phase (also known as the editing phase) is optional in repair structure, that is, some repairs might not include this interval (i.e. filled or silent pauses). This phase might serve the function of holding the floor while planning repair, and also sending notification to the listener that there is a processing problem (Levelt, 1983). If this phase included both filled and silent pauses, the duration of both pauses was calculated as one entity, as can be seen in Figure 4.11 above.

c) Repair (Phase 3)

This phase is defined as 'the correct version of what was wrong before' (Levelt, 1983, p. 44). Repair phase entails executing correction to erroneous or inappropriate utterances. Sometimes, speakers retrace utterances, but no real correction is provided. This phase is measured in seconds from the end of interruption to the end of repair. Some rules of coding this phase are found in Levelt's (1983) as illustrated below:

1- This phase usually starts at the point of the corrected utterance, but, in some cases, a speaker can retrace it to an earlier point (Levelt, 1983). For example, in Figure 4.11 above the speaker retraced to an earlier point which preceded the error in the original utterance, namely, 'they'. Although 'they' itself is not wrong, it was retraced by the speaker, so that it was included in this phase.

2- This phase should include modification to what was said before, otherwise it is 'covert repair' (Levelt, 1983). In Figure 4.11 above, the speaker changes 'they come' to 'they took'.

3- This phase ends with 'the first sentence boundary after the alteration' (Levelt, 1983, p. 45). That is, this phase ends at the first break after the correction. This rule seems vague given that speakers might pronounce some utterances in one go without a break, which suggests that this utterance is planned and produced as one chunk. Thus, it was decided that if an utterance is produced as a chunk, it is calculated as a part of the repair phase as demonstrated in Figure 4.11 above. That is, this phase includes the whole utterance 'they took all of the students' because it was articulated as one utterance. However, if a long pause (longer than 0.30 seconds) occurs directly after the correction, it is considered the end of repair phase. That is, if a long pause appears directly after 'they took,' the part before the pause would be regarded as phase 3, and the rest of the utterance will not be included, as illustrated in Figure 4.11 above.

4- If two or more repairs cluster in one utterance, 'each component of these multiple repairs will be analysed independently' (Levelt, 1983, p. 45). For instance, if a speaker corrects an utterance more than once, each correction is considered an independent repair.

These phases of repair were calculated per 60 seconds. That is, the total durations of a particular phase of repair (e.g., A-repair) was divided by the total number of this repair type (A-repair) per 60 seconds, as demonstrated in Table 4.20 below.

Table 4.20. Temporal phases of repair per 60 seconds

Dimension	Measure	Definition
D-repair	Error-to-Cut-off	The total length of Error-to-cut-off of D-repair divided by the total number of D-repair per 60 seconds.
	Cut-off-to-repair	The total length of Cut-off-to-repair of D-repair divided by the total number of D-repair per 60 seconds.
	Repair	The total length of Repair of D-repair divided by the total number of D-repair per 60 seconds.
A-repair	Error-to-Cut-off	The total length of Error-to-cut-off of A-repair divided by the total number of A-repair per 60 seconds.
	Cut-off-to-repair	The total length of Cut-off-to-repair of A-repair divided by the total number of A-repair per 60 seconds.
	Repair	The total length of Repair of A-repair divided by the total number of A-repair per 60 seconds.
Lexical-repair	Error-to-Cut-off	The total length of Error-to-cut-off of lexical-repair divided by the total number of lexical-repair per 60 seconds.
	Cut-off-to-repair	The total length of Cut-off-to-repair of lexical-repair divided by the total number of lexical-repair per 60 seconds.
	Repair	The total length of Repair of lexical-repair divided by the total number of lexical-repair per 60 seconds.
Grammatical-repair	Error-to-Cut-off	The total length of Error-to-cut-off of Grammatical-repair divided by the total number of Grammatical-repair per 60 seconds.
	Cut-off-to-repair	The total length of Cut-off-to-repair of Grammatical-repair divided by the total number of Grammatical-repair per 60 seconds.
	Repair	The total length of Repair of Grammatical-repair divided by the total number of Grammatical-repair per 60 seconds.
Phonological-repair	Error-to-Cut-off	The total length of Error-to-cut-off of Phonological-repair divided by the total number of Phonological-repair per 60 seconds.
	Cut-off-to-repair	The total length of Cut-off-to-repair of Phonological-repair divided by the total number of Phonological-repair per 60 seconds.
	Repair	The total length of Repair of Phonological-repair divided by the total number of Phonological-repair per 60 seconds.

An Excel sheet was created to calculate the number and duration of different phases of repair types before entering the data into an SPSS file. Table 4.21 below gives a summary of the number of repair phases in the current study.

Table 4.21. Summary of repair phases in the current study

Total number of repairs	**Phase 1** (Error-to-cut-off)	**Phase 2** (Cut-off-to-repair)	**Phase 3** (Repair production)	**Total number of phases**
354	354	329	354	1037

It can be noted in Table 4.21 above that phase 2 includes a lower number than phase 1 and 3. This shows that some repair did not include the second phase. About 25 repair cases did not include filled or silent pauses before repair production. That is, L2 learners produced repair directly after erroneous or inappropriate utterances. This is in line with Levelt's (1983) assumption that this phase is optional in the repair structure (see Section 2.4.2). This is theoretically interesting and will be further discussed in Chapter 6.

4.9.4 Measures of accuracy

Two measures of accuracy were used in the current study, namely, the ratio of error-correction and percentage of error-free clauses. The ratio of error-correction is supposed to give 'a better account of the accuracy of the monitor' as it relates to speakers' conscious decision to correct their errors (Oomen & Postma, 2002, p.168). The ratio of error-correction is calculated by dividing the number of repaired errors by the total number of errors in the speech sample. Some researchers calculate the ratio of error correction separately for each type of error, that is, lexical, grammatical and phonological errors (e.g., Declerck & Kormos, 2012). In the present study, the ratio of error correction for overall errors were calculated as demonstrated in Table 4.22 below.

Table 4.22. Measurement of accuracy

Dimension	**Measure**	**Definition**
Accuracy	Ratio of error-correction	The total number of repairs in speech sample divided by the total number of errors.
	Percentage of error-free clauses	The number of error-free clauses divided by the total number of clauses in speech sample multiplied by 100.

Accuracy was also expressed as a percentage of error-free clauses. That is, the number of clauses that has no error divided by the total number of clauses in speech sample multiplied by 100 as shown in Table 4.22 above. The higher the error-free clauses percentage is, the more accurate a text is anticipated to be. Assessing accuracy is useful in studying self-monitoring as self-monitoring behaviour is regarded as an attempt of L2 learners to promote accuracy in their speech (Kormos, 1999; Gilabert, 2007).

4.9.5 Inter-rater reliability

The researcher coded all of the repair in the data. In order to check the coding reliability, an expert rater coded 10% of randomly selected speech samples. In the case of disagreement between the two raters, a third rater was also consulted to ensure the reliability of the coding process. The repair that could not be classified was added to the category of Rest-repair (see Table 4.19 above). The two raters agreed on 83.43% of the repair. This percentage is high, comparable to the 73% of Levelt (1983), the 76% of Blackmer and Mitton (1991), and the 75% of Kormos and Declerck (2012). Concerning accuracy measures, 10% of the data were second rated by a native speaker of English who is also a researcher. The convergence between the two raters was 83%. The high inter-rater reliability achieved confirmed the consistency of the procedure of data coding. This allowed the researcher to proceed with data analysis. For samples of the coded data, see Appendix 13.

4.10 Data analysis

Data analysis was conducted using IBM Statistical Package, SPSS 24. Descriptive and inferential statistical tests were computed for each of the twenty-seven measures in relation to proficiency level and task condition (see Section 5.2). These measures include the four aspects of self-monitoring behaviour: disfluency (filled pauses, mid-clause, end-clause silent pauses, and repetition); Repair types (A-repair, D-repair E-repair, lexical-repair, grammatical-repair, and phonological-repair); Temporal phases of all repair types (Error-to-cut-off, Cut-off-to-repair, and Repair); and Accuracy (percentage of error-free-clauses and ratio of error-correction).

After checking all the assumptions, between-subject Multivariate Analysis of Variation (MANOVA) was run to identify significant differences between the two proficiency groups and task condition and to check any interaction effects between the two variables (see Section 5.3). This was followed by performing the two-way between-group Univariate Analysis of

Variance (ANOVA) to locate the significant effects and interaction effects of PL and TC on L2 self-monitoring aspects (see Section 5.4). Then, a fine-grained analysis was carried out to examine the impact of PL and TC on temporal aspects of L2 self-monitoring (i.e. temporal phases of repairs) (see Section 4.8). To examine the relationship between working memory capacity and self-monitoring aspects in the two task conditions, Pearson product-moment correlations were run on WMC and L2 self-monitoring aspects (see Section 4.9).

4.11 **Conclusion**

This chapter presented the aim and research questions of the current study. It has also described the procedures of developing the oral narrative picture prompts and dual task condition. This was followed by looking at tests of proficiency and working memory capacity. Detailed descriptions of data collection and coding procedures were provided. This chapter finished with a short description of data analysis. The next chapter will focus on a detailed discussion about analysis of this data.

Chapter 5: RESULTS

5.1 Introduction

The current study used a 2x2 factorial design with proficiency level (PL) and task condition (TC) as independent variables. PL and TC are between-participant variables each with two levels, with the participants belonging to one of these levels (see Section 4.4). Dependent variables are measures of disfluency, frequency of repair types, temporal phases of repair, and accuracy (see Section 4.9). This chapter has ten sections: Section 5.2 presents the statistical analysis of L2 self-monitoring aspects in relation to two proficiency levels and two task conditions. Section 5.3 discusses the use of the Multivariate Analyses of Variance (MANOVA) and presents the results of the MANOVA. This is followed by introducing the two-way between-groups univariate analysis ANOVAs and presenting the effects and interaction effects of PL and TC on L2 self-monitoring aspects. The findings are presented in relation to the first three research questions in Sections 5.5- 5.7. Section 5.8 introduces an explorative analysis carried out to examine the impact of PL and TC on temporal aspects of L2 self-monitoring (i.e. temporal phases of repairs). Then the results of the Pearson product-moment correlations between WMC and L2 self-monitoring aspects in the two task conditions are presented in Section 5.9. This chapter concludes by outlining the key findings of the analyses in relation to the research questions.

5.2 Preliminary analysis

Descriptive statistics are obtained for the dependent variables in relation to Proficiency Level (PL), and Task Condition TC). The means and standard deviations of L2 self-monitoring aspects of disfluency, frequency of the main repair types, and accuracy are presented in Table 5.23 with respect to the two levels of proficiency (Elementary and Intermediate), and the two task conditions (single and dual). Table 5.23 displays preliminary data about the effects of PL and TC on L2 self-monitoring aspects. The data in Table 5.23 demonstrate the differences between performances of the two proficiency groups, and the differences between performances in the two task conditions. The data of each independent variable will be considered separately. With regard to PL, the data indicate that the Elementary learners made more pauses (filled pauses, mid-clause silent pauses) and used repetition compared to the Intermediate learners. The Intermediate group, on the other hand, made more silent pauses at clause boundaries than the Elementary learners. The data, hence, point to two different patterns of disfluency which can be observed with respect to the oral performance of the two proficiency groups.

Table 5.23. Descriptive statistics of the dependent variables

Dimension	Measures	Proficiency level	Task Condition		Total (PL)
			Single	Dual	
			Mean (SD)	Mean (SD)	Mean (SD)
Disfluency	Frequency of filled pauses	Elementary	20.71 (7.82)	18.00 (5.73)	19.31 (6.91)
		Intermediate	13.28 (7.10)	17.93 (7.49)	15.53 (7.60)
		Total (TC)	17.11 (8.31)	17.97 (6.54)	17.54 (7.46)
	Frequency of mid-clause pauses	Elementary	14.24 (3.79)	15.00 (3.90)	14.63 (3.84)
		Intermediate	13.57 (4.77)	14.32 (3.78)	13.93 (4.30)
		Total (TC)	13.91 (4.27)	14.69 (4.06)	14.30 (4.06)
	Frequency of end-clause pauses	Elementary	8.76 (1.96)	8.57 (2.12)	8.75 (2.03)
		Intermediate	9.95 (2.80)	9.90 (2.46)	9.91 (2.62)
		Total (TC)	9.33 (2.46)	9.26 (2.33)	9.30 (2.38)
	Repetition	Elementary	2.06 (2.33)	2.86 (2.14)	2.47 (2.26)
		Intermediate	1.66 (2.12)	2.17 (1.62)	1.90 (1.90)

Dimension	Measures	Proficiency level	Task Condition		Total (PL)
			Single	Dual	
			Mean (SD)	Mean (SD)	Mean (SD)
		Total (TC)	1.86 (2.22)	2.55 (1.94)	2.20 (2.11)
Main repair types	Frequency of A-repair	Elementary	.21 (.48)	.53 (.74)	.37 (.64)
		Intermediate	.75 (1.14)	.83 (.99)	.79 (1.06)
		Total (TC)	.47 (.90)	.67 (.87)	.57 (.88)
	D-repair	Elementary	.15 (.44)	.53 (.56)	.34 (.54)
		Intermediate	.56 (.67)	.57 (.73)	.56 (.69)
		Total (TC)	.35 (.59)	.55 (.64)	.45 (.62)
	E-repair	Elementary	2.00 (1.30)	1.83 (1.58)	1.91 (1.44)
		Intermediate	1.47 (1.24)	1.33 (1.24)	1.40 (1.23)
		Total (TC)	1.74 (1.29)	1.61 (1.45)	1.67 (1.37)
Sub-categories of E-repair	Lexical-repair	Elementary	.65 (.69)	.67 (.93)	.66 (.81)
		Intermediate	.44 (.62)	.33 (.48)	.39 (.55)

Dimension	**Measures**	**Proficiency level**	**Task Condition**		**Total (PL)**
			Single	**Dual**	
			Mean (SD)	**Mean (SD)**	**Mean (SD)**
		Total (TC)	.55 (.66)	.52 (.77)	.53 (.71)
	Grammatical-repair	Elementary	1.06 (.92)	.92 (1.02)	.99 (.97)
		Intermediate	.69 (.86)	.70 (.92)	.69 (.88)
		Total (TC)	.88 (.90)	.82 (.98)	.85 (.94)
	Phonological-repair	Elementary	.29 (.58)	.22 (.42)	.24 (.49)
		Intermediate	.38 (.66)	.30 (.53)	.34 (.59)
		Total (TC)	.33 (.62)	.26 (.47)	.29 (.55)
Accuracy	Percentage of error-free clauses	Elementary	31.53 (21.01)	32.18 (14.85)	31.86 (17.97)
		Intermediate	43.36 (15.08)	53.17 (18.00)	48.11 (17.15)
		Total (TC)	37.26 (19.19)	41.72 (19.35)	39.49 (19.32)
	Ratio of error-correction	Elementary	.19 (.11)	.22 (.13)	.21 (.12)
		Intermediate	.23 (.14)	.28 (.17)	.25 (.16)

Dimension	Measures	Proficiency level	Task Condition		Total (PL)
			Single	Dual	
			Mean (SD)	Mean (SD)	Mean (SD)
		Total (TC)	.21 (.13)	.25 (.15)	.23 (.14)

That is, there was an increase in the frequency of filled pauses, mid-clause silent pauses, and repetition in the speech of the Elementary group, whereas the Intermediate group produced more end-clause silent pauses. This is illustrated in Figure 5.12 below.

Regarding repairs frequency, the Intermediate learners produced more A-repair and D-repair in their speech than the Elementary learners (details of repair types can be found in Section 4.9.2). However, the Elementary group produced a higher proportion of overall E-repair and E-repair sub-categories as compared to the Intermediate group. As such, the data suggest that there are two different patterns of repairs with respect to PL. That is, the frequency of E-repair sub-categories decreased, whereas A-repair and D-repair increased in the speech of the Intermediate group as compared to the Elementary group as can be seen in Figure 5.13 below.

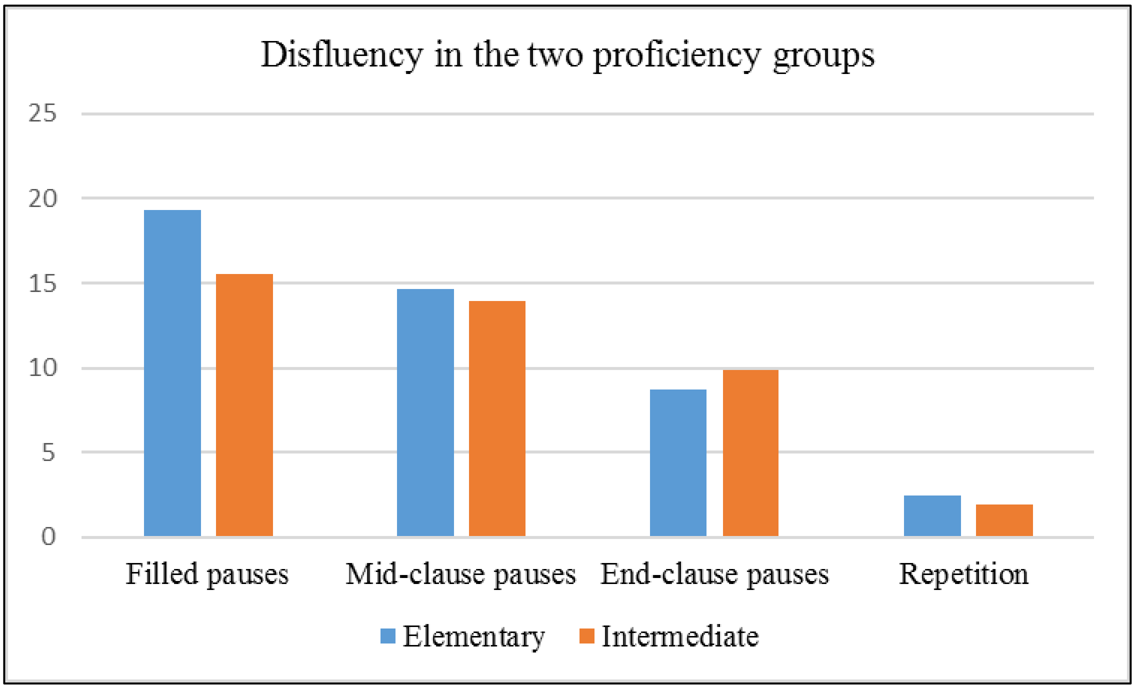


Figure 5.12. Disfluency in the two proficiency groups

With regard to accuracy measures, Table 5.23 illustrates that the Intermediate learners produced a higher rate of error-free clauses and more error-correction as compared to the Elementary learners.

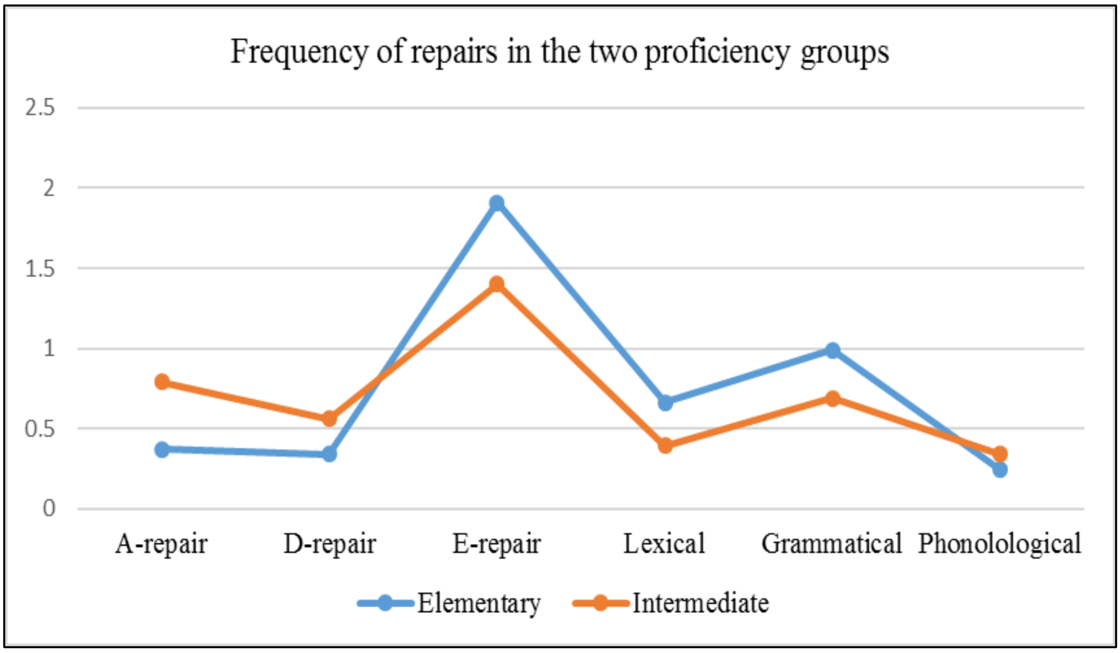


Figure 5.13. Repairs in the two proficiency groups

This result suggests that the accuracy rate was higher in the speech of the Intermediate group compared to the Elementary group. In short, the descriptive statistics in Table 5.23 show that there were differences between the two proficiency groups in terms of disfluency, repair frequency, and accuracy, with the Intermediate group being more accurate, making more A-repair and D-repair, and fewer E-repairs. Table 5.24 below shows some examples of repair in the current data.

Table 5.24. Repair types in the current data

Types of repair	Examples
Different-information-repair	Uh there uh /th/ the help comes to us 1.24|
Appropriateness-repair	|when the stor uh the thunderstorm comes|
Error-repair	|all of them was uh 0.28 all of them were |

Concerning the variances between the two task conditions in disfluency, repairs frequency, and accuracy, it can be noted in Table 5.23 that there were differences between the two task conditions. As for disfluency, it seems that there were small differences between the two task conditions as illustrated in Figure 5.14 below.

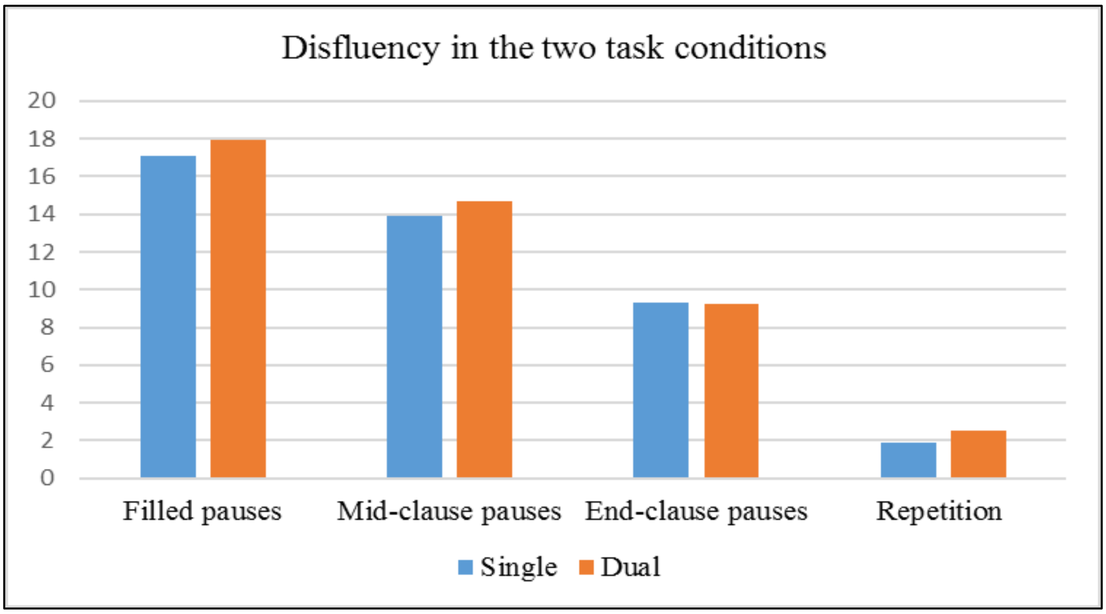


Figure 5.14. Disfluency in the two task conditions

The pattern of disfluency that emerged in relation to TC is similar to that observed in PL. That is, filled pauses, mid-clause silent pauses and repetition increased, whereas end-clause pauses decreased in the dual task as compared to the single task condition. This pattern suggests that these disfluency features (filled-pauses, mid-clause silent pauses, and repetition) showed a similar pattern in relation to PL and TC, whereas end-clause pauses did not conform to such a pattern. With respect to repairs, the data demonstrate that the frequency of A-repair and D-repair increased slightly, and E-repair sub-categories decreased in the dual task compared to the single task condition as illustrated in Figure 5.15 above. This suggests that the dual task may trigger more A-repair and D-repair on the one hand, and decrease E-repair sub-categories on the other. Similar to PL, it can be noted that two patterns of repair emerged in TC. That is, A-repair and D-repair increased, and E-repair sub-categories decreased in the dual task as compared to the single task condition. Finally, the results of the accuracy measures indicate that there was an increase rather than decrease in the percentage of error-free clauses and ratio of error-correction in the dual task rather than the single task condition. This increase in accuracy measures is surprising and in contrast to the predictions stated in Section 4.3. The data suggest that there were small differences between the two task conditions in terms of accuracy.

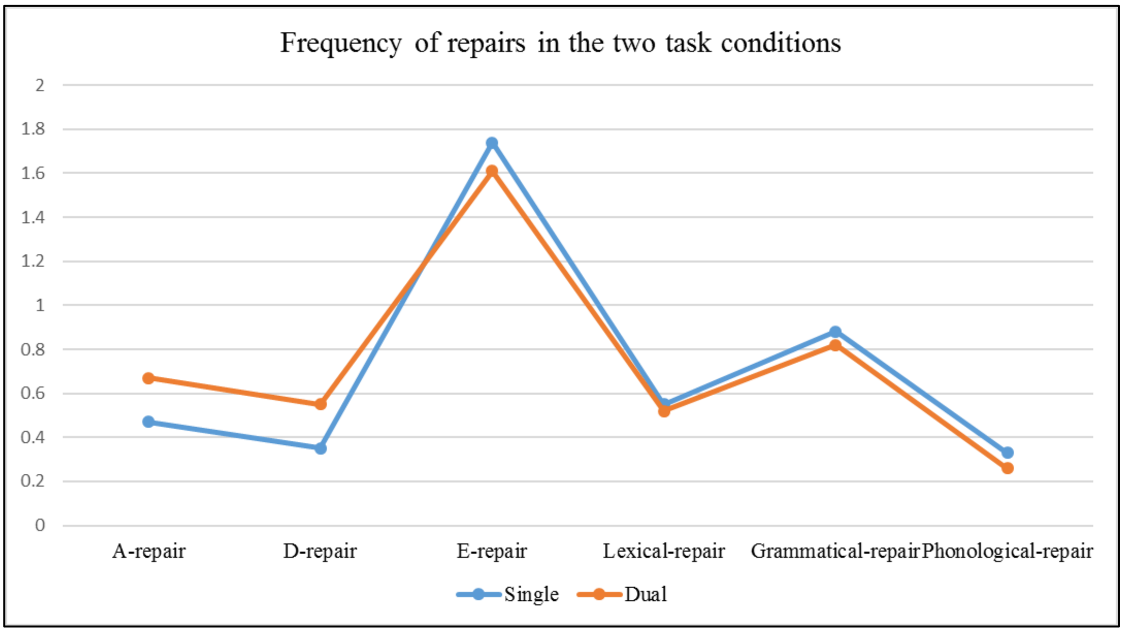


Figure 5.15. Repairs in the two task conditions

To summarize, the data in Table 5.23 suggest that there were some differences between disfluency, repair types, and accuracy under the two levels of PL. However, the differences between the repair, disfluency, and accuracy measures were smaller between the two task conditions. As such, these variances needed to be statistically analysed to determine whether such variations are meaningful. To do so, a multivariate analysis of variance (MANOVA) was used. Section 5.3 discusses the use of MANOVA, and the effects of the two independent variables on the combined measures of dependent variables (disfluency, repairs frequency, and accuracy).

5.3 Multivariate analysis

As noted in the previous section, the descriptive statistics revealed some differences between L2 learners' oral performances under the two levels of proficiency and task conditions. To explore whether there are statistically meaningful differences between the dependent variables under PL and TC, it was decided to run inferential statistics of multivariate analysis of variance (MANOVA). One advantage of MANOVA is that it can minimise Type 1 error which may occur when running a series of one-way ANOVAs. Type 1 error means that significant results may be obtained when in reality there are no differences between performances in different groups (Pallant, 2016). However, the use of MANOVA is controlled by a number of assumptions that need to be checked prior to proceeding with the analysis. These assumptions include sample size, normality, outliers, linearity, and multicollinearity. First, multivariate normality was checked through a linear regression, the results of Mahalanobis distances (the distance between a particular score and the rest of scores on the same measure “e.g., accuracy”) showed that only one measure had a score (34.25) that exceeded the critical value of 27.88 suggested for a 9-dependent variable test. Fortunately, MANOVA tolerates a few outliers, particularly when the score is not too high, and when there is a large sample size (Pallant, 2016). Hence, it was decided to keep the score of this participant because its value was not extreme, and there is a reasonable sample size in the current study. To check the assumption of equality of variance, Levene’s Test of Equality of Error Variances was used. It demonstrated that, for most measures, equal variances can be assumed. However, where any violation of equality of variances happened, a more conservative alpha level was set as recommended by Tabachnick and Fidell (2013). Furthermore, to ensure that the number of dependent variables submitted to MANOVA was compatible with the current sample size (sixty-six participants), it was important to have more cases than dependent variables in each cell (Pallant, 2016). In the current study, there are two independent variables with two levels for each. By dividing the number of participants (sixty-six) by number of cells (four cells), the outcome is 16.5, which represents the maximum number of dependent variables that can be submitted to MANOVA. However, in the current study, the number of dependent variables to be submitted to MANOVA (9 dependent variables) is far less than 16. Therefore, it can be safely assumed that the data set conforms to all assumptions listed above. To avoid duplicating the data, the overall number of E-repair was used to represent E-repair sub-categories (lexical-repair, grammatical-repair, and phonological-repair). The multivariate analyses of variance (MANOVA), hence, will be conducted using nine dependent variables of disfluency, repair frequency, and accuracy:

1. Frequency of filled pauses (disfluency)
2. Frequency of mid-clause silent pauses (disfluency)
3. Frequency of end-clause silent pauses (disfluency)
4. Frequency of repetition (disfluency)
5. Frequency of A-repair (repair frequency)
6. Frequency of D-repair (repair frequency)
7. Frequency of E-repair (repair frequency)
8. Percentage of error-free clauses (accuracy)
9. Ratio of error correction (accuracy)

The two independent variables were entered to MANOVA as between-subject factors as presented in Table 5.25 below. Effect sizes were calculated to assess the magnitude of the obtained effects. Following Cohen's (1988) guidelines, eta squared values at .01 are regarded as small, .06 as medium, and .14 as large effect size. To reduce the chance of Type 1 error, Bonferroni adjustment was applied to the analyses. That is to say, alpha level .05 is divided by 2 (the number of comparisons at each level). Thus, the new alpha level for the main effects of PL and TC on the composite dependent variable is .025. Likewise, the alpha level for interaction effects is set at .0125 (by dividing .05 by four). The results are presented in Table 5.25 below.

Table 5.25. Results of multivariate analysis of variance (MANOVA)

Effect	Wilks' Lambda Value	F	Sig.	Partial Eta Squared
Proficiency	.699	5.738	.000*	.301
Task Condition	.903	1.425	.185	.097
Proficiency x Task Condition	.842	2.498	.012**	.158

*p < .025, **p < .0125

Table 5.25 presents the results of the MANOVA in terms of the effects of PL and TC on the composite dependent variable. The analysis reveals a statistically significant difference between levels of proficiency on the combined dependent variable, F (5, 120) = 5.738, p = .000; Wilks Lambda = .699; partial eta squared (η^2) = .301. The results suggest that there was a significant difference between the Intermediate and Elementary groups in terms of L2 self-monitoring. The effect size associated with this difference is large suggesting that about 30.1 per cent of variances in the composite dependent variable can be explained by PL. However, a significant result was not observed for task condition, F (1, 120) = 1.425, p = .185; Wilks Lambda = .903; η^2 = .097. In terms of the interaction effects, a significant interaction effect was found between PL and TC, F (2, 120) = 2.498, p = .012; Wilks Lambda = .842; η^2 = .158. This suggests that L2 oral performance may have been affected by an interaction effect from the two variables. The effect size associated with this result is large, which suggests that about 15.5 per cent of variances in the combined dependent variable can be explained by the interaction effects between the two variables. These results imply that further analyses are needed to locate significant variances in L2 self-monitoring aspects in relation to PL and TC.

In other words, the results of the MANOVA allowed a series of two-way between-groups univariate analyses of variance (ANOVAs) to examine the effects and the interaction effects of PL and TC on the dependent variables presented above. Section 5.4 presents the findings of ANOVAs in relation to the first three research questions.

5.4 Two-way Between-Groups ANOVAs

In the previous section, the results of the MANOVA revealed a significant effect for PL on the combined dependent variable. There was also a significant interaction effect between PL and TC on the composite dependent variable. To investigate the effects and interaction effects of PL and TC on L2 self-monitoring aspects, a series of two-way between-groups ANOVAs was administered. As was the case in the MANOVA above, the alpha level was adjusted at .025 for the main effects of PL and TC (see Table 5.26 below). With regard to the interaction effects, the alpha level was adjusted at .0125 by diving .05 by the total number of levels (four). Where significant results were obtained, effect sizes of partial eta squared were assessed using Cohen's (1988) guidelines which provide effect size statistics by calculating the difference between groups in terms of standard deviation units. Cohen (1988) proposed the following standards to interpret the strength of Cohen's d effect size statistics: a) below .01 is a small effect size; b) .06 is medium; and c) above .14 is a large effect size. The results of the Two-way Between-groups ANOVAs displayed in Table 5.26 below are used to answer the first three research questions in terms of the effects of PL and TC on L2 self-monitoring aspects in the Sub-sections 5.5, and 5.6 respectively. This is followed by presenting the answer to the third research question with respect to the interaction effects of PL and TC on L2 self-monitoring in the Sub-section 5.7. These three sub-sections present the analyses of the first three research questions with respect to the effects and interaction effects of PL and TC on disfluency, repair frequency, and accuracy. This is followed by an exploratory analysis of the effects and interaction effects of PL and TC on the temporal phases (Error-to-cut-off, Cut-off-to-repair, and repair) of repair types (A-repair, D-repair, E-repair sub-categories); this completes the answers to the first three research questions in Section 5.8.

Table 5.26. Two-way Between-groups Analyses of Variance (ANOVAs)

Measures	Proficiency Level			Task Condition			PL*TC		
	F	Sig.	η^2	F	Sig.	η^2	F	Sig.	η^2
Filled pauses	9.282	.003*	.068	.626	.430	.005	8.954	.003**	.065
Mid-clause pauses	.894	.346	.007	1.136	.288	.009	.000	.993	.000
End-clause pauses	8.011	.005*	.059	.012	.931	.000	.008	.929	.000
Repetition	2.277	.134	.017	3.260	.073	.025	.161	.689	.001
A-repair	8.014	.005*	.059	1.823	.179	.014	.632	.428	.005
D-repair	4.668	.033	.035	3.350	.070	.026	3.207	.076	.024
E-repair	4.744	.031	.036	.407	.525	.003	.004	.947	.000
Lexical-repair	4.818	.030	.036	.117	.733	.001	.250	.618	.002
Grammatical-repair	3.250	.074	.025	.158	.692	.001	.225	.636	.002
Phonological-repair	.962	.328	.007	.375	.541	.003	.029	.866	.000
Error-free clauses	29.237	.000*	.186	2.969	.087	.023	2.275	.134	.017
Ratio of error-correction	3.170	.077	.024	2.551	.113	.020	.142	.707	.001

*p = .025, **p = .0125, PL df (1, 128), TC df (1.28), PL*TC df (1.128)

5.5 Research Question 1: Effects of PL on L2 self-monitoring

Research Question 1 investigated the effects of PL on L2 self-monitoring behaviour in terms of disfluency, frequency of repair types, temporal phases of repairs, and accuracy. There are two proficiency groups in this study: Intermediate and Elementary groups. The first research question is divided to four sub-questions. This section addresses three sub-categories of this research question (disfluency, frequency of repair types, and accuracy). The effects of PL on temporal phases of repairs is addressed in Section 5.8.

5.5.1 Effects of PL on disfluency

It was hypothesised that the Elementary learners would pause more frequently (filled pauses, mid-clause pauses), and produce more repetitions compared to the Intermediate learners. The Intermediate learners, on the other hand, were predicted to produce more silent pauses at clause boundaries compared to the Elementary learners. Four measures were employed to investigate the influence of PL on disfluency (filled pauses, mid-clause silent pauses, end-clause silent pauses, and repetition). The results in Table 5.26 above show that a significant difference ($F = 9.28$, $p =. 003$, $\eta^2 = .068$) was observed for the number of filled pauses with the Elementary group producing significantly more filled pauses ($M = 19.31$, $SD = 6.91$) than the Intermediate group ($M = 15.53$, $SD = 7.60$). The effect size of the variation is moderate and indicates that PL accounts for about 6.8 per cent of variances in filled pauses. This result suggests that the Elementary learners paused more frequently compared to the Intermediate learners. End-clause silent pauses, on the other hand, increased significantly ($F =8.011$, $p =. 005$, $\eta^2 = .059$) in the speech of the Intermediate group ($M = 9.91$, $SD = 2.62$) compared to the Elementary group ($M = 8.75$, $SD = 2.03$). It should be noted that the effect size of this variation is moderate, which means that around 5.9 per cent of differences in end-clause pauses can be explained by PL. This result suggests that the Intermediate group was distinguished by pausing at clause boundaries compared to the Elementary group. Although the results showed that the number of mid-clause silent pauses and repetition was lower in the Elementary group than the Intermediate group, the difference did not reach statistical significance. This result indicates that repetition and mid-clause pausing behaviour was not different in the two proficiency groups. Thus, the hypothesis is partially confirmed with the significant differences between the two proficiency groups in terms of filled pauses and end-clause silent pauses. These results suggest that two patterns of disfluency emerged in relation to proficiency. That is, the speakers at elementary level made frequent use of filled pauses and few clause-final pauses when they speak in L2. However, as proficiency develops, L2 intermediate learners tend to make fewer filled pauses, and more clause-final pauses as compared to elementary learners. The emergence of these disfluency patterns in relation to proficiency level suggest potential changes in L2 monitoring behaviour which will be further discussed in Section 6.2.1.

5.5.2 Effects of PL on repair types

It was predicted that PL would affect the frequency of repair types (A-repair, D-repair, E-repair, lexical-repair, Grammatical-repair, and Phonological-repair). That is, it was expected that L2 Intermediate learners would produce a greater frequency of A-repair, and a smaller numbers of E-repair sub-categories (lexical-repair, grammatical-repair, and phonological-repairs) compared to the Elementary learners. No hypothesis could be formulated for D-repair. The results of the impact of PL on repair types (A-repair, D-repair, and E-repair sub-categories) are presented in this sub-section. The results in Table 5.26 above demonstrate that the Intermediate learners produced significantly more A-repair (M = 0.79, SD = 1.06), than the Elementary learners (M = 0.37, SD = .64), with the difference reaching a statistically significant level (F = 8.014, p =. 005, η^2 = .059). The effect size associated with the difference is moderate, suggesting that about 5.9 per cent of variances in A-repair can be explained by PL. This result suggests that the Intermediate learners produced substantially more A-repair compared to the Elementary learners. As for D-repair, Table 5.26 shows that the difference between the two groups was approaching statistical significance (F = 4.67, p =. 033, η^2 = .035) at the new adjusted alpha level .025. That is to say, the proportion of D-repair slightly increased in the speech of the Intermediate learners (M = 0.56, SD = .69) compared to the Elementary learners (M = 0.34, SD = .54), but this increase was not significant. This finding indicates that the two proficiency groups made almost the same proportion of D-repair in their speech. Likewise, it was illustrated in Table 5.26 that the variance in the overall number of E-repair was approaching significance (F = 4.67, p =. 033, η^2 = .035). The Intermediate group produced a smaller overall number of E-repair in their speech (M = 1.40, SD = 1.23) compared to the Elementary group (M = 1.91, SD = 1.44), but the difference was not statistically significant. The finding shows that the differences between the two proficiency groups was also approaching significance in lexical-repair (F = 4.82, p =. 03, η^2 = .036), with the Intermediate group producing fewer lexical-repair as compared to the Elementary group. No significant differences were observed between the two groups with respect to Grammatical-repair and Phonological-repair. Therefore, the hypothesis is partially confirmed with the meaningful increase of A-repair in the Intermediate group compared to the Elementary group. The data in this section suggest that with the development of proficiency level, repair frequency might not greatly change. However, L2 learners at intermediate level appear to have an increased interest in modifying inappropriate aspects of their utterances, which is significantly different from the behaviour of the Elementary group. This increase in A-repair will be explained in relation to previous research and the Perceptual Loop Theory in Section 6.2.2.

5.5.3 Effects of PL on accuracy

The hypothesis for this sub-category of RQ1 stated that the Intermediate learners would produce a higher percentage of error-free clauses and higher ratio of error correction than the Elementary learners. The results in Table 5.26 above show that the percentage of error-free clauses increased significantly ($F = 29.24$, $p =. 000$, $\eta^2 = .186$) in the speech of the Intermediate group ($M = 48.10$, $SD = 17.15$) compared to the Elementary group ($M = 31.86$, $SD = 17.97$). It can be noted that the effect size of this result was considerable, suggesting that PL accounts for about 18.6 per cent of the variances in accuracy. This finding indicates that the Intermediate learners produced more accurate speech in terms of the percentage of error-free clauses compared to the Elementary group. In the case of the ratio of error-correction, the results indicate that the ratio of error-correction was small in the two groups: the Intermediate ($M = 0.25$, $SD = 0.16$), and the Elementary group ($M = .23$, $SD = .25$). The difference between the two groups was not significant. Thus, the hypothesis is partially confirmed in terms of the effects of PL on accuracy. The data suggest that with the development of L2 proficiency, L2 Intermediate learners corrected the same amount of errors, but their speech became significantly more accurate.

To outline the effects of PL on L2 self-monitoring aspects, the results show that PL had a meaningful effect on certain aspects of disfluency, repair types, and accuracy. The data in this section was in line with the hypotheses suggesting that with the development of L2 proficiency from elementary to intermediate levels, certain disfluency patterns emerge. That is to say, there was a shift from a pattern of frequent filled pause and few clause-final pauses in the elementary level to a pattern of less filled pauses and more clause-final pauses in the intermediate proficiency level. The data also show that as proficiency develops, repair frequency may not change significantly. The Intermediate learners, nonetheless, seem to have greater interest in modifying inappropriate aspects of their utterances compared to the Elementary learners. The data also indicate that with the development of proficiency, speech of L2 learners becomes more accurate. The data suggest that the development of proficiency has an effect on monitoring behaviour of L2 learners, leading to changes in monitoring processes and foci. This will be further discussed in Section 6.2.4.

5.6 Research Question 2: Effects of TC on L2 self-monitoring

One aim of the present study was to explore whether a secondary task affects L2 self-monitoring behaviour. The second research question investigated the effects of task condition on L2 self-monitoring in terms of disfluency, repair types, temporal phases of repair, and accuracy. It was expected that the dual task condition would have detrimental effects on L2 self-monitoring aspects. Task condition was operationalised in terms of whether the learners performed the task under single or dual task conditions. The following sub-sections present the findings in terms of the effects of TC on disfluency, repair types, and accuracy. The effects of TC on temporal phases of repair are presented in Section 5.8.

5.6.1 Effects of TC on disfluency

The findings of the effects of TC on disfluency measures are presented in this sub-section. It was predicted that the dual task condition would increase disfluency (filled pauses, mid-clause silent pauses, repetition, and end-clause pauses) in the speech of L2 learners as compared to the single task condition. The results presented in Table 2.26 above show that TC did not have a statistically significant influence on the disfluency measures employed in the current study (filled pauses, mid-clause silent pauses, end-clause silent pauses, and repetition). The hypothesis in this section is not confirmed as the data show that L2 disfluency was not affected by the increased demand of the task.

5.6.2 Effects of TC on the main repair types

It was predicted that dual task condition would have a detrimental effect on repair frequency. That is to say, it was expected that repair frequency would decrease as a result of performing the dual task condition compared to the single task conditions. The results of the ANOVAs in Table 5.26 above show no significant differences between repair types in the two task conditions (single and dual task conditions). Despite the fact that there was a slight increase in D-repair frequency during the dual task condition (M = .35, SD = .59) compared to the single task condition (M = .55, SD = .64), the difference did not reach a statistically significant level ($F = 3.350$, $p =. 070$, $\eta^2 = .026$). Thus, the hypothesis can be rejected as the data suggest that the increased demand of task condition did not have an effect on repair frequency.

5.6.3 Effects of TC on accuracy

According to the hypothesis, the dual task condition would reduce the percentage of error-free clauses and the ratio of error correction. The findings in Table 5.26 above show there were no significant differences between learners' performance in the two task conditions in terms of the percentage of error-free clauses and ratio of error correction.

To outline the findings in this sub-section, the data unexpectedly demonstrate that TC did not have significant effects on disfluency, repair types, and accuracy. This suggests that self-monitoring behaviour of L2 learners was not affected by the increased demand of task condition. The findings in this section have important implications about the extent to which cognitive resources are employed in L2 self-monitoring which will be discussed in Section 6.3.

5.7 Research Question 3: Interaction effects of PL and TC on L2 self-monitoring

The third research question investigated the interaction effects of PL and TC on L2 self-monitoring behaviour in terms of disfluency, repair types, and accuracy. It was expected that the higher proficiency participants would perform better than the lower proficiency participants under the dual task condition. The following sub-sections present the interaction effects of PL and TC on disfluency, frequency of repair types, and accuracy.

5.7.1 Interaction effects of PL and TC on disfluency

This sub-section reports on the interaction effects of PL and TC on disfluency measures: filled pauses, mid-clause silent pauses, end-clause silent pauses, and repetition. As can be seen from Table 5.26 above, there was a significant effect of PL on filled pauses ($F = 9.282$, $p = .003$, $\eta^2 = .068$), but no main effect of TC was observed. The results show that the impact of PL on filled pauses was significantly mediated by dual task demands ($F = 8.954$, $p = .003$, $\eta^2 = .065$). However, to accurately interpret the interaction effect, it was decided to conduct follow-up tests. One way to do this is by splitting the data file into two groups according to one independent variable, and then to conduct one-way ANOVA to explore the impact of the other variable (Pallant, 2016). For the purpose of the present analysis, the data was split into two groups according to TC, to investigate the differences between the two proficiency groups within each task condition (single versus dual). One-way ANOVA was conducted to investigate the effects of PL on filled pauses in each task condition separately as indicated in Table 5.27 below.

Table 5.27. One-Way ANOVA: The effects of TC on filled pauses

PL	Disfluency	Single			Dual		
		F	Sig.	d	F	Sig.	d
Elementary Intermediate	Filled pauses	16.248	.000	.20	.002	.967	--

*P < .05

The data show that the Intermediate group made significantly fewer filled pauses than the Elementary group in the single task; no differences were observed between the two groups in the dual task (see descriptive data in Table 5.23). These findings suggest that the single task condition triggers variations between the two proficiency groups, whereas these differences disappear when they perform under the dual task condition. Another follow-up test was conducted on proficiency groups to examine variations within each proficiency group in the two task conditions. That is, the data file was divided into two proficiency groups, and one-way ANOVA was conducted on each proficiency group separately, with TC as an independent variable, as shown in Table 5.28 below.

Table 5.28. One-Way ANOVA: The effects of PL on filled pauses

PL	Disfluency	Elementary			Intermediate		
		F	Sig.	d	F	Sig.	d
Single Dual	Filled pauses	2.751	.102	--	6.302	.015	.095

*P < .05

The data show that there was no significant difference between the performances of the Elementary group in the two task conditions in terms of frequency of filled pauses. However, the data indicate that the Intermediate group made significantly more filled pauses in the dual task than in the single task condition; this is illustrated in Figure 5.16 below (descriptive data can be found in Table 5.23 above).

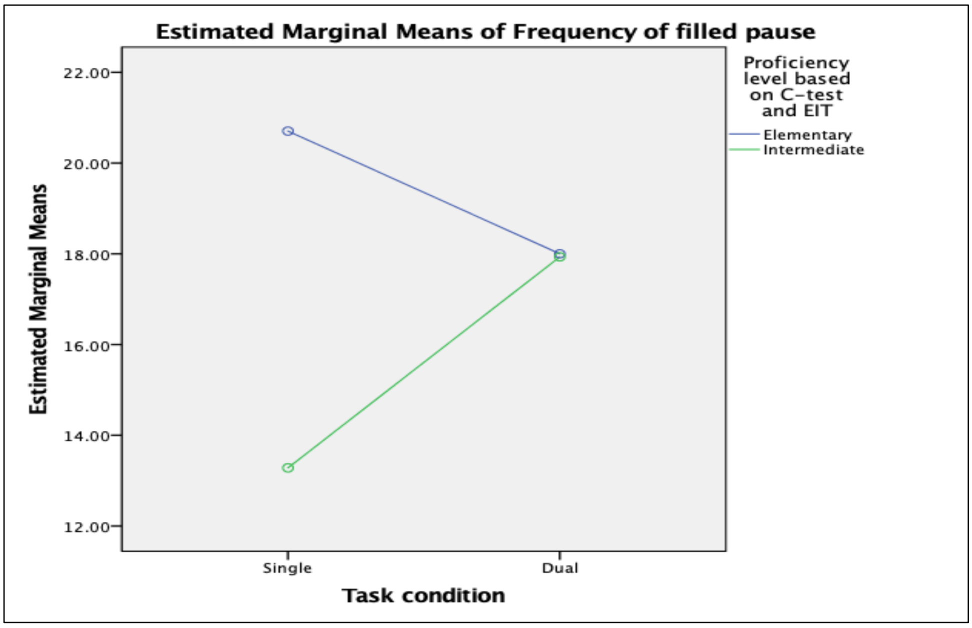


Figure 5.16. Interaction effects of PL and TC on filled pauses

This suggests that the Elementary group may find it equally demanding to speak under the two task conditions as their filled pauses remain the same in the two task conditions. The Intermediate learners, on the other hand, may find the dual task condition more demanding than the single task condition, and this might be the reason for producing more filled pauses in this task condition compared to the single task condition (for further discussion, see Section 6.4).

Section 5.5.1 indicated that PL had a significant impact on end-clause silent pauses (F = 8.011, p = .005, η^2 = .059), but TC did not have a meaningful impact on end-clause pauses. Similarly, the results show no meaningful interaction effect of PL and TC on end-clause pauses. This result suggests that the substantial increase in the end-clause silent pauses in the speech of the Intermediate group was not mediated by TC. Concerning the other disfluency measures (mid-clause silent pauses, and repetition), the results show no main effects or interaction effects of PL and TC on either of these measures. To conclude this section, the examination of the interaction effects of PL and TC on disfluency measures reveals that the significant difference between the two proficiency groups in terms of filled pauses was mediated by dual task demands. However, there was no interaction effects of PL and TC on other measures of disfluency.

5.7.2 Interaction effects of PL and TC on repair types

This sub-section discusses whether there were interaction effects of PL and TC on repairs (A-repair, D-repair, and E-repair sub-categories). Table 5.26 above presented the interaction effects of PL and TC on repairs. As can be seen in Table 5.26, PL had a significant effect on A-repair. That is, there was a significant increase in A-repair in the speech of the Intermediate learners rather than the Elementary learners. However, the data show no meaningful impact of TC, or interaction effects between PL and TC on A-repair. This result suggests that the substantial increase of A-repair in the speech of the Intermediate group was not mediated by TC. Similarly, the data show no interaction effects of PL and TC on D-repair and E-repair sub-categories.

5.7.3 Interaction effects of PL and TC on accuracy

This sub-section displays the interaction effects of PL and TC on the accuracy aspect of L2 self-monitoring. The results presented in Table 5.26 indicate that the percentage of error-free clauses significantly increased ($F = 29.24$, $p =. 000$, $\eta^2 = .186$) in the speech of the Intermediate learners compared to the Elementary learners (see Sub-section 5.5.3). However, the data demonstrate no meaningful impact of TC on the percentage of error-free clauses, or interaction effects on the percentage of error-free clauses. This finding suggests that the greater increase of accuracy in the speech of the Intermediate learners was not mediated by TC. Concerning the effects of the PL and TC on the ratio of error correction, the results indicate that there is no main effect or interaction effects of PL and TC on the ratio of error-correction in the current data.

To summarise the results of the interaction effects of PL and TC on L2 monitoring aspects, the data show one meaningful interaction effect of PL and TC on frequency of filled pauses. This suggests that the single task condition triggers great differences between proficiency groups in terms of filled pauses, and that these differences disappear when they perform in the dual task condition. It also shows that the Intermediate learners produced more filled pauses in the dual task compared to the single task condition. As the present study aims to systematically explore the temporal phases of L2 repair, a fine-grained analysis is carried out in the next section to explore the extent to which PL and TC influence the temporal phases of L2 repair.

5.8 Exploratory analysis of L2 self-monitoring

One of the purposes of the current study is to explore the extent to which L2 self-monitoring processes are affected by levels of L2 proficiency (Elementary and Intermediate). Another aim of the study is to explore whether the task condition (single versus dual tasks) affects L2 monitoring behavior. The previous Sections (5.5- 5.7) present the findings of the effects and interaction effects of PL and TC on L2 self-monitoring behaviour in terms of disfluency, repair types, and accuracy. With the aim of exploring further aspects of L2 self-monitoring behaviour, this section investigates the effects of PL and TC on the temporal phases of L2 self-monitoring which entail the duration of repair phases: Error-to-cut-off (first phase), Cut-off-to-repair (second phase), and Repair (third phase) (see Sections 3.8.3 and 4.9.3 for more details). The rationale for this exploratory analysis is that examining the temporal phases of repair can reveal important information about the two main functions of monitoring (see Section 2.4). That is, examining these aspects can show the duration of detecting an error or inappropriateness in different task conditions and proficiency levels. It can also reveal the duration of planning and making correction or adjustment. These aspects of monitoring can uncover important information about the development of learners' speech processes and how these processes can influence the functioning of self-monitoring (see further discussion in Sections 6.2.3 and 6.3). Moreover, the temporal phases of repairs have rarely been investigated in L2 research. To the best of my knowledge, only one study (i.e. Kormos & Declerck, 2012) has examined the effects of the dual task paradigm and proficiency levels on a particular phase of L2 repair (i.e. Error-to-cut-off). This analysis will provide answers to the sub-categories of the first three research questions in Section 5.5, 5.6, and 5.7.

The following sections examine the effects of PL and TC on the temporal phases of different types of repairs (i.e. A-repair, D-repair, E-repair sub-categories). Two-way between-groups ANOVAs are carried out to examine the effects and interaction effects of PL and TC on the temporal phases of repairs. To reduce the possibility of increasing Type 1 error, a Bonferroni correction alpha level of .025 is considered for the main effect, and .0125 for the significant interaction effects in Tables 5.30, 5.32, 5.33, 5.34, and 5.36 below.

5.8.1 Effects of PL and TC on the temporal phases of A-repair

The effects and interaction effects of PL and TC on A-repair temporal phases are presented in this sub-section. Table 5.29 below demonstrates the descriptive analyses of A-repair intervals in relation to PL and TC. The data show that the Intermediate learners unexpectedly took longer time to execute A-repair temporal phases compared to the Elementary group. It can also be observed that the dual task condition slowed down A-repair temporal phases. To see whether these differences had a statistical significance, the effects and interaction effects of PL and TC on the temporal phases of A-repair were examined (see Table 5.29 below). It can be seen that the duration of the first phase of A-repair increased significantly in the speech of the Intermediate compared to the Elementary groups.

Table 5.29. Descriptive statistics of A-repair temporal phases

Repair type	**Temporal phases**	**Proficiency level**	**Task Condition**		
			Single	**Dual**	**Total (PL)**
			Mean (SD)	**Mean (SD)**	**Mean (SD)**
A-repair	Error-to-cut-off	Elementary	.18 (.49)	.30 .53	.25 (.51)
		Intermediate	.30 (.49)	.69 (.78)	.49 (.67)
		Total (TC)	.24 (.49)	.48 (.68)	.36 (.60)
	Cut-off-to-repair	Elementary	.08 (.21)	.14 (.24)	.11 (.22)
		Intermediate	.14 (.26)	.36 (.53)	.25 (.43)
		Total (TC)	.10 (.24)	.24 (.41)	.17 (.34)
	Repair	Elementary	.27 (.68)	.41 (.63)	.35 (.66)
		Intermediate	.52 (.73)	.80 (1.19)	.66 (.98)
		Total (TC)	.39	.59	.49

Repair type	Temporal phases	Proficiency level	Task Condition		
			Single	**Dual**	**Total (PL)**
			Mean (SD)	**Mean (SD)**	**Mean (SD)**
			(.71)	(.94)	(.84)

In other words, the Intermediate group spent longer time (M = .49, SD = .67) during the Error-to-cut-off phase of A-repair (F = 6.133, p =. 015, η^2 = .046) compared to the Elementary group (M = .25, SD = .51). However, this variation is associated with a small effect size. This finding suggests that producing the first phase of A-repair (Error-to-cut-off) was considerably longer in the Intermediate learners as compared to the Elementary group. The effect size suggests that only 1.5 per cent of variance in this interval can be explained by PL. This suggests that different strategies may be employed by different proficiency groups during the processing of A-repair. The Intermediate learners may develop a certain strategy to process A-repair which may delay the processing of the first phase of A-repair compared to the Elementary learners. Table 5.30 below shows that TC significantly affected the Error-to-cut-off interval of A-repair (the first phase) (F = 6.509, p =. 012, η^2 = .048). That is to say, L2 learners spent a substantially longer time on this interval in the dual task (M = .48, SD = .68) compared to single task condition (M = .24, SD = .49). The variance, nonetheless, was associated with a small effect size which suggests that about 4.8 per cent of differences in this interval can be explained by TC. This finding indicates that the first phase of A-repair (Error-to-cut-off) slowed down in the dual task condition. This suggests that this interval of A-repair was slowed down by the increased demand of the dual task condition.

Table 5.30. Two-way between-group ANOVA: A-repair intervals

Repair types	Temporal phases of repair	Proficiency Level			Task Condition			PL*TC		
		F	**Sig.**	**η^2**	**F**	**Sig.**	**η^2**	**F**	**Sig.**	**η^2**
A-repair	Error-to-cut-off	6.133	.015*	.046	6.509	.012*	.048	1.677	.198	.013

	Cut-off-to-repair	5.936	.016*	.044	6.642	.011*	.049	1.953	.165	.015
	Repair	4.844	.03	.036	2.216	.139	.017	.256	.614	.002

*P < .025, **P < .0125, PL df (1, 128), TC df (1.28), PL*TC df (1.128)

With regard to the interaction effects of PL and TC on the Error-to-cut-off phase, the data demonstrate that there was no interaction effect of the two variables on this interval.

The second phase of A-repair is the Cut-off-to-repair interval, as displayed in Table 5.30 above. The data show that the second phase was significantly affected by PL (F = 6.509, p =. 012, η^2 = .048). The Intermediate learners spent longer time on the Cut-off-to-repair phase (M = .25, SD = .43) than the Elementary learners (M = .11, SD = .22). Yet, the variation was associated with a small effect size which suggests that PL accounts for about 4.4 percent of variance in this interval. This finding suggests that the interruption interval was slightly longer in the speech of the Intermediate learners compared to the Elementary learners. This may indicate that the Intermediate learners may think of more sophisticated adjustments to the inappropriate elements in their speech compared to the Elementary learners. As for the effect of TC on the second phase of A-repair (Cut-off-to-repair), the results show that this interval was significantly affected by TC (F = 6.642, p = .011, η^2 = .049). That is to say, the duration of the Cut-off-to-repair phase was substantially slowed down in the dual task (M = .24, SD = .41) compared to the single task condition (M = .11, SD = .24), but the effect size of this variation was also small, suggesting that around 4.9 per cent of variance in this interval can be explained by TC. The data show no interaction effect of the two variables on this interval.

As for the third phase of A-repair (Repair), the data show that the effect of PL on the Repair was approaching significance (F = 4.844, p = .03, η^2 = .036), with the Intermediate group taking longer time to execute A-repair (M = .66, SD = .98) compared to the Elementary group (M = .35, SD = .66). This finding indicates that there was a short delay in the execution of Repair in the speech of the Intermediate learners than the Elementary learners. Table 5.30 above shows no main effect of TC, or interaction effect of the two variables on the third phase of A-repair. The data in this section shows that the first two intervals of A-repair (Error-to-cut-off and Cut-off-to-repair) follow a similar pattern with respect to PL and TC. This suggests that influences on A-repair may trigger simultaneous and linear reaction from these two phases. That is to say, if Error-to-cut-off interval slows down, Cut-off-to-repair will slow down as a result. Moreover, the longer intervals of A-repair produced by the Intermediate learners suggest that different strategies may be employed by the two proficiency groups to process A-repair. It is likely that the Intermediate learners develop a more sophisticated strategy to process A-repair compared to the Elementary learners. This explanation will be further discussed in relation to previous research and the PLT in Sections 6.3 and 6.4.

5.8.2 Effect of PL and TC on the temporal phases of D-repair

The effects and interaction effects of PL and TC on D-repair temporal phases are presented in this sub-section. Table 5.31 below demonstrates the descriptive analyses of D-repair temporal phases in relation to PL and TC.

Table 5.31. Descriptive statistics of D-repair temporal phases

Repair type	Temporal phases	Proficiency level	Task Condition		
			Single	Dual	Total (PL)
			Mean (SD)	Mean (SD)	Mean (SD)
D-repair	Error-to-cut-off	Elementary	.10 (.32)	.65 (1.09)	.38 (.85)
		Intermediate	.41 (.64)	.32 (.49)	.37 (.57)
		Total (TC)	.25 (.52)	.50 (.88)	.37 (.73)
	Cut-off-to-repair	Elementary	.06 (.22)	.64 (1.27)	.36 (.96)
		Intermediate	.38 (.59)	.21 (.32)	.30 (.49)
		Total (TC)	.22 (.47)	.44 (.98)	.33 (.77)

Repair type	Temporal phases	Proficiency level	Task Condition		
			Single	Dual	Total (PL)
			Mean (SD)	Mean (SD)	Mean (SD)
	Repair	Elementary	.33 (.97)	.88 (1.19)	.61 (1.12)
		Intermediate	.77 (1.02)	.62 (.82)	.70 (.93)
		Total (TC)	.55 (1.02)	.76 (1.04)	.65 (1.03)

It has been indicated in Sections 5.5. and 5.6 that PL and TC did not have significant effects on the frequency of D-repair. The data show that the Intermediate group produced longer intervals of D-repair in the single task condition, but no consistent pattern of differences between the two proficiency groups can be found in the dual task condition in terms of D-repair intervals (Error-to-cut-off, Cut-off-to-repair, and Repair). The data also show some differences in D-repair temporal phases with regard to TC. That is, longer intervals of D-repair can be observed in the dual task compared to the single task condition. Table 5.32 below shows that PL and TC had no meaningful effects on the three temporal phases of D-repair.

Table 5.32. Two-way between-group ANOVA: D-repair intervals

Repair types	Temporal phases of repair	Proficiency Level			Task Condition			PL*TC		
		F	Sig.	η^2	F	Sig.	η^2	F	Sig.	η^2
D-repair	Error-to-cut-off	.009	.924	.000	3.418	.067	.026	6.584	.011**	.049
	Cut-off-to-repair	.170	.680	.001	2.353	.127	.018	8.247	.005**	.061

	Repair	.249	.619	.002	1.212	.273	.00 9	3.94 1	.049	.03 0

*P < .025, **P < .0125, PL df (1, 128), TC df (1.28), PL*TC df (1.128)

However, the data indicate significant interaction effects of PL and TC on the first two temporal phases of D-repair (Error-to-cut-off, Cut-off-to-repair). First, there was a significant interaction effect on the first phase of D-repair (Error-to-cut-off) ($F = 6.584$, $p = .011$, $\eta^2 = .049$). For the Elementary learners, the Error-to-cut-off interval (first phase) was longer in the dual task compared to the single task conditions. This suggests that the first phase of D-repair was slowed down when the Elementary group performed in the dual task condition. In contrast, the Intermediate learners spent less time on the first phase of D-repair (Error-to-cut-off) in the dual task compared to the single task condition as illustrated in Figure 5.17 below. The second phase of D-repair which involves a significant interaction effect is the Cut-off-to-repair interval (second phase) ($F = 8.247$, $p = .005$, $\eta^2 = .061$). It can be noted that there was a moderate effect size associated with this interaction effect which suggests that one variable may moderate the effects of the other variable.

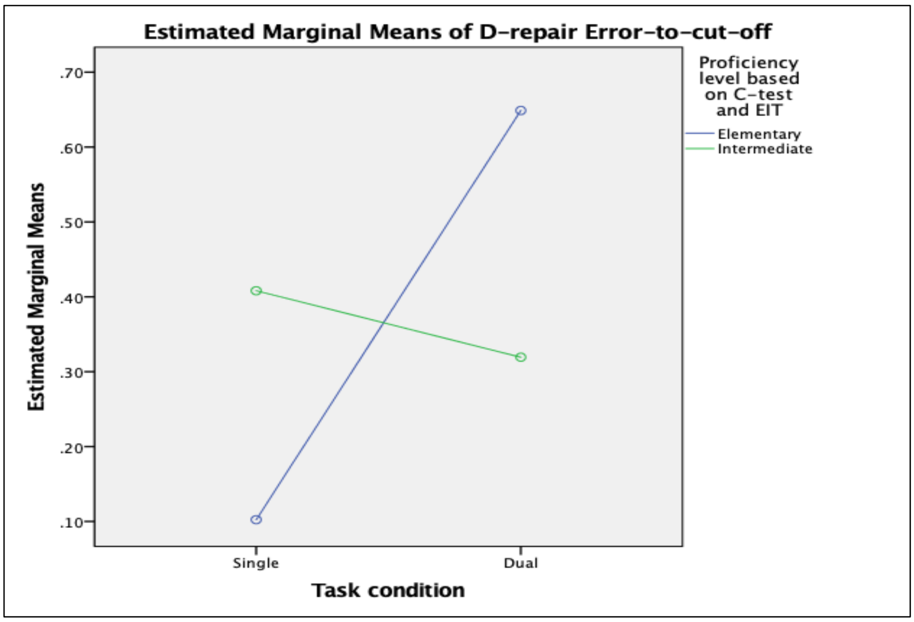


Figure 5.17. Interaction effect of PL and TC on the Error-to-cut-off

It can be observed that the Elementary learners took longer time on the second phase (Cut-off-to-repair) in the dual task rather than the single task condition. This is in contrast to the

Intermediate learners who executed the second phase of D-repair faster in the dual task compared to the single task conditions, as demonstrated in Figure 5.18 below.

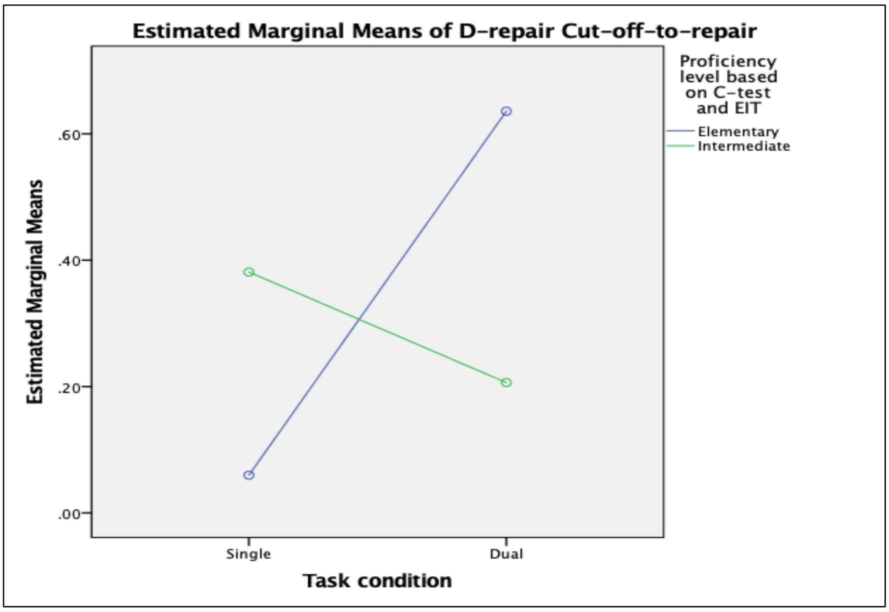


Figure 5.18. Interaction effect of PL and TC on the Cut-off-to-repair

To further explore the interaction effects of PL and TC on D-repair intervals, follow-up tests are carried out (see the Sub-section 5.7.1, for details about the follow-up tests). The data of the follow-up analyses show that in the single task condition, the first two intervals of D-repair (Error-to-cut-off and Cut-off-to-repair) were substantially slower in the speech of the Intermediate learners compared to the Elementary learners. There were no differences between the two proficiency groups in the dual task condition in terms of D-repair intervals as indicated in Table 5.33 below. This suggests that the single task condition triggered great differences between the two proficiency groups, as the Intermediate learners spent longer time on the first two intervals of D-repair (Error-to-cut-off and Cut-off-to-repair) than the Elementary learners.

Table 5.33. One-Way ANOVA for the effects of PL on D-repair intervals

PL	**Temporal phases of D-repair**	**Single task**			**Dual task**		
		F	**Sig.**	**η^2**	**F**	**Sig.**	**η^2**
Elementary	Error-to-cut-off	6.139	.016*	.088	2.351	.130	.035
Intermediate							

Elementary Intermediate	Cut-off-to-repair	8.664	.005*	.119	3.257	.076	.048

*P < .05

This suggests that the Intermediate learners may think of different strategies to replace their messages. The Elementary learners, on the other hand, may interrupt their messages immediately and choose a fast strategy to replace the utterances. It is interesting that these variances, nevertheless, disappeared when the two groups perform in the dual task condition. Concerning the differences within each proficiency group, some meaningful differences were also observed as shown in Table 5.33 above (see descriptive data in Table 5.31 above). The data show that for the Elementary group, these two temporal phases were significantly longer in the dual task condition as compared to the single task condition. As for the Intermediate group, there were no meaningful differences between these two intervals in the two task conditions. This suggests that Elementary group was significantly affected by the increased demand of the dual task condition in terms of D-repair intervals, but no effects were detected on the Intermediate group.

In short, the interaction effects suggest that the two proficiency groups probably had different reasons for slowing down the processing of D-repair in the two task conditions. First, the Intermediate learners may afford themselves high degree of freedom to think of

Table 5.34. One-Way ANOVA: The effects of TC on D-repair intervals

PL	Temporal phases of D-repair	Elementary			Intermediate		
		F	Sig.	d	F	Sig.	d
Single Dual	Error-to-cut-off	7.913	.006	.104	.375	.543	--
Single Dual	Cut-off-to-repair	6.808	.011	.091	2.041	.158	--

*P < .05

replacements of their messages during the single task condition, and that the increased demand of task condition did not affect their processing of D-repair intervals so that there were no differences between their D-repair intervals in the two task conditions. The Elementary group, on the other hand, may choose the fastest way to replace the utterances in the single task condition, but the dual task condition may be so demanding for them that they slowed down the processing of D-repair intervals in the dual task as compared to the single task conditions. The data in this section seem to confirm the explanation provided in Section 5.8.1. that the influences on a particular type of repair may trigger simultaneous and linear reactions in the two intervals of Error-to-cut-off and Cut-off-to-repair (the first and second phase of repair). Furthermore, having great differences between the two proficiency groups in the single task condition which disappear in the dual task condition is further confirmed by the data in this sub-section.

5.8.3 Effects of PL and TC on E-repair temporal phases

This sub-section presents the effects and interaction effects of PL and TC on the temporal phases of E-repair sub-categories (lexical-repair, grammatical-repair, and phonological-repair). First, Table 5.35 below presents descriptive analyses of lexical-repair, grammatical-repair, and phonological-repair intervals. This is followed by presenting the results of the ANOVAs in Table 5.35. It has been indicated in Sections 5.5 and 5.6 that the sub-categories of E-repair was not affected by proficiency level and task condition. As for the temporal phases of E-repair sub-categories, it was predicted that the Intermediate learners would perform better than the Elementary learners under the dual task condition. This means that temporal phases of E-repair would be shorter in the Intermediate group compared to the Elementary group (see Section 4.3). The data in Table 5.35 below indicate that the Intermediate learners spent shorter durations when they executed lexical-repair and grammatical-repairs intervals compared to the Elementary learners. As for the intervals of the phonological-repair, in comparison to the temporal phases of other repairs, the data indicate that the shortest intervals were executed during the processing of phonological-repair. It also shows that the Intermediate learners spent slightly longer time on phonological-repair intervals compared to the Elementary learners. With regard to TC, the data demonstrate no consistent patterns for the effects of TC on lexical-repair and phonological-repair temporal phases. It is, nonetheless, indicated that shorter intervals of grammatical-repair were executed in the dual task condition compared to the single task condition. This indicates that the temporal phases of E-repair sub-categories did not show similar pattern in relation to proficiency level and task condition. The results presented in Table 5.36 below indicate that there were no main effects or interaction effects of PL, TC on the temporal phases of lexical-repair, grammatical-repair, and phonological-repair. The data suggest that the development of proficiency did not affect the time needed to execute E-repair sub-categories. It also suggests that the increased demand of task condition did not have a significant influence on the duration of E-repair sub-categories.

Table 5.35. Descriptive statistics of E-repair temporal phases

Repair type	Measures	Proficiency level	Task Condition		
			Single	Dual	Total
			Mean (SD)	Mean (SD)	Mean (SD)
Lexical-repair	Error-to-cut-off	Elementary	.34 (.40)	.45 (.69)	.40 (.57)
		Intermediate	.30 (.48)	.35 (.62)	.32 (.55)
		Total	.32 (.44)	.40 (.66)	.36 (.56)
	Cut-off-to-repair	Elementary	.32 (.49)	.28 (.54)	.30 (.51)
		Intermediate	.18 (.30)	.12 (.21)	.15 (.26)
		Total	.25 (.41)	.21 (.43)	.23 (.41)
	Repair	Elementary	.47 (.61)	.43 (.58)	.45 (.59)
		Intermediate	.38 (.59)	.28 (.49)	.33 (.54)
		Total	.43 (.60)	.40 (.54)	.39 (.57)
		Elementary	.81	.49	.65

Repair type	Measures	Proficiency level	Task Condition		
			Single	Dual	Total
			Mean (SD)	Mean (SD)	Mean (SD)
Grammatical-repair	Error-to-cut-off		(1.43)	(.61)	(1.09)
		Intermediate	.36 (.40)	.56 (1.07)	.45 (.80)
		Total	.59 (1.08)	.52 (.84)	.56 (.97)
	Cut-off-to-repair	Elementary	.60 (.96)	.50 (.77)	.55 (.86)
		Intermediate	.36 (.61)	.29 (.63)	.33 (.62)
		Total	.48 (.81)	.41 (.71)	.44 (.76)
	Repair	Elementary	1.07 (1.65)	.52 (.55)	.78 (1.24)
		Intermediate	.60 (.84)	.47 (.57)	.54 (.72)
		Total	.84 (1.33)	.49 (.56)	.67 (1.03)
Phonological-repair	Error-to-cut-off	Elementary	.12 (.26)	.12 (.25)	.12 (.25)
		Intermediate	.17 (.35)	.16 (.37)	.17 (.36)
		Total	.15 (.30)	.14 (.31)	.14 (.31)
	Cut-off-to-repair	Elementary	.06 (.14	.14 (.37)	.10 (.28)

Repair type	Measures	Proficiency level	Task Condition		
			Single	Dual	Total
			Mean (SD)	Mean (SD)	Mean (SD)
		Intermediate	.07 (.14)	.16 (.52)	.11 (.37)
		Total	.06 (.14)	.15 (.44)	.11 .33
	Repair	Elementary	.18 (.41)	.15 (.30)	.16 (.35)
		Intermediate	.22 (.40)	.13 (.24)	.18 (.33)
		Total	.20 (.40)	.14 (.27)	.17 (.34)

This means that this aspect of monitoring might not be sensitive to influences of proficiency and task condition. In short, the data in this section suggest that there are two patterns of repair intervals in the current data. The first pattern was affected by PL and TC, and this pattern includes A-repair and D-repair.

Table 5.36. The effects of PL and TC on the temporal phases of repair

Repai r types	Temporal phases of repair	Proficiency Level			Task Condition			PL*TC		
		F	Sig.	η^2	F	Sig.	η^2	F	Sig.	η^2
Lexical-repair	Error-to-cut-off	.564	.454	.004	.606	.438	.005	.088	.768	.00 1
	Cut-off-to-repair	4.228	.042	.032	.473	.493	.004	.027	.869	.00 0
	Repair	1.408	.238	.011	.581	.447	.005	.082	.774	.00 1
Gram matic	Error-to-cut-off	1.374	.243	.011	.126	.724	.001	2.38 9	.125	.01 8

Repai r types	Temporal phases of repair	Proficiency Level			Task Condition			PL*TC		
		F	Sig.	η^2	F	Sig.	η^2	F	Sig.	η^2
	Cut-off-to-repair	2.799	.097	.021	.388	.534	.003	.008	.929	.00 0
	Repair	2.060	.154	.016	3.736	.055	.028	1.36 6	.245	.01 1
Phonological-repair	Error-to-cut-off	.800	.373	.006	.025	.875	.000	.015	.904	.00 0
	Cut-off-to-repair	.051	.822	.000	2.296	.132	.018	.007	.932	.00 0
	Repair	.053	.819	.000	1.070	.303	.008	.273	.602	.00 2

*P < .025, **P < .0125, PL df (1, 128), TC df (1.28), PL*TC df (1.128)

The other pattern, which includes E-repair sub-categories, was not affected by influences of PL, TC and their interaction effects. This suggests that the different types of repair may entail different processing which will be further discussed in Sections 6.2-6.4.

5.9 Research Question 4: Working memory capacity and L2 self-monitoring

The fourth research question asked whether there is a relationship between working memory capacity and L2 self-monitoring in terms of disfluency, frequency of repair types, temporal phases of repair, and accuracy in the two task conditions (single and dual). Working memory capacity was measured using Backward Digits Span Test (see Section 4.6.4 for details). It is predicted that the more demanding task condition has a negative impact on working memory functioning (Michel, Kormos, Brunfaut, & Ratajczak, 2019). The dual task condition is anticipated to have a negative influence on the functioning of working memory because it is expected to tax available cognitive resources. As such, it is hypothesised that a larger WMC would enhance L2 monitoring in the single task rather than in the dual task conditions. To test this hypothesis, the SPSS data file was split into two parts to run correlation analyses in the single and dual task conditions separately.

A Pearson product-moment correlation coefficient is used to determine if there is a relationship between aspects of L2 self-monitoring and WMC in the two task conditions.

Assumptions of normality and linearity were checked to ensure the compatibility of Pearson product-moment correlation analysis with the current data. If significant correlation is obtained, Cohen's (1988) guidelines, that is, r =.10-29 as small, r=.30-.49 as medium, and r = .50-1.0 as large, will be applied to assess the relationship between WMC and L2 self-monitoring aspects. This is followed by calculating the coefficient determination (by squaring r value) to understand how much variance is shared by the two variables. Squared R value indicates how close the data are to the fitted regression line. The following sub-sections will present the correlations between WMC and L2 self-monitoring in terms of disfluency, frequency of repair types, temporal phases of repair, and accuracy in the two task conditions.

5.9.1 The relationship between WMC and disfluency

This sub-section examines whether there is a relationship between WMC and L2 disfluency measures (filled pauses, mid-clause silent pauses, end-clause silent pauses, and repetition) in the two task conditions. It was expected that a larger WMC would enhance self-monitoring in the single task condition. Table 5.37 below presents the results of Pearson product-moment correlation coefficient (r) for WMC and disfluency in the two task conditions.

Table 5.37. Pearson Product-moment Correlations: WMC and disfluency

Single task condition				
WMC	**Filled pauses**	**Mid-clause pauses**	**End-clause pauses**	**Repetition**
Pearson r Sig (2-tailed) N	-.184 .140 66	.202 .104 66	.119 .340 66	-.246* .046 66
Dual task condition				
WMC	**Filled pauses**	**Mid-clause pauses**	**End-clause pauses**	**Repetition**
Pearson r Sig (2-tailed) N	-.192 .122 66	.046 .712 66	.139 .266 66	-.190 .126 66

*. Correlation is significant at the 0.05 level (2-tailed)

The data presented in Table 5.37 above indicate that only one aspect of disfluency correlated negatively with WMC in the single task condition. The correlation was weak, $r = -.246$, $n = 66$, $p = .05$. This indicates that learners with a larger working memory capacity produced a smaller number of repetitions in the single task condition. By calculating the coefficient of determination, the result shows that WMC helps to explain about 6 per cent of the variance in repetition. This is quite a small percentage, which suggests a weak negative relationship between WMC and repetition in the single task condition. On the other hand, there is no significant correlation between WMC and disfluency in the dual task condition, even though the direction of correlation is the same as in the single task condition. That is, there is a negative relationship between WMC, filled pauses and repetition, and a positive relationship between WMC and silent pauses in the dual task condition; however, they are not significant. This suggests that available working memory resources may help L2 learners monitor their speech prior to articulation in the single task condition, as this resulted in fewer repetitions compared to the dual task condition. On the other hand, with the increased cognitive demand in the dual task condition, working memory resources might be consumed, resulting in a negative impact on pre-articulatory monitoring. This finding is in line with the hypothesis which predicted that larger WMC would enhance self-monitoring in the single task condition rather than the dual task condition. Thus, the hypothesis is only partially fulfilled.

5.9.2 The relationship between WMC and repairs

This sub-section examines whether there is relationship between WMC and repair types (A-repair, D-repair, E-repair, lexical-repair, grammatical-repair, and phonological-repair) in the two task conditions. The hypothesis stated that larger WMC would enhance self-monitoring in the single task condition (see Section 4.3). Table 5.38 below presents the results of Pearson product-moment correlation coefficient (r) for the relationship between WMC and repair types in the two task conditions.

Table 5.38. Pearson Product-moment Correlations: WMC and repair types

Single task condition						
WMC	**A-repair**	**D-repair**	**E-repair**	**Lexical-repair**	**Grammatical-repair**	**Phonological-repair**

Pearson r Sig (2-tailed) N	.095 .449 66	.079 .528 66	-.182 .144 66	-.275* .026 66	.150 .228 66	.101 .418 66
Dual task condition						
WMC	**A-repair**	**D-repair**	**E-repair**	**Lexical-repair**	**Grammatical-repair**	**Phonological-repair**
Pearson r Sig (2-tailed) N	-.106 .399 66	.039 .755 66	-.039 .773 66	-.086 .494 66	.051 .682 66	.144 .250 66

*. Correlation is significant at the 0.05 level (2-tailed)

The data demonstrate that in the single task condition there was no relationship between WMC and A-repair, D-repair, grammatical-repair, and phonological-repair. The only exception was lexical-repair as a small negative correlation was found between WMC and lexical-repair, $r = -.275$, $n = 66$, $p = .03$. This indicates that learners with a larger working memory capacity produced fewer lexical-repair in the single task condition. By calculating the coefficient of determination, the result shows that WMC helps to explain about 8 per cent of the variance in lexical-repair frequency in the single task condition. This is quite a small percentage which suggests a weak relationship between WMC and lexical-repair frequency. However, the data indicate no correlation between WMC and repair types in the dual task condition, which suggests that there was not significant relationship between WMC and repair types in the dual task condition. This finding suggests that a larger WMC may enhance L2 self-monitoring prior to articulation in the single task condition compared to the dual task condition. That is, L2 learners with a larger WMC may detect and correct some of their lexical errors prior to articulation in the single task condition, and this resulted in fewer lexical-repair. The data in this section is in line with the hypothesis stated above.

5.9.3 The relationship between WM and the temporal phases of repairs

The correlations between WMC and the temporal phases of repair are presented in this sub-section. It was hypothesised that L2 learners with a larger working memory capacity would benefit from extra working memory resources when they perform in the single task condition rather than the dual task condition. As there is a large number of repair temporal phases in this study, the data are presented in two tables. First, the correlations between WMC and the temporal phases of A-repair and D-repair in the two task conditions are presented in Table 5.39 below. Then, the correlations between WMC and the intervals of E-repair sub-categories in the two task conditions are illustrated in Table 5.40.

Table 5.39. Pearson Correlations: WMC, A-repair and D-repair intervals

Single task condition						
WMC	**Temporal phases of A-repair**			**Temporal phases of D-repair**		
	Error-to cut-off	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**
Pearson r Sig (2-tailed) N	-.142 .257 66	.096 .442 66	.045 .720 66	-.013 .917 66	.063 .613 66	.015 .903 66
Dual task condition						
WMC	**Temporal phases of A-repair**			**Temporal phases of D-repair**		
	Error-to cut-off	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**
Pearson r Sig (2-tailed) N	.090 .473 66	.046 .713 66	-.013 .918 66	-.173 .164 66	.119 .341 66	.142 .254 66

*. Correlation is significant at the 0.05 level (2-tailed)

Table 5.39 above indicates that there was no significant correlation between WMC and the temporal phases of A-repair and D-repair in the two task conditions. This suggests that individual differences in WMC do not help to explain variations in A-repair and D-repair intervals in the two task conditions.

Table 5.40. Pearson Product-moment Correlations: WMC and E-repair intervals

Single task condition

WMC	Lexical-repair			Grammatical-repair			Phonological-repair		
	Error-to cut-off	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**
Pearson r Sig (2-tailed) N	-.090 .474 66	-.079 .526 66	-.184 .140 66	-.115 .359 66	-.129 .302 66	-.085 .497 66	.149 .233 66	.152 .222 66	.173 .165 66
Dual task condition									
WMC	**Lexical-repair**			**Grammatical-repair**			**Phonological-repair**		
	Error-to cut-off	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**	**Error-to cut-off**	**Cut-off-to-repair**	**Repair**
Pearson r Sig (2-tailed) N	-.089 .480 66	-.084 .503 66	-.185 .137 66	.046 .712 66	.036 .774 66	-.048 .701 66	.057 .647 66	.060 .633 66	-.034 .784 66

*. Correlation is significant at the 0.05 level (2-tailed)

Table 5.40 above shows negative correlations between WMC and the temporal phases of lexical-repair and grammatical-repair, and positive correlations between WMC and phonological-repair in the single task condition; however, these correlations did not reach statistical significance. In the dual task condition, the relationship between lexical-repair phases and WMC was in the same direction as in the single task condition. However, there was no significant correlation between WMC and any of temporal phases of E-repair sub-categories in the dual task condition. This suggests that WMC was not implicated in the two task conditions in terms of temporal phases of E-repair. Thus, the hypothesis was not supported in terms of the relationship between WMC and E-repair temporal phases.

5.9.4 The relationship between WMC and accuracy

This sub-section aims to explore whether there is a relationship between WMC and accuracy measures (percentage of error-free clauses and ratio of error-correction) in the two task conditions. It was predicted that extra working memory resources would assist performance in the single task condition where there is less demand on cognitive resources (see Section 4.3).

Table 5.41. Pearson Product-moment Correlations: WMC and accuracy

Single task condition		
WMC	**Percentage of error-free clauses**	**Ratio of error-correction**
Pearson r Sig (2-tailed) N	.019 .882 66	.207 .095 66
Dual task condition		
WMC	**Percentage of error-free clauses**	**Ratio of error-correction**
Pearson r Sig (2-tailed) N	-.145 .245 66	-.126 .312 66

*. Correlation is significant at the 0.05 level (2-tailed)

The data presented in Table 5.41 above demonstrate that the relationship between WMC and accuracy measures in the two task conditions follow different directions. That is, in the single task condition there was a positive non-significant relationship between WMC and accuracy measures. In the dual task condition, there was a negative non-significant relationship between accuracy measures and WMC. However, there was no significant correlation between WMC and accuracy measures in the two task conditions. This suggests that individual differences in working memory capacity do not account for variances in the accuracy of L2 speech production in the single and dual task conditions. Thus, the hypothesis is rejected.

To conclude this section, the findings were in line with the hypothesis stated in Section 4.3. There were weak significant negative correlations between WMC and two aspects of self-monitoring (repetition and lexical-repair). The data suggests that individual differences in working memory capacity could predict only small variations in one aspect of disfluency and one repair. However, most of the correlations in this section are not significant which is not surprising as this is in line with previous research that examined WMC in relation to L2 procession and production (e.g., Georgiadou & Roehr-Brackin, 2017; Awwad, 2017). It can be argued that research in this area needs to employ various measurements of WMC which can properly assess this construct. Update in the field of psychology and neurology in this regard can be useful in developing an appropriate measurement for working memory.

5.10 Summary of the key findings

This section presents a summary of the findings in this study and relate the data to the four research questions.

5.10.1 Research Question 1: Effects of Proficiency

Research Question 1(RQ1) investigated the effects of proficiency level on L2 self-monitoring behaviour in terms of disfluency, frequency of repair types, temporal phases of repairs, and accuracy. Findings indicated that the Intermediate learners produced fewer filled pause and more clause-final silent pauses compared to the Elementary learners. They also produced more A-repair and longer duration of A-repair than the Elementary group. Accuracy rate increased in the speech of Intermediate but not the Elementary learners. This suggests that with the development of proficiency the foci and processes of L2 self-monitoring develop. That is, there was a shift in disfluency from a pattern of more filled pauses and less clause-final pauses in the Elementary group, to a pattern of fewer filled pauses and more clause-final silent pauses in the Intermediate learners. There was also increased tendency for making A-repair in the Intermediate group which suggests a possible change in monitoring foci. A higher percentage of error-free clauses indicates a possible development in monitoring functioning which will be further discussed in Section 6.2.

5.10.2 Research Question 2: Effects of Task Condition

Research Question 2 (RQ2) examined the effects of task condition on L2 self-monitoring in terms of disfluency, repair types, temporal phases of repair, and accuracy. It was expected that the dual task condition would have detrimental effects on L2 self-monitoring aspects. In contradiction to the hypothesis, the data show that task condition had no significant effects on disfluency, repair types, or accuracy. The only exception was that task condition slowed down the first two intervals of A-repair. This in turn may indicate that task condition did not affect monitoring aspects; it only slowed down processing of A-repair. This suggests that A-repair involves different processing from other types of repair which makes it vulnerable to the increased demand of task condition. The data may also signify the extent to which cognitive resources are implicated in L2 monitoring, as discussed in Section 6.3.

5.10.3 Research Question 3: Interaction Effects of PL and TC

Research Question 3 (RQ3) investigated whether there were interaction effects of PL and TC on L2 self-monitoring behaviour in terms of disfluency, repair types, temporal phases of repair, and accuracy. The data demonstrated that there was an interaction effect on filled pauses and the first two intervals of D-repair. It is interesting to find that the single task condition triggered some differences between the two proficiency groups in filled pauses and D-repair intervals. These differences disappeared when they performed in the dual task condition. This suggests that a normal speaking task condition may elicit differences between elementary and intermediate levels of proficiency. However, performing under the more demanding task condition may promote the two proficiency groups to employ different strategies to cope with the increased demand, and this probably reduced the differences between their performances. It is also worth noting that PL and TC triggered simultaneous and linear reactions from the first two intervals of A-repair and D-repair. This will be further discussed in Section 6.4.

5.10.4 Research Question 4: Relationship between WM and L2 Self-Monitoring

Research Question 4 (RQ4) investigated the relationship between WMC and L2 self-monitoring aspects in the two task conditions. It was expected that a larger working memory capacity would enhance self-monitoring behaviour in the single task condition where there is less demand on cognitive resources. The results presented in Section 5.9 were in line with the hypothesis. The data showed that significant negative weak correlations were found between WMC and repetition, as well as WMC and lexical-repair in the single task condition but not under the dual task condition. However, these correlations were quite small, suggesting that individual differences in WMC could explain small variations in repetition and lexical-repair in the single task condition. This may mean that individual differences in working memory capacity might not be the key cognitive factor that accounts for differences in self-monitoring behaviour in the two task conditions. There is further discussion on this in Section 6.5.

Chapter 6: DISCUSSION

6.1 Introduction

The main purpose of the present study was to explore the effects of proficiency level (PL) and task condition (TC) on L2 self-monitoring behaviour in terms of disfluency, repair types, temporal phases of repair, and accuracy. It also aimed to examine whether there was a relationship between individual differences in working memory capacity (WMC) and L2 self-monitoring aspects. Chapter 5 presented the results of this data analysis in relation to the four research questions of the study. In this chapter, the results of the data analysis are interpreted in relation to previous studies, the Perceptual Loop Theory (the PLT), and the L2 speech production model. Visual presentations of the results are provided in this chapter to communicate the main findings. Bar graphs are used to illustrate differences between performances of the two proficiency groups and task conditions, and line graphs to display interaction effects. Section 6.2 presents a discussion about L2 self-monitoring aspects in relation to proficiency level. Section 6.3 examines the data of L2 self-monitoring behaviour and task condition. This is followed by looking at the interaction effects between proficiency and task condition in Section 6.4. Finally, the relationship between working memory capacity and self-monitoring is discussed in Section 6.5.

6.2 Proficiency level and L2 self-monitoring behaviour

The first research question sought to examine the effects of PL on L2 self-monitoring behaviour. Monitoring behaviour was examined in two proficiency groups, namely, the Elementary and Intermediate groups. Four hypotheses were formulated for the first research question in relation to the four aspects of L2 self-monitoring (disfluency, repair types, temporal phases of repairs, and accuracy). The results of the two-way between-groups ANOVAs in Section 5.4 are discussed in the following sub-sections.

6.2.1 Proficiency and disfluency

The first hypothesis stated that there would be more filled pauses, mid-clause silent pauses, and repetition in the Elementary learners' speech compared to the speech of the Intermediate learners. On the other hand, it was expected that silent pauses at clause boundaries would increase in the Intermediate group, compared to the Elementary group. The data indicated that filled pauses significantly decreased in the speech of the Intermediate learners compared to the Elementary learners. On the other hand, end-clause silent pauses significantly increased in the speech of the Intermediate learners but not the Elementary learners. There were no significant differences between the two groups in terms of mid-clause silent pauses and repetition. Thus, the hypothesis was partially supported. That is, the findings showed two patterns of disfluency in the current data: a decrease in filled pauses and an increase in clause-final silent pauses in the Intermediate learners rather than in the Elementary learners, as illustrated in Figure 6.19 below.

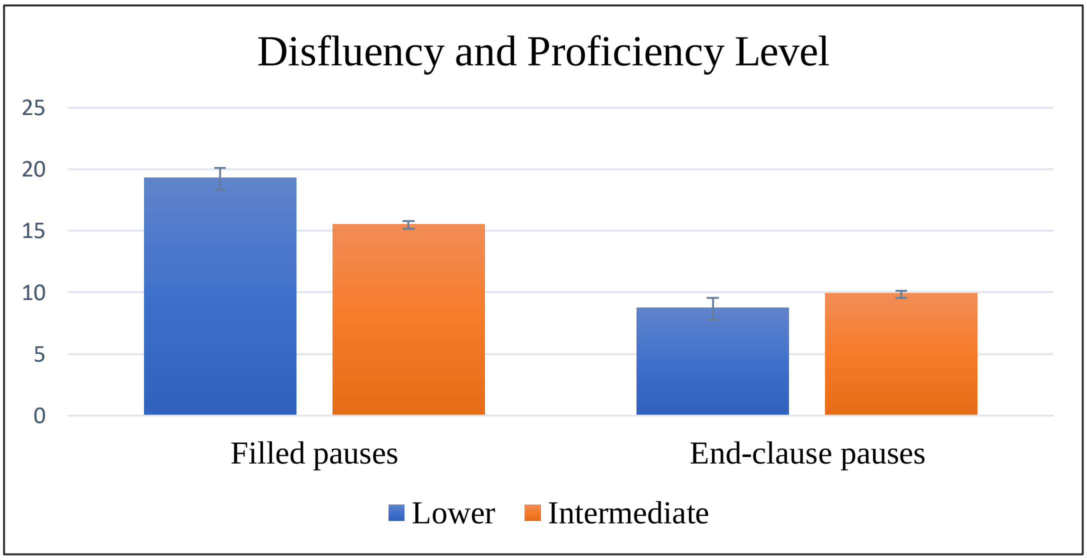


Figure 6.19. The effects of proficiency level on disfluency

L2 research has provided different accounts as to why filled pauses are produced in L2 speech. They have been considered as a strategy for coping with processing pressures during spontaneous speech production (De Jong, Steinel, Florijn, Shoonen, & Hulstijn, 2012; Dornyei & Kormos, 1998; Skehan & Foster, 2005). Filled pauses are supposed to have the same function as silent pauses, and are likely to be used to promote fluency (Kahng, 2014). It has also been argued that filled pauses are planned and produced in order to perform a communicative role in the same way as lexical elements (Clark & Fox Tree, 2002). Other researchers argue that filled pause behaviour might be linked to a personal speaking style, rather than a second language processing behaviour (De Jong, Groenhout, Shoonen, & Hulstijn, 2015; Derwing, Munro, Thomson, & Rossiter, 2009). Among these views, the first one is the most common and is supported by research in the field of L2. However, in the current study, filled pause behaviour will be interpreted in light of the Perceptual Loop Theory (Levelt, 1983, 1989), which is not commonly discussed in L2 research. According to the PLT, disfluency (e.g., filled pauses, silent pauses, and repetition) are produced as corrective actions to anticipated errors (Levelt, 1983). That is, when the monitor detects an error in the inner speech (before articulation), it sends instructions to speech processes and these errors or inappropriateness are corrected before articulation, resulting in disfluency. This is called Covert-repair which entails correcting errors or appropriateness before articulation. Although this type of repair is not examined in the current study, it is important for this investigation to understand the impact of proficiency on disfluency as an aspect of self-monitoring.

The hypothesis for the effects of proficiency on disfluency stated that if disfluency (filled pauses, mid-clause silent pauses, and repetition) is corrective action to anticipated errors, disfluency features are expected to decline in the Intermediate group. That is to say, it is expected that accuracy would improve in the Intermediate learners and disfluency would decline as a result. This assumption was partly supported by the significant increase of percentage of error-free-clauses (see Section 5.5.3), and the significant decrease of filled pauses in the Intermediate learners. Such an assumption, nevertheless, cannot be entirely verified as other disfluency features did not show significant decrease in the Intermediate learners. Furthermore, it is not possible to confirm whether the increase of filled pauses in the speech of the Elementary learners relates to monitoring, or other processing difficulties (e.g., the Formulator processes). That is, filled pauses frequency in L2 speech might not particularly reflect monitoring, as it may also be produced to serve different purposes as discussed above. Thus, the assumption that disfluency is corrective action to anticipated errors might not be particularly precise in the case of L2 speech. However, to investigate the extent to which filled pauses relate to monitoring, L2 researchers need to examine Covert-repair in different levels of proficiency by means of quantitative and qualitative methods.

Concerning clause-final pauses, the result was in line with previous literature (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016; Tavakoli, 2011). That is, more silent pauses at clause boundaries were produced by the Intermediate learners than the Elementary learners. It has been argued that clause-final pauses relate to the Conceptualiser's influences (Lambert, Kormos, & Minn, 2017; Skehan, Foster, & Shum, 2016). In other words, L2 learners seem to pause at clause boundaries to plan the next utterance. According to Levelt (1983), there are three loops of monitoring that inspect the outcomes of speech processes (see Section 2.3.3). The conceptual loop inspects the outcome of the Conceptualiser (preverbal plan); the inner loop inspects the outcome of the Formulator (the phonetic plan); and the auditory loop checks the outcome of the articulator (overt speech) (Levelt,1983). This shows that as the speech plan for the next utterance is being prepared by the Conceptualiser, it is necessarily checked by the monitor before passing it to the next speech component according to the PLT. Thus, end-clause silent pauses may show the influences of both the Conceptualiser and the Monitor.

One might ask why the frequency of end-clause pauses increased in the Intermediate learners rather than in the Elementary learners. Research shows that pauses at clause boundaries are used by L1 speakers to perform more general long-term planning of the next utterance (Kircher, Brammar, Levelt, Bartles, & McGuire, 2004). This general long-term planning is not only restricted to the conceptual preparation of the message; it also involves other aspects of the message such as word-order and syntactic encoding (Kircher et al., 2004). Thus, self-monitoring occurs during clause-final pauses (Kircher et al., 2004; Levelt, 1983). This suggests that pausing at clause boundaries in L1 implies that the monitor checks both speech conceptual plan and other aspects of the message before articulation. This could be considered as macro-level monitoring which is carried out before the articulation of L1 utterance. On the other hand, pausing within clause boundaries in L1 shows that the monitor checks the Formulator outcome (Kircher et al., 2004). This indicates that the monitor checks the Formulator output to determine whether the outcome of the Formulator is compatible with the standards of production. As most L1 speech processes are automatic, within-clause boundaries pauses are not frequently produced compared to L2 learners (Tavakoli, 2011). However, in L2 speech where most of speech processes are still not automatized, silent pauses may indicate monitoring (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Michel, 2011; Tavakoli, 2016), or they could relate to processing difficulties such as processing lexis, grammar, or phonology. That is to say, silent pauses in L2 are likely to relate to monitoring or to other speech stages, particularly the Formulator. The current data displayed no differences between the two proficiency groups concerning the mid-clause silent pauses, and this suggests that the Formulator processes may be similar in the two proficiency groups. This is not surprising because the present study compares elementary to intermediate levels of proficiency, and these groups may not be very different in their L2 processing abilities. However, the significant increase of end-clause pauses in the Intermediate learners suggests that these learners may develop a certain monitoring behaviour which entails long-term planning and monitoring of utterances, which is similar to monitoring behaviour of L1 speakers as discussed above.

To conclude, the data showed that two patterns of disfluency emerged in relation to proficiency level: the Intermediate learners made fewer filled pauses and more clause-final pauses than the Elementary learners. The results were interpreted in light of the PLT which views filled pauses as corrective measures to expected errors (Levelt, 1983). It has been assumed that the number of filled pauses possibly decreased in the Intermediate learners as a result of the development of the linguistic system, which resulted in a higher accuracy rate. However, this assumption cannot be entirely verified as Covert-repair was not part of the current investigation. Finally, the increase of clause-final pauses was explained as a potential shift towards macro-level planning and monitoring of next utterances, which is common in L1 speakers.

6.2.2 Proficiency and repair frequency

The second sub-hypothesis of the first research question predicted that proficiency level would affect the frequency of repair types (A-repair, D-repair, lexical-repair, grammatical-repair, and phonological-repair). That is, it was expected that L2 Intermediate learners would produce more A-repair, and fewer number of E-repair sub-categories (lexical-repair, grammatical-repair, and phonological-repairs) compared to the Elementary learners. The data analysis showed that the Intermediate learners produced indeed significantly more A-repair compared to the Elementary learners. Concerning D-repair, the differences between the two proficiency groups were approaching significance with the Intermediate learners producing more D-repair compared to the Elementary learners.

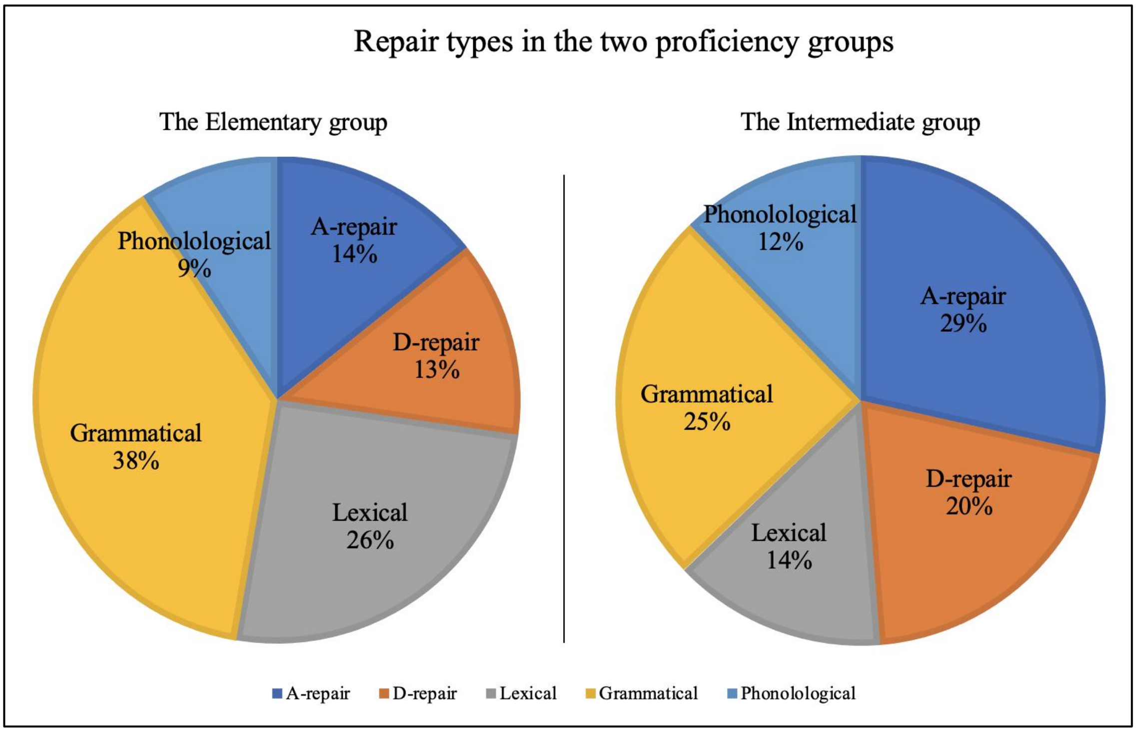


Figure 6.20. Repair types in the two proficiency groups

The data analysis also indicated that E-repair was approaching significance with the Intermediate learners producing fewer E-repair sub-categories (lexical and grammatical-repair) compared to the Elementary learners. The data was in line with the hypothesis although only A-repair showed a significant increase in the Intermediate group compared to the Elementary group. It can be seen in Figure 6.20 above that in the case of the Elementary group, the overall E-repair (lexical-repair, grammatical-repair, and phonological-repair) made about 73% of all repairs, whereas A-repair and D-repair together made up about 27% of repairs. In the speech of the Intermediate group, the percentage of E-repair decreased to about 51% of all repair. On the other hand, A-repair and D-repair together reached about 49% of repairs in the speech of the Intermediate group, as demonstrated in Figure 6.20 above.

The data indicates that two patterns of repair types emerged in relation to proficiency level: there was a significant increase of A-repair, and a slight decrease in E-repair in the speech of the Intermediate learners rather than the Elementary learners. This finding is in line with the previous studies which examined the effects of proficiency on repair types (e.g., Kormos, 2000; Van Hest, 1996; Zuniga & Simard, 2019). In Kormos's (2000) study, advanced learners made more A-repair, and fewer lexical-repair and grammatical-repair than pre-intermediate learners. However, there were no significant differences between pre-intermediate and upper-intermediate learners on the frequency of A-repair and E-repair sub-categories. Kormos (2000) found that having higher proficiency scores correlated positively with A-repair, and negatively with lexical-repair, grammatical-repair types, and the total number of E-repair. Similarly, E-repair correlated negatively with proficiency level in Zuniga and Simard's (2019) study. That is to say, the higher proficiency score L2 learners obtained, the less E-repair they produced. Van Hest (1996) also found that advanced learners produced significantly more A-repair and fewer lexical-repair types than beginners and intermediate learners, but no differences were reported between beginners and intermediate learners. It should be noted that there were differences between beginners and advanced learners, but no differences were reported between beginners and intermediate learners in Kormos's (2000) and Van Hest's (1996) studies. In comparison to previous studies, it can be assumed that the Intermediate group in the present study is possibly comparable to Kormos's and Van Hest's advanced groups rather than their intermediate groups. This may explain the significant increase of A-repair in the speech of the Intermediate group in the current study. In comparison to the findings of the previous studies, the current study showed two different patterns of repairs in the L2 speech: a significant increase in A-repair and a slight decrease in

E-repair in the Intermediate group. It is worth noting that D-repair involved the same pattern as A-repair in the current data, but it was mediated by task condition, as will be discussed in Section 6.4.

One possible interpretation of the substantial increase of A-repair in the speech of the Intermediate learners is found in previous literature. It has been assumed that with the automatization of the Formulator sub-processes, E-repair decreases, thus freeing attentional resources, so that these resources can attend to the Conceptualiser and the monitoring processes (Gilabert, 2007; Kormos, 2000, 2006; Van Hest, 1996). This means that the availability of attentional resources, as a result of the automatization of the Formulator sub-processes, accounts for the increase of A-repair in advanced learners (Gilabert, 2007; Kormos, 2000, 2006; Van Hest, 1996). This suggests that available cognitive resources are the key factor in the increased number of A-repair. However, the present study provides empirical evidence which shows that consuming available cognitive resources did not affect the frequency of A-repair as presented in Section 5.6.2. As such, the increase of A-repair might be triggered by other factors. It might be assumed that the monitoring functioning might be affected by the development of the declarative knowledge of L2 rules. The data showed that in the case of the Elementary learners about three quarters of repairs related to correcting linguistic errors (lexical, grammatical, and phonological) as shown in Figure 6.20 above. It was assumed that L2 lower proficient learners focus on linguistic errors due to limited knowledge of the target language (O’Connor, 1988). This means that due to their limited knowledge of English, the Elementary learners may check their speech against a limited number of rules during the monitoring process. The data also showed that there was about 27% A-repair and D-repair types in the speech of the Elementary learners which suggests that they made few A-repair and D-repair types. It could be assumed that the Elementary learners may produce A-repair and D-repair when they realized that certain inappropriate utterances may lead to inaccurate information during task performance. On the other hand, with the development of knowledge of the target language in the Intermediate learners, the monitor checked utterances against a more developed L2 linguistic repertoire, resulting in a significant increase of A-repair in the speech of the Intermediate learners as compared to the Elementary learners. In short, it seems that in the Elementary group, the monitor focused on correcting linguistic errors, whereas the development of L2 knowledge allowed L2 intermediate learners to check and revise their linguistic errors as well as appropriateness of their utterances. This suggests that with the development of proficiency, L2 learners' monitoring behaviour is likely to develop so that they can monitor their speech for both form and meaning.

To conclude, it was predicted that the Intermediate learners would produce more A-repair, and fewer E-repair compared to the Elementary learners. The findings showed that the Intermediate learners produced significantly more A-repair compared to the Elementary learners. The development of the declarative knowledge in the Intermediate group needs be considered when interpreting the substantial increase of A-repair in the Intermediate group. Further discussion about the development of declarative knowledge can be found in the next Section (6.2.3)

6.2.3 Proficiency and temporal phases of repairs

It was predicted that repair intervals would decrease in the Intermediate group compared to the Elementary group. The data in the sub-section 5.8 showed that two patterns of repair intervals emerged in relation to proficiency level. The first pattern showed that two temporal phases of A-repair (and D-repair in the single task condition) substantially slowed down in the Intermediate rather than Elementary groups. The second patterns showed that temporal phases of E-repair sub-categories were not affected by proficiency level, as illustrated in Figure 6.21 below. It should be noted that D-repair intervals were mediated by task condition so that they appear almost identical in the two proficiency groups in the chart below (for further discussion, see Section 5.8.2).

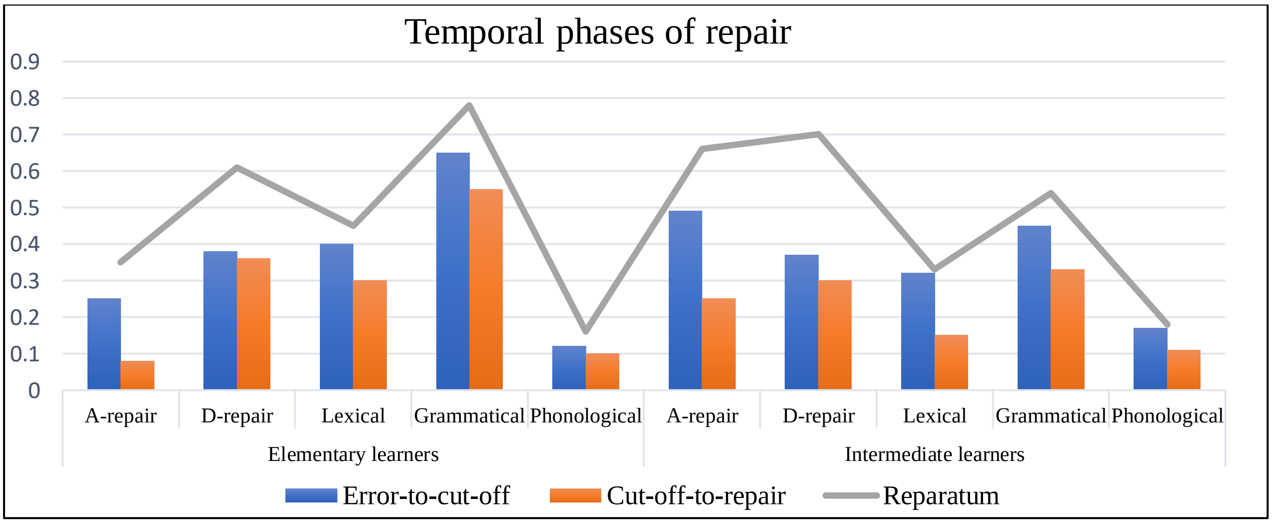


Figure 6.21. Temporal phases of repair in the two proficiency groups

The discussion in this section will focus on three points: first, the emergence of two different patterns of temporal phases in relation to proficiency. Second, the longer duration of two intervals of A-repair (and D-repair in the single task condition) in the Intermediate rather than the Elementary learners. Third, the parallel reaction of the first two phases of repair (Error-to-cut-off and Cut-off-to-repair) in relation to proficiency. These three points will be interpreted in light of the Perceptual Loop Theory (PLT).

According to Levelt (1983, 1989), there are two functions of monitoring: first, checking the speech against speakers' intentions, and the linguistic standards. The second function of monitoring is issuing instructions for correction when a mismatch is discovered (Levelt, 1983, 1989). The focus here is on the first function of monitoring, namely, comparing the speech with the communicative intentions and the linguistic standards. The matching function of the monitor entails checking the inner and outer speech against the outcomes of two stages of speech production, namely, the Conceptualiser and the Formulator. For the purpose of this research, the focus here is on monitoring the overt speech. The first comparison carried out by the monitor is between the utterance and the communicative intention. If a mismatch is detected between the intention and the utterance, the speech is interrupted, and an instruction is sent to the Conceptualiser to either replace the preverbal plan or modify it, and this leads to D-repair or A-repair respectively (Details about these types of repair are in Sections 2.4.1 and 2.5.1). Different-information-repair (D-repair) occurs when the monitor discovers a mismatch between speaker's intention and his/her utterance, and this mismatch requires re-ordering, or replacing the message, as demonstrated in Example 1 below. The following examples are taken from the current data.

Example 1. |Uh there uh /th/ the help comes to us 1.24|

In this example, it seems that there was a mismatch between the speaker's intention and the utterance. The speaker possibly changed her mind and decided to talk about a different idea, thus the speech was interrupted. In this case, the monitor is likely to have issued an instruction to the Conceptualiser to produce a different preverbal plan which matches the speaker's new intention. This type of repair shows that the speaker's intention keeps changing and that the monitor is in a constant state of checking and sending feedback to speech processes.

Another outcome of comparing the utterance to the speaker's intention is A-repair. This kind of repair occurs when the monitor discovers that the utterance does not express the intended message appropriately or accurately, or that the utterance may sound incoherent or ambiguous. Thus, the monitor issues an instruction to the Conceptualiser to reproduce the same preverbal plan with some modifications as indicated in Example 2.

Example 2. |when the stor uh the thunderstorm comes|

In this example, the speaker was describing a picture story which shows a thunderstorm (see Appendix 8-B). The speaker may realise that she needed to say 'thunderstorm' instead of 'storm' so that her speech would clearly describe the scene. As the monitor detected this mismatch, the speech was halted, and it is likely that an instruction was given to the Conceptualiser to modify the preverbal plan. In this case, the preverbal plan was not replaced; only the way to produce this message was modified. In the above examples, it should be noted that the adjustment entails that the Conceptualiser either issues a new preverbal plan (Example 1) or modifies the way to communicate the existing one (Example 2). This shows that D-repair and A-repair involve deviations from the originally intended message which requires replacement or modification.

On the other hand, when the monitor detects a mismatch between the utterance and the linguistic rules (lexical, grammatical, or phonological), instructions are sent to the Formulator to modify the phonetic plan (Levelt, 1983, 1989). This shows that the preverbal plan does not change, only the phonetic plan is edited. see example 3 below.

Example 3. | all of them was uh 0.28 all of them were |

This example shows that as the utterance was produced, the speaker may have realized that there was a grammatical error. The monitor in this case checked the utterance against L2 linguistic system. As an error was detected, an order is likely to have been given to the Formulator to make a correction (i.e. encoding 'were' instead of 'was'). It has been argued that correction of E-repair, which involves real errors, is short and quick (Levelt, 1983, 1989), suggesting that it is processed at the Formulator level rather than involving both the Conceptualiser and the Formulator (Kormos, 2000a). On the other hand, A-repair involves revising a speech plan at the Conceptualiser (Boland, Hartsuiker, Pickering, & Postma, 2005), which is then sent to the Formulator to be re-encoded. This show that A-repair and D-repair involve more complicated processes than E-repair. That is to say, A-repair and D-repair entail replacing or modifying a preverbal plan at the Conceptualiser, then sending the modified preverbal plan to the Formulator to be re-processed before speech articulation. E-repair, on the other hand, involves only processing the error at the Formulator. This likely explains the emergence of two different patterns of temporal phases in the current data.

However, it is still not clear why the temporal phases of A-repair (and D-repair in the single task condition) were delayed in the speech of the Intermediate rather than the Elementary learners. It is worth noting that A-repair is expected to be different from other types of repair because it is not made for corrections but to add further specification to the message (Levelt, 1983). As such, L1 speakers allow themselves greater freedom when making adjustments in A-repair (Levelt, 1983). This suggests that L1 speakers may think of a variety of adjustments in the case of A-repair which makes them spend longer time on A-repair compared to other types of repair. As A-repair is considered demanding in L1 processing (Levelt, 1983), it would be even more so in L2 processing, especially that some encoding processes of L2 speech (e.g., Lexical, grammatical, phonological) are still unstable and lack automatization (Segalowitz, 2010). That is to say, the Intermediate learners may have a more developed linguistic system than the Elementary learners which may afford them a variety of modifications to A-repair compared to the Elementary learners. However, the lack of automatization of speech processes may delay the processing of A-repair in the Intermediate group. On the other hand, due to limited knowledge of the target language of the Elementary learners, they may choose a quick strategy to make modification to A-repair. This finding lends support to the working model of L2 speech production which proposed a separate module for declarative knowledge (Kormos, 2011). The rules that are stored in the declarative knowledge store are not automatized, thus accessing these rules needs conscious effort (Kormos, 2011).

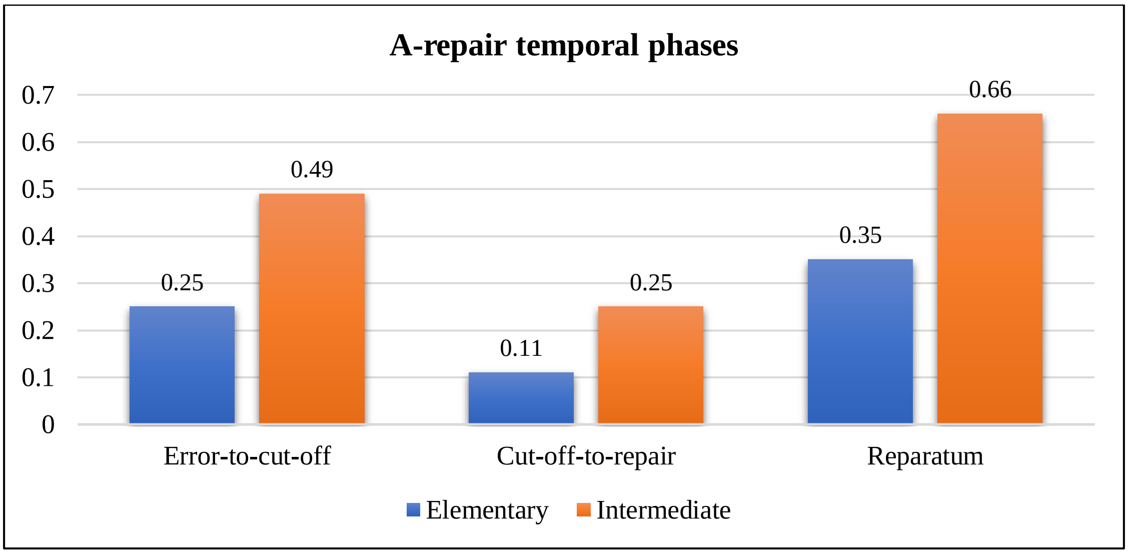


Figure 6.22. A-repair intervals in the two proficiency groups

In other words, as the Intermediate group has potentially acquired a developed knowledge of the target language, their monitoring behavior possibly resembles L1 speakers (to a certain extent), so that they afforded themselves greater freedom in modifying A-repair than the Elementary group. Yet, as their L2 linguistic knowledge is still not automatized, it takes them a long time to process A-repair leading to longer intervals of A-repair compared to the Elementary learners (see Figure 6.22 above).

However, this finding contradicts previous L2 research which found that the Cut-off-to-repair interval correlated negatively with proficiency (Kormos, 2000b). This means that higher proficiency learners produced a shorter interval of Cut-off-to-repair than lower proficiency learners. It has been assumed that due to the automatization of the sub-processes of the Formulator, the higher proficiency learners produced shorter intervals of this phase (Kormos, 2000b). It is worth noting that in Kormos's study, the comparison was made between beginners and advanced learners who may have some of the sub-processes of the Formulator automatized. The Intermediate learners in the current study might have some development in their declarative knowledge, but this knowledge may still not be automatized so that they took a long time to execute A-repair. As these rules become automatized, repair duration is likely to have decreased, and this could explain why the advanced L2 learners in Kormos's (2000b) study spent less time on A-repair interval (Cut-off-to-repair) than the lower proficiency group. This may provide an explanation for the delay in the processing of A-repair intervals (and D-repair in the single task) in the Intermediate group. Further explanation of D-repair intervals will follow in Section 6.4.

The last point to be discussed in this section is the processing of the first two temporal phases of repair (Error-to-cut-off and Cut-off-to-repair) (see Figure 6.23 below). There is an unresolved argument in self-monitoring research about the functions of the first two phases of repair. That is, there is disagreement on whether the Error-to-cut-off (the first phase) is exclusively used for error detection (Levelt, 1983; Van Hest, 1996; Kormos, 2000b), or whether it involves both error detection and planning of repair (Blackmer & Mitton, 1991; Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008; Oomen & Postma, 2001; Pillai, 2006). The former view suggests that these two intervals are processed as two separate units with Error-to-cut-off (the first phase) used mainly for error-detection and that Cut-off-to-repair (the second phase) is used for planning of repair (see Figure 6.23 below).

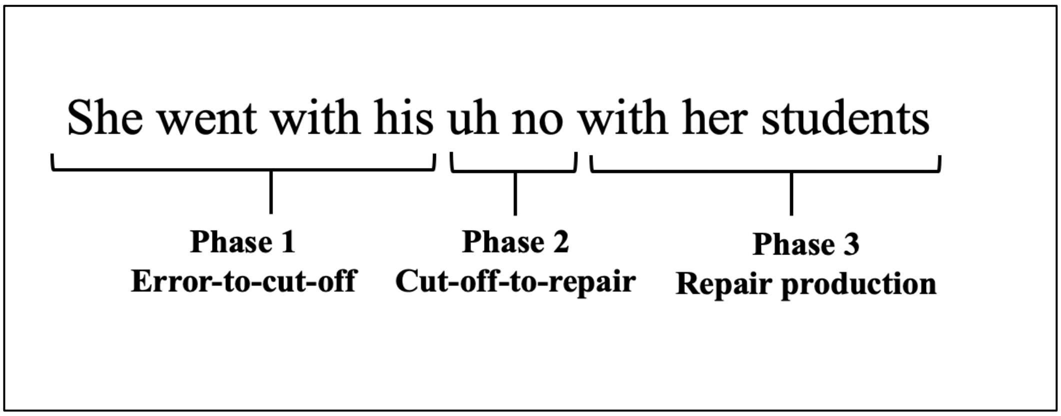


Figure 6.23. Temporal phases of repair

The latter opinion, on the other hand, argues that these two intervals are planned and processed concurrently as a single unit. That is, Error-to-cut-off (the first phase) is not used exclusively for error-detection, but also involves planning of repair which starts at this interval, and continues to the end of the next phase, namely, Cut-off-to-repair. The present data indicated that the two intervals of Error-to-cut-off and Cut-off-to-repair showed similar responses in relation to proficiency and task condition as presented in Section 5.8. The current data suggest that these two intervals involve similar processing, and that any influence on a particular type of repair triggers similar reaction in the two intervals. In other words, if the first phase (Error-to-cut-off) is used solely for error-detection, it should have decreased rather than increased with the development of proficiency as it is expected to be easier for the Intermediate learners to detect an inappropriateness in their speech than the Elementary learners.

However, as this interval (Error-to-cut-off) increased in the Intermediate rather than Elementary learners, this suggests that this interval is not exclusively used for error-detection but probably for planning as well. That is, when the Intermediate learners found it difficult to process a repair, both intervals slowed down. This shows that the Intermediate learners did not immediately interrupt their utterance when detecting an error. They may have interrupted their utterances when they were ready to produce repair. This finding lends support to previous research which assumes that L2 learners only interrupt their utterances when they are ready to make repair (Seyfeddinipur, Kita, & Indefrey, 2008). It also supports the assumption which states that these two intervals of Error-to-cut-off and Cut-off-to-repair are planned and processed concurrently (Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008). Furthermore, the current data lends support to the assumption that difficulty of repair type slows down the first phase of repair (Error-to-cut-off) (Broos, Duyck, & Hartsuiker, 2018; Hartsuiker, Catchpole, De Jong, & Pickering, 2008). That is, the more demanding repair in the current data (A-repair and D-repair) involved slowing down of the first phase (Error-to-cut off). However, no firm conclusion concerning the concurrent planning and processing of the first two phases of repair (Error-to-cut-off and cut-off-to-repair) could be provided based on the current data. Further research is needed to uncover more features of planning and processing of repair.

To conclude, three points have been discussed in this sub-section: the two patterns of temporal phases emerged in the data suggest that A-repair and D-repair involve demanding processing which relates to both the Conceptualiser and the Formulator. E-repair involves processing at the Formulator level. The temporal phases of A-repair (and D-repair in the single task condition) increased in the Intermediate learners due to development of their declarative knowledge. However, as this knowledge has not reached automatization, processing these repair types may be more demanding, and may require a long time. Finally, the first two phases of repair (Error-to-cut-off and Cut-off-to-repair) are likely to be planned and processed concurrently, which suggests that re-planning of the repair starts at the first phase (Error-to-cut-off), and continues to the end of the second phase (Cut-off-to-repair). The next section will discuss the effects of PL on accuracy.

6.2.4 Proficiency and accuracy

It was predicted that the Intermediate learners would produce a higher percentage of error-free clauses and higher ratio of error-correction rather than the Elementary learners. The data showed that the Intermediate group produced a significantly higher percentage of error-free clauses, whereas the ratio of error-correction was similar in the two groups (see Section 5.5.3). The result of the error-free clauses was in line with the hypothesis and with previous studies (e.g., Awwad, 2017). The increase in error-free clauses suggests that there was a potential decrease in errors which probably means that there was development in the underlying speech processes (DeKeyser, 2013), particularly the Formulator sub-processes where most errors occur (Kormos, 2006). That is, the significant increase of the percentage of error-free clauses means that the lexical, grammatical, and phonological encoding processes are likely to have developed so that errors consequently decreased in the Intermediate learners as compared to the Elementary learners. As such, this measure may be sensitive to the development of the underlying speech processes, particularly the Formulator. Concerning the ratio of error-correction, it was expected that it would increase in the same way as the percentage of error-free clauses. That is to say, as accuracy increased in the Intermediate learners, they were expected to detect and correct higher ratio of their errors rather than the Elementary learners. On the contrary, the ratio of corrected errors was the same in the two proficiency groups. This result was in line with the data obtained in Kormos's (2000a) study which examined the ratio of error-correction in three proficiency groups (pre-intermediate, intermediate and advanced L2 learners). The results showed that there were no differences between the three proficiency groups in terms of the ratio of error-correction. On the other hand, in Declerck and Kormos's (2012) study, the ratio of error-correction increased in the advanced rather than the intermediate learners. This was interpreted as meaning that monitoring processes were functioning more efficiently in the advanced than in the intermediate learners (Declerck & Kormos, 2012). However, the lack of agreement in the results of these studies suggests that this measure possibly relates to conscious decisions taken by L2 speakers during L2 processing. In other words, the ratio of error-correction, which is supposed to refer to the accuracy of error-detection (Oomen & Postma, 2001; Kormos, 2006), may increase when L2 learners prioritize accuracy over fluency. In fact, it has been argued that L2 speakers make a conscious decision about correcting errors and the type of errors to be corrected (Mackay, 1992). Speakers' conscious decision may be affected by different factors during speaking such as task type or task condition. In Declerck and Kormos's (2012) study, the use of controlled-network description may prioritize accuracy so that a higher ratio of error correction was produced by the advanced learners in the single task condition (further discussion about the network description task can be found in Section 2.4.3 above and Section 6.4 below). With the increased demand of the dual task condition, cognitive resources may be consumed along the concurrent tasks resulting in less concentration on accuracy. Hence, a lower ratio of error-correction was produced by the advanced learners in the dual task condition compared to single task condition.

To conclude this sub-section, the measures of error-free clauses and ratio of error-correction appear to involve different processing. The percentage of error-free clauses seems to be sensitive to the development of the sub-processes of the Formulator. The ratio of error-correction, which is assumed to be a reliable measure of monitoring efficiency, may relate to the conscious decision taken by L2 speakers. The lack of agreement in the studies examining the ratio of error-correction is partly due to the differences in proficiency levels of participants in these studies, and the wide variations in speaking tasks. That is, this measure might be affected by two main factors which are the degree of automatization and the conscious decisions of L2 speakers. Few L2 studies have employed this measure to examine the differences between self-monitoring behaviour in different proficiency levels (Kormos, 2000; Declerck & Kormos, 2012). The field would benefit from more research that examines ratio of error correction in different levels of proficiency and L1 speech. It would also benefit from stimulated recalls which can highlight the conscious decisions taken by speakers to either correct or ignore their errors.

6.3 Task condition and L2 self-monitoring

The second research question sought to examine the effects of TC on L2 self-monitoring behaviour in terms of disfluency, repair types, temporal phases of repairs, and accuracy. Based on the PLT, it was hypothesized that dual task condition would have detrimental effects on L2 self-monitoring behaviour. That is, there would be more disfluency, longer intervals of repair, fewer number of repairs (A-repair), and a lower rate of accuracy and error-corrections in the dual task compared to the single task conditions. In contrast to the hypothesis, the data showed that task condition had no meaningful effects on disfluency, repair types, temporal phases of most repairs, and accuracy. The only exception was that task condition slowed down the first two intervals of A-repair, namely, Error-to-cut-off and Cut-off-to-repair. This section focuses on the following points: first, it discusses the current data in relation to previous research, and then explains the lack of effect of task condition on most monitoring aspects. This is followed by presenting an explanation for the delay in A-repair temporal phases in the dual task compared to the single task condition.

It was hypothesised that consuming cognitive resources along the dual task condition would have a negative impact on different aspects of monitoring. However, the data indicated that no effect was observed on monitoring aspects apart from slowing down the first two intervals of A-repair. These findings are in line with the data found in Declerck and Kormos (2012). In their data, the impact of TC was only observed on the ratio of the error-correction and lexical errors. That is, there was a significant decrease in the ratio of error-correction and an increase in lexical errors in the dual task condition compared to the single task condition. This was interpreted as meaning that the efficiency of monitoring deteriorated in the dual task as compared to the single task condition. As the current study did not examine the frequency of errors, the discussion will only focus on contradictory results in terms of the ratio of error-correction in the two studies. The lack of agreement between the two studies can be explained in terms of the tasks used. Declerck and Kormos's (2012) task involved a restricted network-description which required participants to produce a limited set of utterances which focused mainly on colours and directions. In such a task, utterances can be either right or wrong (Van Hest, 1996). It has been argued that tasks with a controlled setting elicit more errors than narrative tasks which afford abundance of description and specification (Van Hest, 1996). In the case of producing a wrong expression in network-description, participants need to correct it immediately to be able to complete the task successfully. Thus, this task is likely to have encouraged speakers to prioritize accuracy, resulting in a great ratio of error-correction in the single task condition. Then, with the increased demand of the dual task condition, the participants encountered a demanding situation that is likely to have divided their cognitive resources along the concurrent tasks, hence a lower ratio of error correction was observed. However, the task in the current study seems to allow participants more freedom in choosing what aspects of their speech to attend to. This could account for the lack of considerable differences between the two task conditions in terms of the ratio of error-correction in the current study. A significant question to ask in this context is why L2 monitoring aspects are not affected by the increased demand of the dual task condition despite the fact that self-monitoring is largely drawing on cognitive resources.

There are two possible explanations for the lack of influence of task condition on L2 monitoring aspects. On the one hand, it has been argued that even in the single task condition L2 speech processes require substantial cognitive resources, and therefore, performing in the dual task condition might not lead to noticeable effects on these speech processes (Declerck & Kormos, 2012). That is to say, in the case of L2 speech where cognitive resources are already consumed, the increased demand of task condition would have no considerable impact on L2 monitoring. However, introducing cognitively demanding tasks in the L1 context where most speech processes are automatic (running independently of cognitive resources), led to the same results. For instance, when Oomen and Postma (2001) examined the effects of normal and fast conditions on monitoring aspects, no effect was reported on ratio of error correction or disfluency; only temporal phases of repair slowed down. The only L1 study which found an effect of dual task condition on ratio of error-correction used restricted network-description (Oomen & Postma, 2002). This lends support to the argument made above about the potential role of the network-description task in increasing the ratio of error-correction in Declerck and Kormos's (2012) data. In short, it could be assumed that the lack of effect of the dual task condition on monitoring aspects is not exclusive to the L2 context, as similar results were found in L1 research. This assumption contradicts the theory of Declerck and Kormos (2012) that L2 processing resources are already consumed, thus the dual task condition would not lead to an observable impact on monitoring aspects. Another possible explanation for the lack of differences between the two task conditions might be that with the increased cognitive demand of the task condition, the cognitive resources is taxed but the monitor becomes active in the demanding task condition, so that, no noticeable differences were observed between the two task conditions. That is to say, in the dual task condition, the monitor was able to correct the same amount of errors, make the same amount of repair, maintain the rate of accuracy, and fluency even with the increased demand of the task condition. Self-monitoring, as a conscious process, might be alert to the difficulty of the situation so that cognitive resources are directed towards some aspects of speech. Support for this assumption is discussed in the next section (Section 6.4).

The only effect of the dual task condition in this study was observed on the duration of the first two phases of A-repair. The present study showed that A-repair temporal phases were substantially affected by the dual task condition as illustrated in Figure 6.24 below.

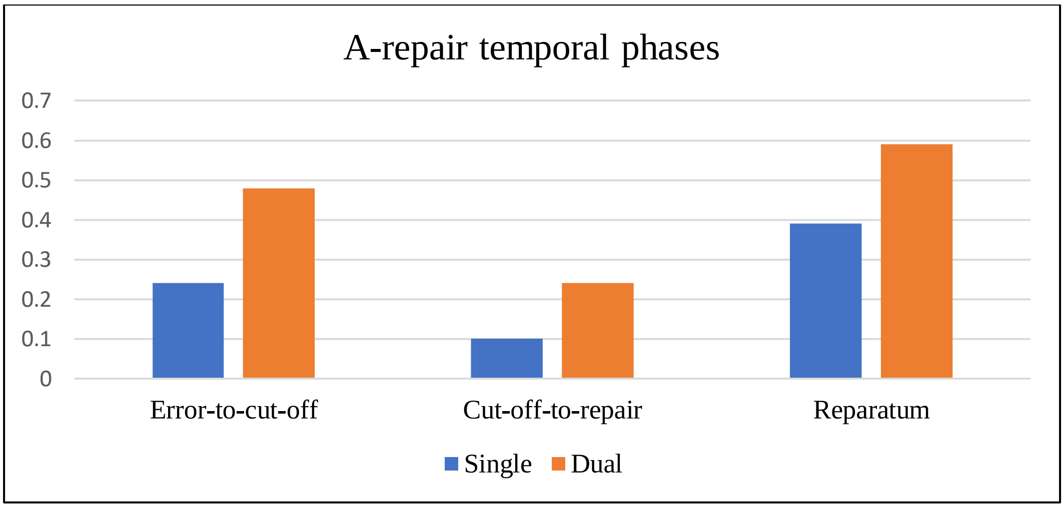


Figure 6.24. Effects of TC on A-repair temporal phases

It was shown in Section 6.2.3 that A-repair and D-repair involve processing at the Conceptualiser level. The Conceptualiser is widely presumed to be the slowest stage of speech production because it involves controlled processing, that is, processing that draws on cognitive resources (Levelt, 1989; Felker, Klockmann, & De Jong, 2019). Therefore, A-repair, which is processed at the Conceptualiser level, entails slower processing than other repair types (Blackmer & Mitton, 1991; Oomen & Postma, 2001; Van Hest, 1996). This suggests that in a situation where cognitive resources are depleted, it is likely that A-repair temporal phases become vulnerable to the effects of the demanding task. Although D-repair is supposed to draw on the Conceptualiser as well, the impact of TC on D-repair interacted with the influence of PL. This will be discussed in the next section (Section 6.4). This data suggest that E-repair sub-categories are less dependent on cognitive resources, thus their temporal phases were not affected by the dual task demand as much as A-repair and D-repair. To conclude, the lack of noticeable effects of task condition on monitoring aspects reveals that monitoring functions did not deteriorate as a result of the increased demand of task condition. In contrast, the monitor may become active and make the same amount of repair, and maintain the same rate of accuracy and fluency. Despite the fact that A-repair involves controlled processing, the dual task condition did not affect the frequency of A-repair, it only slowed down A-repair temporal phases.

6.4 Interaction effects between PL and TC on L2 self-monitoring

The third research question sought to examine whether there was an interaction effect of PL and TC on L2 self-monitoring behaviour in terms of disfluency, repair types, temporal phases of repairs, and accuracy. It was expected that there would be interaction effects of proficiency level and task condition on monitoring behaviour. That is to say, it was expected that the higher proficiency participants would perform better than the lower proficiency participants under the dual task condition. The data demonstrated that there were interaction effects of PL and TC on filled pauses and D-repair intervals (see Section 5.4). It is interesting to find that the single task condition triggered differences between the two proficiency groups in terms of filled pauses and D-repair intervals, and these differences disappeared when they performed in the dual task condition (see Section 5.7). This section explains the meaning of the interaction effects of PL and TC (on filled pauses and D-repair intervals), and the processing of D-repair. It then examines an assumption about the function of the monitor under the dual task condition.

To explain the meaning of the interaction effects on filled pauses and D-repair, it is important to look at proficiency groups separately. The data indicated that the Intermediate group produced significantly more filled pauses in the dual task condition compared to the single task condition, which could be interpreted in two different ways as discussed in Section 6.2.1 above. First, filled pauses may be used by the Intermediate group as a strategy to cope with the increased demand of the dual task condition. Alternatively, the Intermediate learners might be concerned with accuracy, thus they used filled pauses as corrective measures to expected errors (Levelt, 1983). That is to say, due to the increased demand of the dual task condition, the Intermediate learners might expect errors to occur, and so they produced a substantial number of filled pauses to monitor their speech and maintain accuracy. This increase in filled pauses in the Intermediate group reduced the difference that was observed between the two groups in the single task condition, as can be seen in Figure 6.25 below.

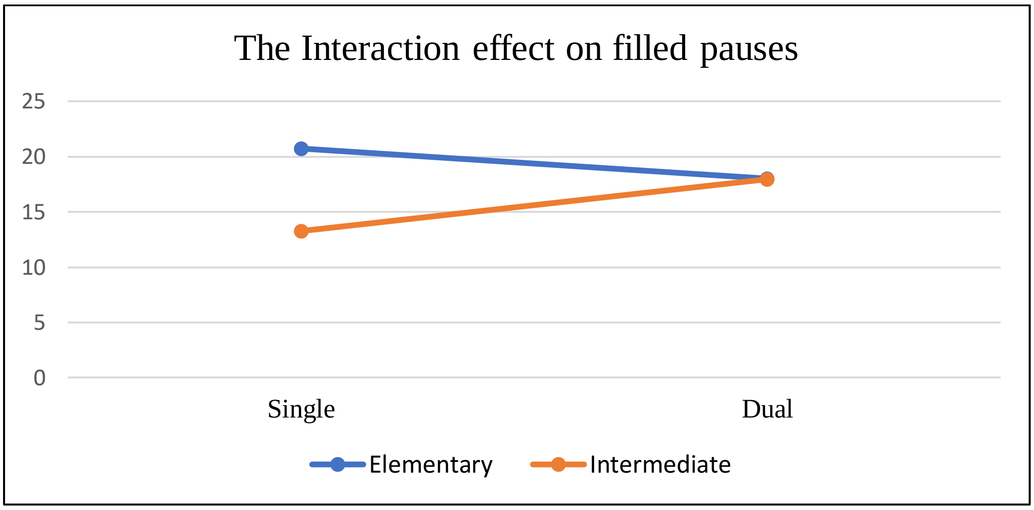


Figure 6.25. The interaction effects of PL and TC on filled pauses

This suggests that lack of overall differences between the single and the dual task conditions does not necessarily mean that there were no meaningful variations in the monitoring behaviour of the two groups in the two task conditions. It could, nonetheless, mean that each proficiency group may adjust their monitoring behaviour to cope with the dual task demand, resulting in substantial variances between the two task conditions within a group (either Elementary or Intermediate). This can be seen also in D-repair temporal phases. The data showed that there were significantly shorter intervals of D-repair in the Elementary group compared to the Intermediate group in the single task condition, despite the fact that the overall result showed that there were no significant differences between D-repair intervals in the two task conditions. In short, the reason for the lack of differences between the two task conditions was that the differences between the two proficiency groups disappeared when they performed in the dual task condition. The follow-up tests indicated that each proficiency group resorted to a different strategy to cope with the increased demand of the dual task condition, resulting in substantial differences in certain monitoring aspects in the two task conditions within a group. The overall results, which combined the data of the two proficiency groups as one cohort, concealed important information about monitoring behaviour of the two proficiency groups in the two task conditions.

Concerning D-repair processing, the data demonstrated that the Intermediate group produced longer intervals of D-repair than the Elementary group in the single task condition, which could be interpreted in the same way as A-repair in Section 6.2.3.

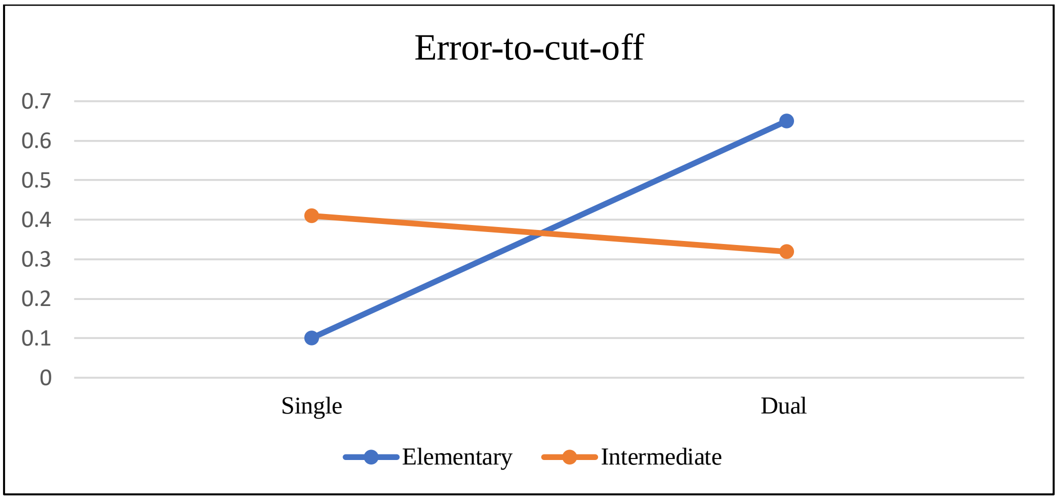


Figure 6.26. The interaction effects of PL and TC on Error-to-cut-off of D-repair

Like A-repair, D-repair involves controlled processing which is considered demanding. Similarly, because D-repair does not involve correcting errors, L2 Intermediate learners may afford themselves greater freedom in making adjustment to D-repair compared to the Elementary learners as discussed in 6.2.3. However, with the increased cognitive demand in the dual task condition, their processing of D-repair may be slowed down, which concealed the differences that were observed between the two groups in the single task condition, as demonstrated in Figures 6.26 above and 6.27 below.

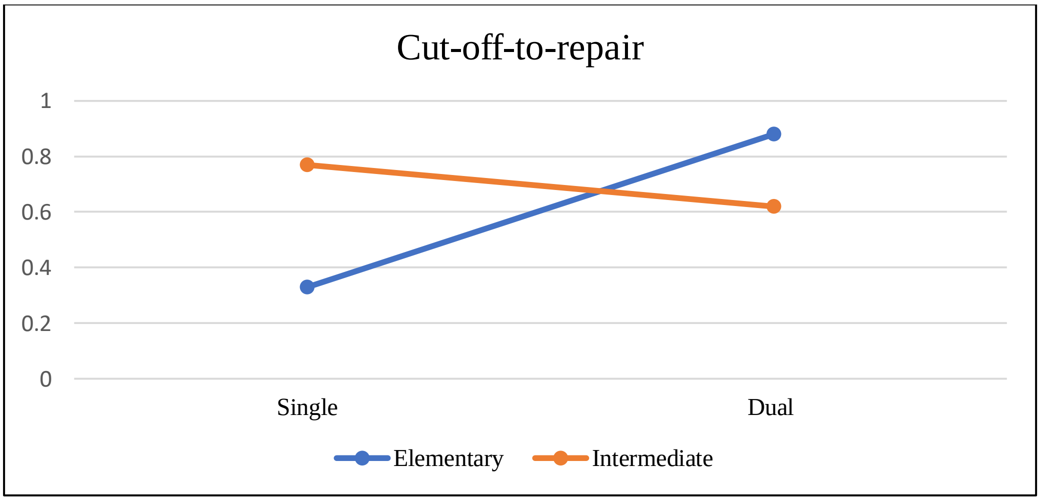


Figure 6.27. The interaction effects of PL and TC on Cut-off-to-repair of D-repair

It should be noted that the differences between the two proficiency groups in the dual task condition in terms of D-repair temporal phases were not significant. This is because the performance of the Intermediate learners was not affected by the dual task condition in terms of D-repair intervals, whereas the Elementary learners was substantially influenced by the increased demand of the dual task condition.

The parallel reaction of D-repair intervals (Error-to-cut-off and Cut-off-to-repair) in the Elementary group in the two task conditions lends support to the assumption made in the previous section (Section 6.3) that these two intervals are planned simultaneously and that the first phase of repair (Error-to-cut-off) can be used for both error-detection and planning of repair (for information about these phases, see Figure 6.23 above). A question that arises here is whether D-repair intervals were affected by TC in the same way as A-repair, given that they possibly entail the same processing as shown in Section 6.3. It seems that D-repair may involve certain variances from A-repair. Figures 6.26 and 6.27 above showed that the first two phases of repair slightly dropped when the Intermediate group performed in the dual task condition compared to the single task condition. This was not the case for A-repair intervals which showed a parallel increase in the two proficiency groups when they performed in the dual task condition compared to the single task condition. It could be argued that the nature of D-repair processing, which involves replacing the preverbal message rather than modifying the way to communicate it (as discussed in Section 6.2.2), may trigger different conscious decisions from the two proficiency groups which, therefore, affect the processing of D-repair intervals.

The differences between the two proficiency groups in the single task condition suggest that the less demanding task condition may trigger differences between the Elementary and Intermediate groups. The fact that there were no differences between monitoring aspects in the two task conditions suggests that the monitor may become robust with the increased cognitive demand in the dual task condition (see Section 6.3). That is, it maintained fluency and accuracy. Moreover, it made the same number of repair and corrected the same ratio of errors. As self-monitoring draws on the comprehension system and is affected by the conscious decisions of speakers, it could be assumed that the monitor may become sensitive with the increased cognitive demand of speaking task. This assumption was first introduced by Oomen and Postma (2001) when they found that L1 speakers corrected the same number of errors in the fast and normal task conditions. They argued that with the increased demand of the task, the monitor becomes so sensitive to errors that it detected the same amount of errors in both task conditions and adjusted its speed according to the speed of the task condition (Oomen & Postma, 2001). It is worth noting that in the current data, further monitoring aspects apparently increased in the dual task condition when examining each proficiency group separately. Despite the fact the no further statistical analyses were conducted to examine these aspects, it could be noted, in the descriptive statistics in Section 5.2, that a similar pattern emerged with regard to A-repair, D-repair, and the percentage of Error-free clauses. It could be seen in Figure 6.28 below that A-repair and D-repair increased in the Elementary group when they performed in the Dual task compared to the Single task condition.

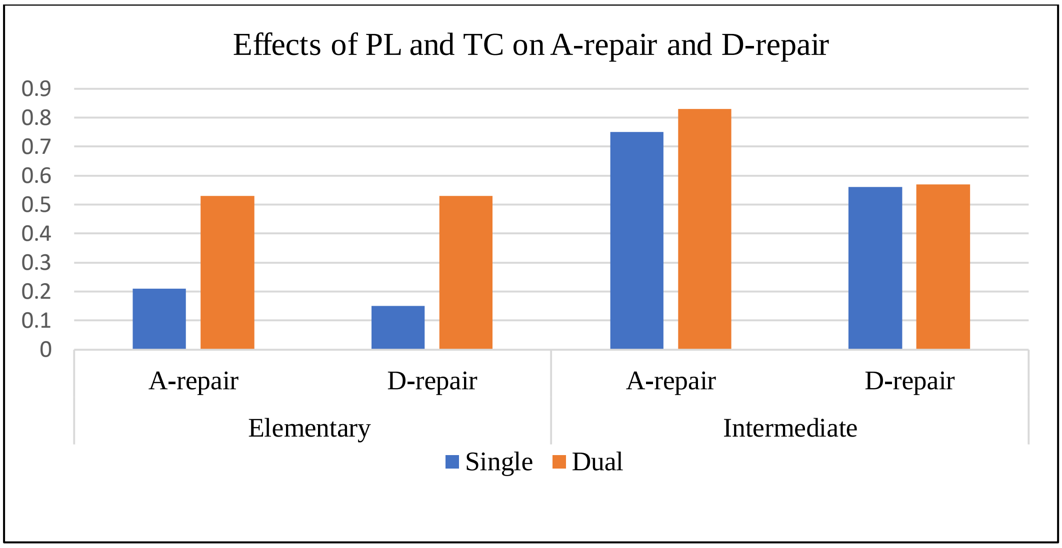


Figure 6.28. Effects of PL and TC on A-repair and D-repair

There was also a slight increase in A-repair and D-repair in the performance of the Intermediate group in the dual task condition compared to the single task condition. More importantly, there are observable differences between the two groups when they performed in the single task condition, but these differences seem to be reduced in the dual task condition.

Furthermore, Figure 6.29 below shows that the percentage of error-free clauses apparently increased in the performance of the Intermediate group in the dual task condition compared to the single task condition. This data is in line with the pattern detected in the interaction effects which suggest that the monitor becomes active with the increased demand of the dual task condition. This is also in line with the data of Levelt et al., (1999) which reported that the monitor becomes intense in the more demanding task condition. This means that the auditory loop of the monitor may operate actively with the increased cognitive demand of task condition so that it detects the same number of errors and maintains accuracy and fluency. Employing the interaction analyses in the current study highlights an important aspect of self-monitoring behaviour. That is, the interaction analyses demonstrated the differences in self-monitoring behaviour within each proficiency group as well as between the two groups in the two task conditions.

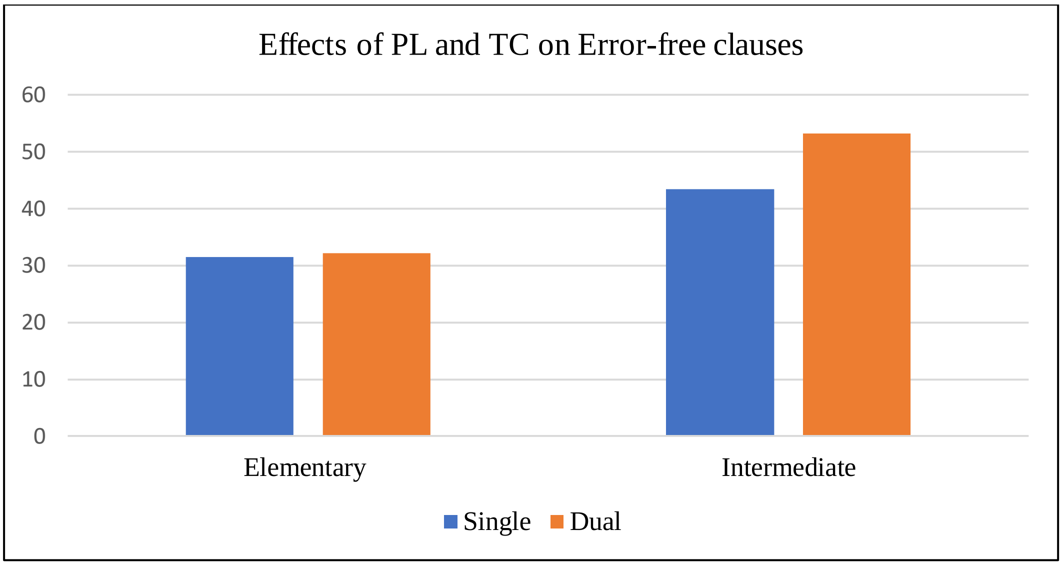


Figure 6.29. Effects of PL and TC on percentage of error-free clauses

This highlights the importance of using interaction analysis in investigating self-monitoring. The interaction analysis reveals not only the impact of individual difference and the contextual factor on performance, but also the process that associates them (DeKeyser, 2012). The process under investigation is treated like 'a black box', and that understanding it is useful for generalizability of findings (DeKeyser, 2012, p. 190). In other words, in the current context, the interaction effects between PL and TC reveal important information about self-monitoring behaviour in L2 learners that can be useful for theory, methodology, and pedagogy as will be discussed in Sections 7.4 and 7.5.

In short, one of the main premises of the PLT is that self-monitoring largely draws on cognitive resources (Levelt, 1983). It was expected that taxing cognitive resources along a dual task paradigm would have a negative impact on L2 self-monitoring (Oomen & Postma, 2002). Despite the fact that the overall findings showed no considerable differences between monitoring aspects in the single and dual task conditions, the interaction effects revealed that the single task condition triggered differences in filled pauses and D-repair intervals between the performances of the two groups. However, these differences disappeared when performed in the dual task condition. Examining the effects of TC on the two proficiency groups separately suggested that with the increase of cognitive demand, the monitor probably became active and previously observed differences disappeared. In fact, this suggests that the functioning of L2 self-monitoring might not necessarily deteriorate with the increased demand on cognitive resources. Self-monitoring as a conscious process might be affected by the decisions taken by L2 learners while they perform in the demanding task condition. Figures 6.28 and 6.29 above demonstrated that the monitor may become intense in the dual task condition as compared to the single task condition in the two proficiency groups. However, this assumption needs to be examined in future studies by studying the effects of the dual task condition on each proficiency group separately. The findings in this section support the supposition that interaction analysis is a helpful tool in exploring language processing (DeKeyser, 2012).

6.5 Working memory capacity and L2 self-monitoring behaviour

The fourth research question investigated the relationship between working memory capacity (WMC) and L2 self-monitoring in terms of disfluency, repair types, temporal phases of repair, and accuracy in the two task conditions. It was hypothesised that a larger WMC would enhance L2 self-monitoring in the single task condition rather than dual task condition (see Section 4.3). Back Digit Span Test (BDST) in L2 English was used to assess WMC (see Section 4.6.4). The results (in Section 5.9) were in line with the hypothesis stated above. The data showed that significant negative correlations were found between WMC and repetition on the one hand, and WMC and lexical-repair on the other in the single task condition rather than in the dual task condition. That is, L2 learners with a larger WMC produced a fewer number of repetition and lexical-repair in the single task condition than in the dual task condition. However, these correlations were small, suggesting that individual differences in WMC explain small variations in repetition and lexical-repair in the single task condition.

As the BDST is supposed to measure the temporary storage and processing of information (William, 2012), the results could be interpreted in light of the Perceptual Loop Theory. Working memory is the cognitive system that is crucial to the functioning of L2 self-monitoring. That is to say, during L2 processing, a phonetic plan is stored and held in the pre-articulatory buffer to be inspected by the inner loop of monitoring before articulation (Levelt, Roelofs, & Meyer, 1999; Postma, 2000; Postma & Noordanus, 1996). When the monitor detects an error in the internal speech, it sends an alarm signal to the working memory which issues instructions to the speech components to make a repair (Levelt, 1989). Some errors are likely to be corrected before articulation (Pillai, 2006; Postma & Kolk, 1993). It could be assumed that larger WMC, which affords more cognitive resources, may facilitate the monitoring's function, particularly the inner loop in detecting and correcting lexical errors during the pre-articulatory stage (see Section 2.3.3), so that less lexical-repair is produced. One may ask why WMC was implicated in lexical-repair and not any other type of repair in the current data. As self-monitoring depends on the limited-capacity resources system, namely, working memory, it can attend to only few items at a time, namely, 'items in the working memory' (Levelt, 1989, p, 20). According to VanPatten (1996, 2002), L2 learners start by processing lexical elements (content words) before anything else during L2 processing, to comprehend meaning in the input. This might also be true for L2 speech production. It could be assumed that to ensure that meaning is communicated effectively, L2 learners may focus on lexis more than other aspects of speech. As such, during the pre-articulatory monitoring, the inner loop is likely to focus on revising and correcting lexical errors resulting in less lexical-repair in the overt speech.

The PLT considers disfluency as a corrective measure to expected errors (Levelt, 1983). In other words, frequent disfluency (pauses and repetition) is expected to be used to provide speakers with time to deal with any anticipated processing problem. In this context, the larger capacity of working memory should give an advantage to pre-articulatory monitoring, so that some processing problems might be solved prior to articulation. Hence, less repetition was produced in the overt speech. The analyses yielded significant weak correlations between lexical-repair, repetition and WMC in the single task condition, but not in the dual task condition. This suggests that during the single task condition, available working memory resources might have supported the function of the inner loop (pre-articulatory monitoring), which led to reducing repetition and lexical-repair in L2 overt speech as explained above. However, during the dual task condition, working memory resources were greatly consumed along different aspects of processing; therefore, no extra resources were available for the pre-articulatory monitoring. This may explain why lexical-repair and repetition did not decrease in the overt speech in the dual task condition.

Taken together, these results contradict previous studies that investigated the relationship between self-repair and WMC (e.g., Georgiadou & Roehr-Brackin, 2017; Mojavezi & Ahmadian, 2014). In Mojavezi & Ahmadian's (2014) study, there was a positive significant relationship between WMC and E-repair which was interpreted as meaning that L2 learners with larger WMC produced more E-repair (lexical repair included). Georgiadou and Roehr-Brackin (2017) did not find any relationship between WMC and self-repair which was interpreted as implying that self-repair may be mainly affected by the conscious decisions of L2 speakers as to whether to correct errors or not. However, in Georgiadou and Roehr-

Brackin (2017), it was found that WMC correlated negatively with one aspect of disfluency (filled pauses) in the intermediate L2 learners. They assumed that larger WMC ‘should constitute an advantage, resulting in less need to pause in order to create extra time for conceptualisation and formulation’ (Georgiadou & Roehr-Brackin, 2017, p. 891). This particular finding is in line with the negative correlation between WMC and repetition in the current data which suggests that larger WMC may have an association with disfluency.

The mixed findings in these studies in terms of self-repair suggest that different factors may confound the results. For example, the test used to measure WMC in the different studies might produce confusing effects. Mojavezi & Ahmadian (2014) used a working memory listening span test where L2 learners were given sets of sentences in L1 Farsi. They were also required to write down the last word in each sentence of the set. Georgiadou and Roehr-Brackin (2017) used a word recall test in L2 English, a backward digit span test in L1 Arabic, and a listening span test in L2 English. The current study used BDST in L2 English. It can be seen that the relationship between WMC and self-repair was assessed through a variety of tests which could probably assess various constructs as well as WMC. Task type is another factor that may have an effect on the result. As it is suggested that task type affects the foci of self-repair (Van Hest, 1996), it could be assumed that a variety of task types and task conditions could mediate the relationship between WMC and self-repair as well. In Mojavezi & Ahmadian’s (2014) study, participants were asked to narrate video cartoons in careful online planning, whereas in Georgiadou and Roehr-Brackin (2017), L2 speech was elicited through interviews. In the current study, L2 learners were asked to describe picture stories under single and dual task conditions. Different contexts could probably trigger different reactions from working memory that could be depicted differently in the correlation analyses. The mixed results of these studies did not help to understand how WMC is implicated in L2 self-monitoring functions.

To conclude, the lack of agreement in L2 studies concerning the relationship between WMC and L2 self-repairs suggests that working memory as a cognitive construct might be implicated differently in different contexts, and such reaction might not follow consistent patterns in relation to different speech aspects. L2 research in this area mainly uses correlation analysis to examine the relationship between WMC and different aspects of L2 speech. Using different statistical analyses may yield important information about the way WMC functions in different contexts. For instance, exploring the interaction effects between WMC and other factors (e.g., proficiency level, task condition) may uncover how WMC interacts with other factors. Interaction analysis can enrich our understanding of both WMC, and self-monitoring behaviour as explained in Section 6.4.

6.6 Conclusion

This chapter has discussed the results of the study which were obtained to answer the four research questions. The findings were explained and discussed in detail with regards to the hypotheses, the PLT, L2 speech production working model, and previous research on self-monitoring and working memory. The next chapter will present the implications and contributions of this study to an L2 speech production model and research. The limitations of the study will be mentioned, and suggestions for future research made.

Chapter 7: CONCLUSION

7.1 Introduction

This chapter draws conclusions from the findings reported in this study. It then highlights the significance of the results obtained and their contribution to the theory and research on L2 self-monitoring behaviour. This chapter further discusses the implications of the findings for pedagogy and assessment of L2 speech. The chapter concludes with a discussion of the limitations of the study, and proposes some potential areas for future research on L2 self-monitoring.

7.2 Summary of the findings

This study aimed to examine the extent to which L2 self-monitoring behaviour can be affected by proficiency development and task condition (single and dual). The current study also set out to investigate the relationship between working memory capacity and L2 self-monitoring behaviour. To do so, two proficiency groups (elementary and intermediate) were recruited to perform in two task conditions (single and dual) (see Section 4.6.2). Also, participants' working memory capacity was assessed using BDST (see Section 4.6.4). Four research questions and hypotheses were formulated to guide the current investigation (see Section 4.3). It was expected that a higher proficiency level and larger WMC would support the functioning of the inner loop of monitoring, and therefore some aspects of monitoring such as disfluency, certain repair types (E-repair sub-categories), and temporal phases of repair would decrease in overt speech. A higher proficiency level and larger WMC were expected to promote accuracy in L2 speech. The increased demand of task condition, on the other hand, was expected to consume the available cognitive resources and thus increase disfluency, E-repair sub-categories, and temporal phases of repair (see Section 4.3 for further details about the hypotheses).

As indicated in the findings in Chapter 5, there were significant differences between the two proficiency groups in almost every aspect of L2 self-monitoring, which suggests that L2 self-monitoring behaviour is affected by proficiency development. The data showed that the Intermediate group made significantly fewer filled pauses, and more end-clause pauses than the Elementary group. The finding of the filled pauses was in line with the hypothesis. The increase of end-clause pauses in the Intermediate learners as compared to the Elementary learners may mark a shift towards a macro level monitoring, namely, planning and revising the next utterance prior to articulation which is common in L1 speakers (see Section 6.2.1).

Concerning repair types, it was found that the Intermediate group produced significantly more A-repair compared to the Elementary group. The results showed that development of declarative knowledge in the Intermediate group may enable them to check further aspects of their utterances such as the appropriacy of their message. Unexpectedly, the Intermediate group also spent significantly longer time in executing the first two phases of A-repair (the Error-to-cut-off and Cut-off-to-repair phases) (see Section 6.2.3). This suggests that the learners had a higher level of declarative knowledge, yet this knowledge was not automatized, and as such it took the intermediate learners a long time to access this knowledge. On the other hand, due to limited knowledge of the target language of the Elementary learners, they may be choosing a quick strategy to make modification to A-repair. That is, as they had learned a limited number of L2 items, these were familiar to them, probably because of frequent use; thus, accessing these items might not be demanding while speaking.

With regard to the accuracy measures, it was found that the Intermediate learners produced a higher percentage of error-free clauses than the Elementary learners. This suggests that there was development in the L2 linguistic system in the Intermediate group as they were able to produce more accurate language than the Elementary group; this finding is also in line with the hypothesis which stated that disfluency was produced as a corrective action to anticipated errors. In the current data, accuracy significantly increased and filled pauses significantly decreased in the Intermediate group compared to the Elementary group. However, these findings are not conclusive because only filled pauses (among other disfluency measures) significantly decreased in the Intermediate group. In short, the data analysis of the first research question suggests that L2 self-monitoring aspects (disfluency, repair types, temporal phases of repair, and accuracy) were considerably influenced by the development of proficiency level.

The findings of the second research question showed that dual task condition considerably slowed down the first two temporal phases of A-repair (the Error-to-cut-off and the Cut-off-to-repair) (see Section 5.8). This suggests that A-repair involves controlled processing, thus its temporal phases were vulnerable to the increased demand of the dual task condition. The data in this section lends support to the argument that the two temporal phases of the Error-to-cut-off and the Cut-off-to-repair are simultaneously planned and executed. Other aspects of monitoring were not affected by the increased demand of task condition. It was argued that the monitor might become robust to deal with the increased demand of the dual task condition. That is, it could detect and correct the same number of errors, maintain fluency and accuracy in the dual task condition as in the single task condition.

The data analysis for the third research question showed that there were significant interaction effects between proficiency level and task condition on filled pauses and D-repair temporal phases, namely, the Error-to-cut-off and the Cut-off-to-repair intervals (see Section 5.7). The interaction effects indicate that there were considerable differences between the two proficiency groups when they performed in the single task condition in terms of filled pauses and D-repair intervals. However, these differences disappear when they performed in the dual task condition which shows that the single task condition triggered greater differences between the two proficiency groups in terms of filled pauses and D-repair intervals than the dual task condition. The interaction analyses also showed that there were within-group variations. That is, the increase of filled pauses suggests that in the dual task condition participants might expect errors to occur, and so they made more filled pauses to buy time for checking their utterances before articulation. The Elementary group, on the other hand, produced longer intervals of D-repair which suggests that in the more demanding task condition, Elementary learners may slow down their speech to be able to plan their D-repair appropriately. This is in line with the hypothesis that expected that the dual task condition would affect the inner loop. In this case, it resulted in more filled pauses and longer D-repair intervals in the overt speech. This suggests that the auditory loop, to a certain extent, may become intense in the dual task condition compared to the single task condition. This is in line with the data in Levelt et al., (1999) which reported that monitoring becomes intense in the more demanding task condition. In such a case, being intense may means that the auditory loop of the monitor operates actively with the increased cognitive demand of task condition. The data also supports the argument that interaction analysis is a useful tool to highlight the functioning of language processes (DeKeyser, 2012).

Concerning the relationship between WMC and L2 self-monitoring aspects in the two task conditions, significant, but weak, correlations were found in the single task condition between WMC and lexical-repair on the one hand, and WMC and repetition on the other. This was interpreted as meaning that a larger working memory capacity supported the functioning of the inner loop in the single task condition which resulted in less repetition and lexical-repair in the overt speech. The findings are in line with the premises of WMC. That is, as working memory has a limited capacity, only few aspects of monitoring can benefit from the extra cognitive resources, particularly during the stage of pre-articulatory monitoring (via the inner loop). However, the inconsistent data of the relationship between WMC and repair in previous studies and the current study restricts the generalisation of the current findings. The examination of L2 self-monitoring behaviour in the current study has yielded important information for the theory, research, pedagogy, and language assessment as will be discussed in the following sections.

7.3 Contributions of the study

One strength of this study lies in testing some assumptions of the PLT on L2 learners. There have been calls for L2 researchers to test L1 self-monitoring theories in the L2 context (e.g., Kormos, 2000a; Van Hest, 1996). Although there is a number of studies in the field of SLA and TBLT that have investigated self-repair (e.g., Gilabert, 2007; Wang & Skehan, 2014), few of these studies have related their findings to theories of self-monitoring (e.g., Kormos, 2000). Testing the assumptions of the monitoring theories in the L1 context has helped in validating the theoretical principles of these theories (e.g., the PLT). With regard to the L2 context, it has been assumed that the L1 speech production model, which the PLT was based on, can be applied to the L2 context without major changes (Kormos, 2000). This suggests that the theoretical principles of the PLT can be applied to L2 monitoring, as this theory is also based on the monolingual speech production model. The current study found that L2 monitoring behaviour changes with the development of proficiency, which suggests that the PLT's assumptions need to be examined in L2 learners at different proficiency levels. This is an important factor to consider in the development of an L2 speech production model.

The present study also contributes to the bulk of self-monitoring literature by studying self-monitoring as a multifaceted construct. That is, the current study has not used self-repair as the only proxy of monitoring, but it also included a number of aspects including disfluency, temporal phases of repair, and accuracy measures. It is useful to include both disfluency and accuracy when investigating L2 self-monitoring behaviour because one of the premises of the PLT is that disfluency is expected to be a corollary of error occurrence. Repair types and temporal phases of different repair are important aspects of self-monitoring as they enrich our understanding of the development and functioning of L2 self-monitoring (for further details, see Section 7.4). Part of this study has been exploratory in nature as it has employed a systematic examination of the temporal phases of repair. There are only few studies examining temporal phases of L2 repair. The potential reason for lack of research in this area is that the measurement of the temporal phases of repair is time consuming as it entails calculating these phases in seconds using PRAAT software. The fine-grained analysis of different aspects of monitoring in the present study has provided original contributions to SLA research and an L2 speech production working model as can be seen in Section 7.4.

Another contribution of the current study is the investigation of the relationship between WMC and self-monitoring behaviour which has rarely been conducted in L2 research (e.g., Georgiano & Roehre-Brackin, 2017; Mojavezi & Ahmadian, 2014). Individual differences in WMC are particularly relevant to self-monitoring because self-monitoring as a controlled process draws on working memory resources in its functioning. Previous studies investigated few aspects of self-monitoring in relation to WMC such as disfluency and self-repair. The current study, on the other hand, examined several aspects of L2 self-monitoring in relation to WMC (i.e. disfluency, repair types, temporal phases of repair, and accuracy). The current findings are in line with the premises of the limited capacity system of working memory and the controlled processing of self-monitoring. That is, due to the limited capacity of working memory, it can hold only few items at a time (Baddeley, 1996). The present data suggests that the less demanding task condition may provide a beneficial situation for the working memory in relation to the pre-articulatory monitoring.

The final contribution worth highlighting is the operationalisation and assessment of proficiency level in the present study. The operationalisation of proficiency in the current study draws on Hulstijn's definition of proficiency as 'knowledge of language and the ability to access, retrieve and use that knowledge in listening, speaking, reading and writing' (Hulstijn, 2015, p. 21). This definition highlights the important elements of proficiency, relevant to the current investigation, such as linguistic knowledge, and spontaneous processing and production of a language. The aim of the current study was to employ tests that could assess linguistic knowledge, processing and production of L2 speech. Thus, the decision was taken to employ both the C-test and EIT (see Section 4.6.3). Another reason for combining the C-test with EIT was to assess language in both written and oral modalities. Such assessment could address the issue of construct validity which refers to providing an adequate operationalisation of a construct in question as discussed in Section 4.6.3. It could be assumed that using both the C-test and the EIT is likely to provide a comprehensive assessment of language proficiency.

7.4 Theoretical implications

The present study sought to advance our knowledge about L2 self-monitoring behaviour manipulated along proficiency level, dual task paradigm, and working memory capacity. The study has provided a novel contribution to the Perceptual Loop Theory (the PLT), and the L2 speech production working model. One of the main premises of the PLT is that self-monitoring draws largely on cognitive resources (Levelt, 1983, 1989, 1999; Levelt et al., 1999). As L2 speakers' interlanguage is in a constant state of development, it was hypothesised that L2 self-monitoring behaviour would change as a result of proficiency development. The result was in line with this assumption, as different aspects of self-monitoring changed considerably with proficiency development. First, two patterns of disfluency emerged in relation to proficiency level: there was a considerable decrease in filled pauses and an increase in clause-final pauses in the Intermediate group compared to the Elementary group (see Section 5.5). The results were interpreted in light of the PLT which views filled pauses as corrective measures to expected errors (Levelt, 1983). It has been assumed that filled pauses possibly decreased in the Intermediate learners as a result of development of the linguistic system. This finding was supported by a considerable increase of accuracy rate in the Intermediate learners. However, this assumption cannot be entirely verified as Covert-repair was not part of the current investigation. The increase of clause-final pauses was explained as a potential shift towards macro-level planning and monitoring of next utterances which is common in L1 speakers (see Section 6.2.1). The findings also showed that the Intermediate learners produced significantly more A-repair compared to the Elementary learners. It was assumed that the increase of A-repair in the Intermediate learners could partly be explained as a result of the automatization of the sub-processes of the Formulator, and partly because of the development of the declarative knowledge in the Intermediate learners (see Section 6.2.3). The first two temporal phases of A-repair considerably slowed down in the Intermediate learners which suggests that declarative knowledge of the Intermediate learners has developed, but this knowledge may not yet be automatized. In short, it could be assumed that L2 self-monitoring behaviour changes as a result of the automatization of some speech processes and the development in declarative knowledge.

Another principal premise of the PLT states that self-monitoring is sensitive to contextual effects (Levelt, 1983). This principle follows from the assumption that self-monitoring draws on cognitive resources. As L2 processing largely consumes cognitive resources (Kormos, 2006), self-monitoring behaviour is expected to be affected by the influences of different contexts. To examine the impact of resource limitation on L2 self-monitoring behaviour, dual task condition was used in the current study to deplete the available cognitive resources. The data in Section 5.6 showed that there were no significant effects of dual task condition on most aspects of L2 self-monitoring. The only exception was that dual task condition significantly slowed down the first two temporal phases of A-repair. Lack of a significant impact of the dual task condition on most monitoring aspects suggests that the monitoring process was robust enough for the increased demand of task condition. That is, the monitor was able to detect and correct the same number of errors, and maintain fluency and accuracy in the demanding task condition. Despite the fact that the temporal phases of A-repair slowed down, due to its controlled processing, the monitor maintains the same frequency of A-repair in the dual task as compared to the single task condition. Interestingly, the follow-up analyses of the interaction effects revealed important information about the functioning of L2 self-monitoring. It showed that the single task condition triggered variations between the two proficiency groups, but these variations disappeared when they performed in the dual task condition. This suggests that in the single task condition, the learners performed as they were expected, that is, the Intermediate group performed better than the Elementary group. However, the introduction of the dual-task condition pushed the learners out of their comfort zone in using their L2 ability with a particular impact on their monitoring behaviour. Section 5.7 showed that L2 Intermediate learners produced significantly more filled pauses in the more demanding task condition, which suggests that they were actively monitoring their speech, given that filled pauses are assumed to be produced when errors are anticipated to occur (see Section 6.2.3). Although the finding in this regard is not conclusive, it is in line with the assumption that the monitor becomes intense with the increased cognitive demand of task condition (Levelt et al., 1999).

The prediction of the PLT concerning the association between WMC and self-monitoring has received support from the current data. According to the PLT, self-monitoring is restricted by the limited capacity of working memory as it can deal with only few items of internal or external speech at a time (Levelt, 1983). In the L2 speech processes, cognitive resources are already exhausted along several processes (grammatical, lexical, phonological) as these processes are not yet automatized especially at the lower level of proficiency (Kormos, 2006). This means that L2 speakers may be paying attention to few aspects of their speech during self-monitoring. The data showed that significant negative weak correlations were found between WMC and repetition, as well as WMC and lexical-repair in the single task condition rather than the dual task condition. That is, L2 learners with larger WMC produced a fewer number of repetition and lexical-repair in the single than in the dual task conditions. The results suggest that during the single task condition, extra working memory resources might enhance the function of the inner loop resulting in less repetition and lexical-repair in L2 overt speech, as explained in Section 6.5. However, during the dual task condition, as working memory resources were more actively engaged, the pre-articulatory monitoring was not supported.

These variables draw on the main principle of the PLT which states that self-monitoring is dependent on cognitive resources. As these variables relate to cognitive resources in different ways (see Sections 2.8- 2.10), it was expected that these variables will affect the functioning of L2 self-monitoring behaviour. PL and WMC are individual difference factors that relate to learners themselves whereas TC is a contextual factor that was manipulated to increase the demand on cognitive resources. The data revealed that WMC as an individual factor does not have a considerable impact on self-monitoring behaviour. TC as a contextual factor also did not produce significant effects on different aspects of monitoring with the exception of slowing down some temporal phases of certain types of repair. The only factor that is considerably important in influencing self-monitoring behaviour is PL as it significantly affects the main aspects of L2 self-monitoring behaviour (disfluency, repair types, temporal phases of repair, and accuracy). The question that can be asked here is how proficiency development leads to changes in L2 self-monitoring behaviour? Researchers argue that the development of proficiency level leads to automatization of some sub-processes of speech production, and thus freeing some cognitive resources to support the functioning of self-monitoring (Kormos, 2006; Van Hest, 1996). Proficiency development, however, influences not only availability of cognitive resources but also the development of learners' knowledge of the target language (see Section 6.2.2). The current data suggests that development of declarative knowledge accounts for the significant increase of A-repair in the Intermediate learners (see Section 6.2.2). Thus, it cannot be claimed that the availability of cognitive resources is the only link between L2 self-monitoring and PL. As such, researchers need to consider the development of declarative knowledge as an important factor when investigating L2 self-monitoring behaviour.

The results of the study also lent support to the L2 speech production working model. It has been assumed that L2 speech system has a separate module called declarative knowledge store which holds rules and standards of production which are recently acquired by L2 learners, but still not automatized (Kormos, 2011). According to the L2 speech production model, accessing the rules in the declarative knowledge store needs conscious effort (Kormos, 2011). The Intermediate learners in the current study produced substantially more A-repair and spent significantly longer time processing A-repair as compared to the Elementary learners. The declarative knowledge in the Intermediate learners might have developed but this knowledge may not be automatized yet. As such, these rules are probably held in the declarative knowledge store where accessing these rules would need more time. This could explain why executing A-repair is longer in the Intermediate learners as compared to the Elementary learners (see section 6.2.3). The data also indicated that A-repair is different from other types of repair. A-repair involves more complicated processing that makes it vulnerable to the increased demand of the task condition. This suggests A-repair is likely processed at the Conceptualiser level. This is in line with the processing of A-repair in L1 speech (see Section 6.2.3).

Moreover, this study has argued that the substantial increase in A-repair in the Intermediate learners might likely be affected, among other factors, by the development of declarative knowledge of the target language (see Section 6.2.3). Previous studies have assumed that the automatization of the sub-processes of L2 speech leads to freeing cognitive resources which likely result in an increase in the number of A-repair (Kormos, 2000a; Van Hest, 1996). The current study has provided an empirical evidence that the frequency of A-repair was not affected by consuming available cognitive resources along the dual task condition (see Section 6.3). As such, it has been assumed that availability of cognitive resources is not be the only factor that accounts for the considerable increase of A-repair. The development of knowledge of the target language needs to be considered when interpreting the increase of A-repair in L2 speech. To conclude, the implications of the current data should be considered with cautions as more studies are still needed to validate the current findings.

7.5 Implications for pedagogy and assessment

The current data have important implications for different dimensions of L2 pedagogy and language assessment. One pedagogical implication relates to the role of working memory in language learning. Teaching and learning a second language need to benefit from the principles of WMC into implementing teaching and learning materials. The present data demonstrated that the less cognitively demanding task condition (single task condition) created slightly less load on working memory resources. As such, the single task condition slightly enhanced the pre-articulatory monitoring by providing extra resources. The results suggest that the less demanding task condition induces L2 learners to revise and check few aspects of their speech prior to articulation which perhaps contributes to noticing gaps in their linguistics repertoire. Previous research has argued that noticing gaps in the knowledge of the target language can trigger further learning processes (DeKeyser, 2007; Izumi, 2003; Kormos, 1999, 2006; Swain, 1985). Tasks that are excessively taxing the cognitive resources of working memory might not help the development of the interlanguage (Michel et al., 2019). Moreover, the cognitively demanding tasks could negatively affect L2 learners with lower WMC in classroom and testing contexts (Michel et al., 2019). As it was discussed earlier, the more L2 learners monitor their speech, the more likely language acquisition occurs (DeKeyser, 2007, 2010). It could then be argued that less demanding task condition could have beneficial effect for language development.

The current finding has another important implication for EFL classroom practices. As self-monitoring is supposed to enhance language acquisition, self-monitoring needs to be encouraged in EFL classroom. L2 learner may notice their errors but choose to ignore these errors for different reasons. For example, L2 learners might want to sound more fluent (Lennon, 1990); do not want to slow down their speech (Kormos, 2006); or want to avoid embarrassment as self-correction might direct listeners' attention to their errors (Krashen, 1981). In fact, self-repair is classified as an aspect of disfluency which negatively affects the smooth flow of speech (Lennon, 1990; Skehan, 2003; Tavakoli & Skehan, 2005). One of the main goals of classroom practices is to develop fluency of L2 speech production which necessarily means reducing self-repair. Hence, disfluency is not encouraged in EFL classroom, and often avoided by L2 learners (Kormos, 2006; Krashen, 1981; Lennon, 1990). However, the present data showed that the Intermediate learners produced significantly more A-repair than the Elementary learners, but they need longer time to execute A-repair. This suggests that their declarative knowledge has developed but accessing this knowledge still needs conscious effort. It is recommended that EFL teachers need to encourage self-corrections and allow L2 learners ample time to plan and execute their self-repair. For example, teachers can use some strategies that promote self-corrections such as making face expressions, gestures, or asking students to repeat what they said. Using these strategies can direct learners to their errors and give them a chance to think of correcting these errors themselves.

Furthermore, the present findings have implications for language assessment. The data indicated that A-repair considerably increased with the development of proficiency level. That is, the Intermediate group produced significantly more A-repair than the Elementary group (see Section 5.5.2). This data is in line with the previous research which argues that self-repair can be used as a useful predictor of language development (DeKeyser, 2010; Kormos, 2006; Tavakoli, Nakatsuhara, & Hunter, 2017). As A-repair is mainly produced to enhance the way messages are communicated, the increase in A-repair suggests that there is development in L2 learners' linguistic system. Thus, language assessment needs to distinguish between self-repair that is made to correct linguistic errors, and self-repair that targets conceptual aspects of the message. The current assessment criteria give credit to producing fluent speech which is characterized by a smaller number of self-repair (e.g., the CEFR). L2 learners, who are keen on enhancing the way their messages are communicated, may unfairly be assessed according to the current assessment criteria. Language assessment needs to take into account repair types to ensure that reliable measurement of learners' linguistic ability is achieved.

The current study has implications for TBLT as it provided a detailed description of designing two comparable tasks following the guidelines proposed by De Jong and Vercelloti (2016). The practical application of this guideline in the present study can be a useful example to curriculum designers and TBLT practitioners whose goal is to design and sequence tasks within syllabus design. Narrative picture prompts can be made more difficult in several aspects including storyline, familiarity, number of elements, and vocabulary. For instance, familiarity of the story in the narrative task could add a further degree of complexity to the task. The pilot study showed that differences between the stories of the two picture prompts resulted in differences in some speech features such as mid-clause silent pauses and phases of repair (see Section 3.10.1). As the goal of language pedagogy is to foster L2 acquisition, teachers, test makers, and task designers need to take note of the empirically grounded argument in L2 monitoring research to enhance language learning, teaching, and assessment practices.

7.6 Limitations and suggestions for future research

The present study has a number of limitations and suggestions that need to be considered in future research. As the current study did not examine Covert-repair, the results of disfluency could not accurately be interpreted. Future research needs to obtain in-depth investigation of Covert-repair. To do this, retrospective interviews need to be employed to explore the conscious decisions taken by L2 learners when they produce disfluency, especially filled pauses. This could highlight whether disfluency is produced to monitor L2 speech or as a result of other processing difficulties. Examining Covert-repair in different proficiency levels could provide important information about the development of L2 self-monitoring behaviour. Furthermore, the present study only examined two levels of proficiency (elementary and intermediate), and the results indicated that L2 self-monitoring considerably changes with the development of proficiency level. Future research needs to examine a greater range of proficiency levels as well as L1 baseline data, as this will enrich our knowledge of L2 self-monitoring behaviour. Future studies also need to examine individual traits and their effect on self-monitoring behaviour such as anxiety, self-perception, and introversion/extroversion.

As self-monitoring is often examined in the monologic mode, it would be useful for future studies to look at self-monitoring in dialogue. In Levelt's (1983) account of self-repair, he explained that once speakers detect an error in their speech, they interrupt utterances and produce hesitation to signal to the listener that there is a problem in their speech. This suggests that producing hesitation can be motivated by the presence of the addressee. However, it is still not clear how self-monitoring can be affected by an interactive interlocutor. It has been assumed that self-monitoring in dialogue is more complicated than in monologue because interlocutors could interrupt others' speech and correct their errors (Hartsuiker, 2014). This means that dialogic performance/speech could enhance self-monitoring because the presence of the interlocutor may reduce the cognitive demand during processing. In fact, some researchers found that dialogic talk results in fewer aspects of disfluency compared to monologue (Michel, 2011; Tavakoli, 2016; Witton-Davies, 2014), which could be indicative of the fact that dialogue may enhance self-monitoring. Furthermore, it has been assumed that having an interlocutor from different linguistic or social backgrounds could have an impact on self-monitoring (Hartsuiker, 2014). That is, speakers may adjust their utterances to be suitable for the addressees. There is still a lack of research in this area, perhaps due in part to the difficulty of measuring dialogue (Tavakoli, 2016). It would be interesting for future studies to investigate the effects of dialogic versus monologic modes on different aspects of self-monitoring (disfluency, repair types, temporal phases of repair, and accuracy).

As working memory is mainly examined in correlational research (Juffs & Harrington, 2011), there is a need for caution in interpreting correlational analyses. For example, researchers need to be aware about the sensitivity of correlation to sample sizes. The likelihood of obtaining a significant correlation is higher as the sample size increases (Juffs & Harrington, 2011). The data obtained from the correlation analyses may differ according to age group, task condition, or proficiency level as discussed in Section 6.5. It would be useful for L2 research to examine the interaction effects between WMC and different factors (e.g., proficiency level, task type, and L1), to uncover how WMC is implicated in relation to these different variables. Interactional research is useful in this regard because it can show not only the combined impact of individual differences and contextual factors on L2 performance, but also the potential effects on the L2 learning processes (DeKeyser, 2012). That is, interaction effects are likely to reveal whether the contextual factors such as task type and task condition can reduce the cognitive load on WMC during L2 processing. This has important implications because such data can highlight whether or not various task types can create an unnecessarily high working memory load and thereby restrict the potential of L2 learning through self-monitoring.

Finally, future research needs to increase the demand of dual task condition by reducing the time of stimulus spent on the secondary task. The current study had a stimulus appear every 5 seconds; future research needs to reduce the timing to 3, 2, or 1 second. Future research also needs to consider setting a stricter rule for reaction times by considering responses accurate only if they occur within a very short time, e.g., 1.00 ms of stimulus presentation, since this practice is being followed in some psychological research (e.g., Eichorn, Marton, Schwartz, Melara, & Pirutinsky, 2016). Future research needs also to implement some of the methods developed in the TBLT field for assessing task complexity (e.g., Révész, 2014; Révész et al., 2016; Révész, Sachs, & Hama, 2014). For example, it would be useful to assess the validity of cognitive task demand by examining the accuracy and reaction time of a secondary task on a control group, and then to compare the results of the reaction time and accuracy in the secondary tasks in the control and main groups. Evidence for the validity of task complexity can also be assessed using other measures such as self-rating and expert judgements (Révész, 2014; Révész et al., 2014). These techniques can give an accurate impression about the extent to which task demand is operationalised.

7.7 Final remarks

As the current study has provided some insights into the functioning of the monitor in L2 speech production, it is hoped that it will contribute to further investigations into L2 self-monitoring behaviour. Examining how L2 self-monitoring works in relation to task conditions, proficiency levels, and WMC has important consequences for language processing, production, acquisition and development. It is hoped that examining some assumptions of the PLT on L2 learners will encourage future research in the L2 field to test the existing theories of self-monitoring, particularly the PLT. Much remains to be understood about L2 self-monitoring in relation to various individual traits and cognitive systems. Self-monitoring research could greatly benefit from theorising and methods in other fields, particularly cognitive psychology. Interdisciplinary research is certainly needed in self-monitoring investigation as it is a complex construct that cannot be accurately understood without combining perspectives from different fields. Hopefully, this study has provided a reference for future research on L2 self-monitoring through testing some assumptions of the PLT in the L2 context, besides introducing novel methods and practices in studying L2 self-monitoring behaviour.

List of references

Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. Bilingualism: Language and Cognition, 4(2), 179-190.

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), 165-175.

Ahmadian, M. J., Abdolrezapour, P., & Ketabi, S. (2012). Task difficulty and self-repair behavior in second language oral production. International Journal of Applied Linguistics, 22(3), 310-330.

Ahmadian, M. J., & Tavakoli, M. (2014). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. Language Teaching Research, 15(1), 35-59.

Albarqi, G. (2016). The effects of manipulating cognitive task complexity along +/- immediacy of information on hesitation aspect of L2 fluency. (Unpublished dissertation). University of Reading, Reading, UK.

Albarqi, G. (2018). What can dual task paradigm tell us about second language self-monitoring behaviour? Language Studies Working Papers, 9, 3-13.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological review, 89(4), 369-406.

Awwad, 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 doctoral thesis). University of Reading, Reading, UK.

Babaii, E., & Ansari, H. (2001). The C-test: a valid operationalization of reduced redundancy principle? System, 29(2), 209-219.

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford: Oxford University Press.

Baddeley, A. (1986). Working memory. Oxford: Oxford University Press.

Baddeley, A. (1996). Exploring the central executive. The Quarterly Journal of Experimental Psychology Section A, 49(1), 5-28.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in cognitive sciences, 4(11), 417-423.

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

Baddeley, A. (2012). Working memory: Theories, models, and controversies. Annual review of psychology, 63, 1-29.

Baddeley, A. (2015). Working memory in second language learning. In Z. Wen, M. Mota, & A. McNeill (Eds.), Working memory in second language acquisition and processing (pp. 17-28). Bristol: Multilingual Matters.

Baddeley, A., & Hitch, G. (1974). Working memory. Psychology of learning and motivation, 8, 47-89.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control (pp. 28-61). New York: Cambridge University Press.

Blackmer, E. R., & Mitton, J. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. Cognition, 39(3), 173-194.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. Research methodology in second-language acquisition, 245-261.

Boersma, P., & Weenink, D. (2008). Praat, a system for doing phonetics by computer. Computer software, downloaded from htt://www.fon.hum.uva.nl/praat/.

Boland, H. T., Hartsuiker, R. J., Pickering, M. J., & Postma, A. (2005). Repairing inappropriately specified utterances: Revision or restart? Psychonomic bulletin & review, 12(3), 472-477.

Brédart, S. (1991). Word interruption in self-repairing. Journal of psycholinguistic research, 20(2), 123-138.

Broos, W. P., Duyck, W., & Hartsuiker, R. J. (2018). Monitoring speech production and comprehension: Where is the second-language delay? Quarterly journal of experimental psychology, 1-19.

Canale, M., & Swain, M. (1980). Theoretical bases of com-municative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1-47.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. Behavioral and brain Sciences, 22(1), 77-94.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. Psychological review, 92(3), 414-423.

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. System, 72, 85-98.

Cobb, T. (2017). Web Vocabprofile. An adaptation of Heatley, Nation & Coxhead's (2002) Range, http://www.lextutor.ca/vp/

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New Jersy: Lawrence Erlbaum Associates.

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological bulletin, 104(2), 163.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 62–101). Cambridge, England: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and brain sciences, 24(1), 87-114.

Cowan, N. (2015). Second-language use, theories of working memory, and the Vennian mind. In Z. Wen, M. Mota, & A. McNeill (Eds.), Working memory in second language acquisition and processing (pp. 49-60). Bristol: Multilingual Matters.

Council of Europe. (2001). Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge, UK: Cambridge University Press.

Daller, M., & Yixin, W. (2017). Predicting study success of international students. Applied Linguistics Review, 8(4), 355-374.

De Bot, K. (2003). A bilingual production model: Levelt’s ‘speaking’model adapted. In The bilingualism reader (pp. 399-420).

De Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. Language Teaching Research, 20(3), 387-404.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS), 17– 20. Stockholm, Sweden: Royal Institute of Technology.

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

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

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

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. Bilingualism: Language and Cognition, 15(4), 782-796.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), Cognition and second language instruction (pp. 125-151). Cambridge: Cambridge University Press.

DeKeyser, R. (2007). Practice in a second language: Perspectives from applied linguistics and cognitive psychology. Cambridge: Cambridge University Press.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. Foreign Language Annals, 43(1), 80-92.

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. Language learning, 62, 189-200.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. Studies in Second Language Acquisition, 19(2), 195-221.

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. Language learning, 63, 52-67.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological review, 93(3), 283.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. Studies in Second Language Acquisition, 31(4), 533-557.

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

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. Language Testing, 9(2), 187-206.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. Studies in Second Language Acquisition, 20, 349–385.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. Psychological review, 87(3), 272.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. Bilingualism: Language and Cognition, 12(3), 273-289.

Dunn, A. L., & Tree, J. E. F. (2009). A quick, gradient bilingual dominance scale. Bilingualism: Language and Cognition, 12(3), 273-289.

Eichorn, N., Marton, K., Schwartz, R. G., Melara, R. D., & Pirutinsky, S. (2016). Does working memory enhance or interfere with speech fluency in adults who do and do not stutter? Evidence from a dual-task paradigm. Journal of Speech, Language, and Hearing Research, 59(3), 415-429.

Ellis, R. (2005a). Planning and task performance in a second language (Vol. 11). Amsterdam: John Benjamins Publishing.

Ellis, R. (2005b). Principles of instructed language learning. System, 33(2), 209-224.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. Applied Linguistics, 27(3), 464-491.

Feldmann, U., & Stemmer, B. (1987). Thin_ aloud a_ retrospective da_ in C-te_ taking: diffe_ languages-diff_ learners-sa_ approaches. Introspection in second language research, 251-267.

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

Finardi, K., & Weissheimer, J. (2009). On the Relationship between working memory capacity and L2 speech development. Signótica, 20(2), 367-391.

Fortkamp, M. B. M. (2000). Working memory capacity and L2 speech production: An exploration study. (Unpublished doctoral thesis), Universidade Federal de Santa Catarina.

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

Gaillard, S., & Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: The elicited imitation task. Language learning, 66(2), 419-447.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. Trends in cognitive sciences, 3(11), 410-419.

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. Developmental Psychology, 29(4), 770.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? European Journal of Psychology of Education, 8(3), 259.

Georgiadou, E., & 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), 877-895.

Georgiadou, E. S. (2016). The role of proficiency, speaking habits and error-tolerance in the self-repair behaviour of Emirati EFL learners. Asian EFL Journal, 18(4), 102-126.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. International Review of Applied Linguistics in Language Teaching, 45(3), 215-240.

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

Golonka, E. M. (2006). Predictors revisited: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. Modern Language Journal, 90, 496-505.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. Studies in Second Language Acquisition, 34(3), 445-474.

Grabowski, J. (2010). Speaking, writing, and memory span in children: Output modality affects cognitive performance. International Journal of Psychology, 45(1), 28-39.

Guará-Tavares, M. G. (2008). Pre-task planning, working memory capacity, and L2 speech performance. (Unpublished doctoral thesis). Universidade Federal de Santa Catarina, Brazil.

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

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. Cognitive psychology, 44(4), 339-387.

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R. J. Harris (Ed.), Cognitive processing in bilinguals (pp. 123-135). Amsterdam: North Holland.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. Studies in Second Language Acquisition, 14(1), 25-38.

Hartsuiker, R. (2014). Monitoring and control of the production system. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), The Oxford Handbook of Language Production. (pp. 417-436). Oxford: Oxford University Press.

Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. Cognition, 108(3), 601-607.

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. Cognitive psychology, 42(2), 113-157.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. Available at http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Hood, L., & Lightbown, P. (1978). What children do when asked to “say what I say”: Does elicited imitation measure linguistic knowledge? Allied Health and Behavioral Sciences, 1, 195–220.

Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen & M. Pierrard (Eds.), Investigations in instructed second language acquisition (pp. 1-27). New York: Mouton de Gruyter.

Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. Bilingualism: Language Cognition, 15(2), 422-433.

Hulstijn, J. H. (2015). Language proficiency in native and non-native speakers: Theory and research (Vol. 41). Amsterdam: John Benjamins.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? Applied Psycholinguistics, 30(4), 555-582.

IBM SPSS Statistics for Mac (Version 24). (2012). Armonk, NY: IBM Corp.

Ishikawa, T. (2007). The effect of manipulating task complexity along the (±Here-and-Now) dimension on L2 written discourse. In M. P. García Mayo (Ed.), Investigating tasks in formal language learning (pp. 136–156). Toronto: Multilingual Matters.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. Applied Linguistics, 24(2), 168-196.

Jou, J., & Harris, R. J. (1992). The effect of divided attention on speech production. Bulletin of the Psychonomic Society, 30(4), 301-304.

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

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

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

Kane, M. J., & Engle, R. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. Journal of experimental psychology: General, 132(1), 47.

Kircher, T. T., Brammer, M. J., Levelt, W., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: engagement of left temporal cortex during pauses in speech. NeuroImage, 21(1), 84-90.

Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. Language Testing, 14(1), 47-84.

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-Test. Language Testing, 1(2), 134-146.

Kormos, J. (1998). A new psycholinguistic taxonomy of self-repairs in L2: A qualitative analysis with retrospection. Even Yearbook, ELTE SEAS Working Papers in Linguistics, 3, 43-68.

Kormos, J. (1999). Monitoring and self-repair in L2. Language learning, 49(2), 303-342.

Kormos, J. (2000a). The role of attention in monitoring second language speech production. Language learning, 50(2), 343-384.

Kormos, J. (2000b). The timing of self-repairs in second language speech production. Studies in Second Language Acquisition, 22(2), 145-167.

Kormos, J. (2006). Speech production and second language acquisition. New York: Routledge.

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

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. System, 32(2), 145-164.

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

Kovač, M., & Milatović, M. (2013). Analysis of repair distribution, error correction rates and repair successfulness in L2. Studia linguistica, 67(2), 225-255.

Krashen, S. D. (1981). Second language acquisition and second language learning. Oxford: Oxford University Press.

Lambert, C., & Kormos, J. (2014). Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition. Applied Linguistics, 35(5).

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

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. Applied Linguistics, 16(3), 307-322.

Laver, J. D. (1973). The detection and correction of slips of the tongue. Speech errors as linguistic evidence, 132-143.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistical control of speech production. In V. A. Fromkin (Ed.), Errors in linguistic performance: Slips of the tongue, ear, pen, and hand (pp. 287-305). New York: Academic Press.

Lennon, P. (1984). Retelling a story in English as a second language. In H. W. D. D. Mohle & M. Raupach (Eds.), Second language productions (pp. 50-68). Tu˙bingen, Germany: Gunter Narr Verlag.

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

Levelt, W. (1983). Monitoring and self-repair in speech. Cognition, 14(1), 41-104.

Levelt, W. J. (1989). Speaking: From intention to articulation (Vol. 1). Cambridge: MIT press.

Levelt, W. J. (1992). The perceptual loop theory not disconfirmed: A reply to MacKay. Consciousness and Cognition, 1(3), 226-230.

Levelt, W. J. (1999). Language production: A blueprint of the speaker. Neurocognition of language, 83-122.

Levelt, W. J., & Cutler, A. (1983). Prosodic marking in speech repair. Journal of semantics, 2(2), 205-218.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral brain sciences, 22(1), 1-38.

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. Applied Psycholinguistics, 40(3), 693-722.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological review, 95(4), 492-527.

Mackay, D. C. (1992). Awareness and error detection: New theories and research paradigms Consciousness and Cognition, 1, 199-225.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. Psychological review, 89(5), 483-506.

McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: A question of validity. Journal of Speech and Hearing disorders, 47(1), 19-24.

McNamara, T. F. (1997). ‘Interaction’in second language performance assessment: Whose performance? Applied Linguistics, 18(4), 446-466.

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners’ written performances. Journal of Second Language Writing, 45, 31-45.

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (pp. 141-174). Amsterdam: John Benjamins.

Mitchell, A. E., Jarvis, S., O’Malley, M., & Konstantinova, I. (2015). Working Memory Measures and L2 Proficiency. In Z. Wen, M. Mota, & A. McNeill (Eds.), Working memory in second language acquisition and processing (Vol. 87, pp. 270-283). Bristol: Multilingual Matters.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. R. Bourne (Eds.), Foreign language learning: Psycholinguistic studies on training and retention (pp. 339-364). Mahwah, NJ: Erlbaum.

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

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), Errors in linguistic performance: slips of the tongue, ear, pen, and hand. New York: Academic Press.

O’Connor, N. (1988). Repairs as indicative of interlanguage variation and change. In T. J. Walsh (Ed.), Synchronic and diachronic approaches to linguistic variation and change (pp. 251-259). Washington, D.C.: Georgetown University Press.

Oller, J. W., Jr. (1979). Language tests at school. London: Longman.

Oller, J. W., Jr. , & Conrad, C. (1971). The cloze technique and ESL proficiency. Language learning, 21, 183-196.

Oomen, C., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. Journal of psycholinguistic research, 30(2), 163-184.

Oomen, C., & Postma, A. (2002). Limitations in processing resources and speech monitoring. Language Cognitive Processes, 17(2), 163-184.

Ortega, L., Iwashita, N., Norris, J., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. Paper presented at the Second Language Research Forum, Toronto.

Pallant, J. (2016). SPSS survival manual. Berkshire: McGraw-Hill Education.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. The Quarterly Journal of Experimental Psychology Section A, 44(1), 47-67.

Pashler, H. (1998). The Psychology of Attention. Cambridge, MA: MIT Press.

Pellicer-Sánchez, A. (2015). Developing automaticity and speed of lexical access: the effects of incidental and explicit teaching approaches. Journal of Spanish Language Teaching, 2(2), 126-139.

Phakiti, A., & Paltridge, B. (2015). Approaches and methods in applied linguistics research. In A. Phakiti & B. Paltridge (Eds.), Research methods in applied linguistics: A practical resource (pp. 5-25). London: Bloomsbury.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. Studies in Second Language Acquisition, 11(1), 63-90.

Pillai, S. (2006). Error-detection, self-monitoring and self-repair in speech production. Paper presented at the Proceedings of the 9th Australian International Conference on Speech Science & Technology, Melbourne.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. Cognition, 77(2), 97-132.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech, Language, and Hearing Research, 36(3), 472-487.

Postma, A., Kolk, H., & Povel, D.-J. (1990). On the relation among speech errors, disfluencies, and self-repairs. Language and Speech, 33(1), 19-29.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. Language and Speech, 39(4), 375-392.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. Applied Linguistics, 15(1), 36-57.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. Applied Linguistics, 35, 87–92.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. Studies in Second Language Acquisition, 38(4), 703-737.

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. Language learning, 64, 615–650.

Richards, B., Malvern, D. D., Meara, P., Milton, J., & Treffers-Daller, J. (2009). Vocabulary studies in first and second language acquisition: The interface between theory and application. Basingstoke, Hampshire: Palgrave Macmillan.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. Studies in Second Language Acquisition, 19(2), 223-247.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. IRAL-International Review of Applied Linguistics in Language Teaching, 45(3), 193-213.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (pp. 3-38). Amsterdam: John Benjamins.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. IRAL-International Review of Applied Linguistics in Language Teaching, 49(4), 295-319.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic/Control Processing and Attention. In R. Parasuraaman & D. R. Davies (Eds.), Varieties of attention (pp. 1-27). New York: Academic Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 3.0). Computer Software and Manual. Pittsburgh, PA: Psychology Software Tools Inc.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty (Ed.), The handbook of second language acquisition (pp. 383-408). Malden, MA: Blackwell.

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

Segalowitz, N. (2012). In P. J. Robinson (Ed.), The Routledge encyclopedia of second language acquisition. London: Routledge.

Segalowitz, N., & Hulstijn, J. H. (2005). Automaticity in bilingualism and second language learning In J. F. Kroll & A. M. B. D. Groot (Eds.), Handbook of bilingualism: Psycholinguistic approaches (pp. 371-388). Oxford: Oxford University Press.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. Applied Psycholinguistics, 14(3), 369-385.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. Cognition, 108(3), 837-842.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological review, 84(127), 127-190.

Sigott, G. (2004). Towards identifying the C-Test construct. Frankfurt am Main: Peter Lang.

Simard, D., Fortier, V., & Zuniga, M. (2011). Attention et production d'autoreformulations autoamorcées en français langue seconde, quelle relation? Journal of French Language Studies, 21(3), 417-436.

Simard, D., French, L., & Zuniga, M. (2017). Evolution of Self-Repair Behaviour in Narration Among Adult Learners of French as a Second Language. Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée, 20(2), 71-89.

Simpson, R., Eisenchlas, S., & Haugh, M. (2013). The functions of self-initiated self-repair in the second language Chinese classroom. International Journal of Applied Linguistics, 23(2), 144-165.

Skehan, P. (2003). Task-based instruction. Language teaching, 36(1), 1-14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. Applied Linguistics, 30(4), 510-532.

Skehan, P. (2015). Working memory and second language performance: A commentary. In Z. Wang, M. Mailce, & M. Arthur (Eds.), Working memory in second language acquisition and processing (Vol. 87, pp. 189-201). Bristol: Multilingual Matters.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), Planning and task performance in a second language (Vol. 11, pp. 193-216). Amsterdam: John Benjamins.

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

Skehan, P., Xiaoyue, B., Qian, L., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. Language Teaching Research, 16(2), 170-187.

Stebbins, R. A. (2001). Exploratory research in the social sciences. London: Sage.

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. Language learning, 65(4), 860-895.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), Input in second language acquisition (Vol. 15, pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning In C. Cook & B. Seidlhofer (Eds.), Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson (pp. 125-144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. Applied Linguistics, 16(3), 371-391.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics. Boston: Pearson.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. ELT journal, 65(1), 71-79.

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), 133-150.

Tavakoli, P. (2019). Automaticity, fluency and second language task performance. In Z. Wen & M. J. Ahmadian. (Eds.), Researching L2 task performance and pedagogy: In honour of Peter Skehan. Amsterdam: John Benjamins.

Tavakoli, P., Nakatsuhara, F., & Hunter, A.-M. (2017). Scoring validity of the Aptis Speaking Test: Investigating fluency across tasks and levels of proficiency. London: British Council.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), Planning and task performance in a second language (Vol. 11, pp. 239-273). Amsterdam: John Benjamins.

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. Language learning, 44(2), 307-336.

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

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research:“Clozing” the gap. Studies in Second Language Acquisition, 33(3), 339-372.

Van Hest, E. (1996). Self-repair in L1 and L2 production. Tilburg: Tilburg University Press.

VanPatten, B. (1996). Input processing and grammar instruction: Theory and research. Norwood, NJ: Albex.

VanPatten, B. (2002). Processing instruction: An update. Language learning, 52, 755-803.

Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective. In P. Skehan (Ed.), Processing perspectives on task performance (Vol. 5, pp. 155-185). Amsterdam: John Benjamins.

Wickens, C. D. (2007). Attention to the second language. IRAL-International Review of Applied Linguistics in Language Teaching, 45(3), 177-191.

Williams, J. N. (2012). Working memory and SLA. In M. Gass & A. Mackey (Eds.), The Routledge handbook of second language acquisition (pp. 427–441). New York: Routledge.

Williams, S. A., & Korko, M. (2019). Pause behavior within reformulations and the proficiency level of second language learners of English. Applied Psycholinguistics, 40, 723-742.

Willis, J. (1996). A Framework for Task-Based Learning. Harlow: Longman Addison-Wesley.

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

Wright, C. (2010). Role of Working Memory in SLA. Saarbrucken: VDM Publishing House.

Wright, C. (2015). Working Memory and L2 Development Across the Lifespan: A Commentary. In Z. Wen, M. Mota, & A. McNeill (Eds.), Working memory in second language acquisition and processing (Vol. 87, pp. 285-298). Bristol: Multilingual Matters.

Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. Foreign Language Annals, 46(4), 680-704.

Yixin, W., & Daller, M. (2014). Predicting Chinese Students' academic achievement in the UK. Learning, Working Communicating in a Global Context, 217.

Zuniga, M. (2015). Attention and second language speech production: The relationship between attention and the production of self-initiated self-repairs. (Doctoral thesis). University of Auebec, Montreal, Canada.

Zuniga, M., & Simard, D. (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. Journal of psycholinguistic research, 48(1), 43-59.

School of Literature and Languages
Department of English Language and Applied Linguistics

Appendix 1: Consent form

Project title: **Second Language Fluency Development**

I understand the purpose of this research and understand what is required of me.

I have read and understood the Information Sheet relating to this project, which has been explained to me by Ghadah Albarqi. I agree to the arrangements described in the Information Sheet so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 2: Information sheet

School of Literature and Languages
Department of English Language and Applied Linguistics

Department of English Language and Applied Linguistics
EM Building
The University of Reading
Whiteknights, PO Box 219
Reading RG6 6AW

Researcher:
Ghadah Ahmad Albarqi
Email: g.a.i.albarqi@pgr.reading.ac.uk

INFORMATION SHEET

The purpose of this study is to explore second language speech production under varying degrees of task complexity and task conditions and the relationship between speech and individual differences.

The participants have been selected because their second language is English. They will be asked to perform some speaking tasks, proficiency tests and test of working memory. Their performances are recorded as they complete these tasks and tests. They may also be asked to take part in an interview about the tasks that they completed.

The collected data will be securely kept on a password-protected computer or in a locked drawer. Only the researcher and the supervisors will have access to the data. The data will be used for academic purposes only. The privacy and confidentiality will be carefully observed, and the participants have the right to withdraw from the study at any time they wish to.

This project has been subject to ethical review by the School Ethics and Research Committee and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's Notes for Guidance on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at p.tavakoli@reading.ac.uk

Ghadah Albarqi [redacted]

Appendix 3: VocabProfilers

A-The Ship

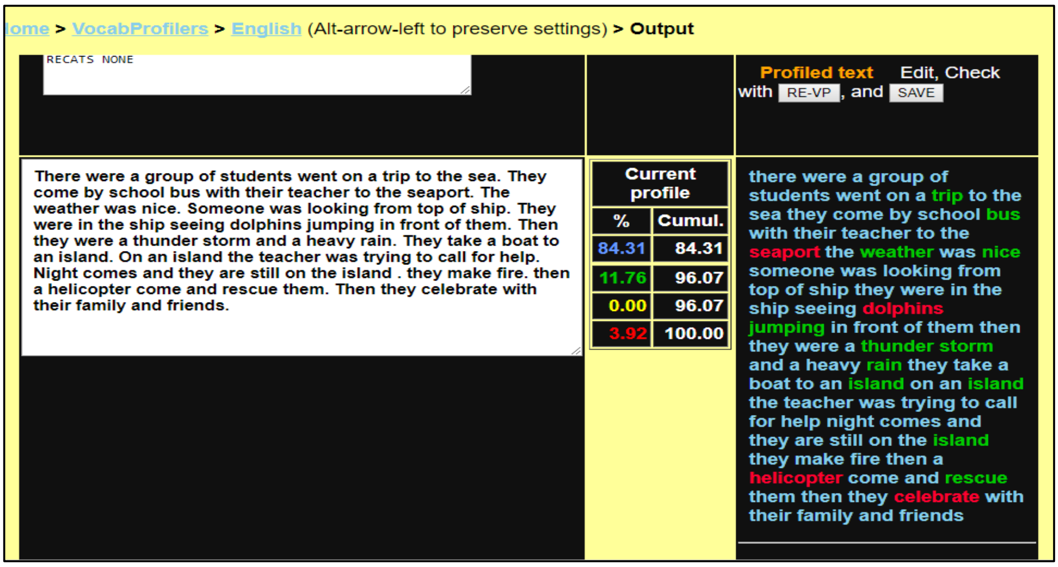


Figure 30. VovabProfilers (the Ship)

B- The Fire

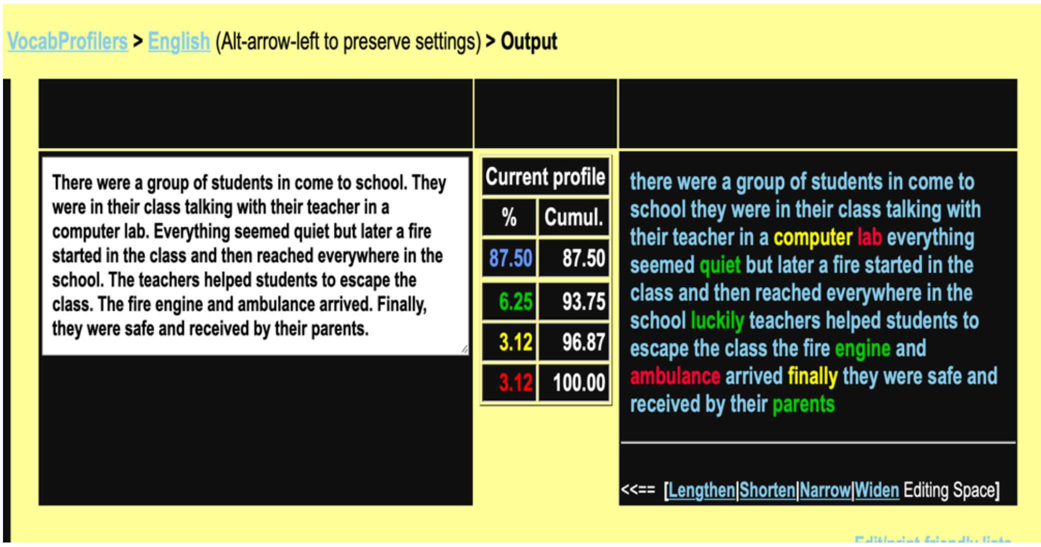


Figure 31. VocabProfilers (the Fire)

Appendix 4: Designing the picture prompts

A-The Fire

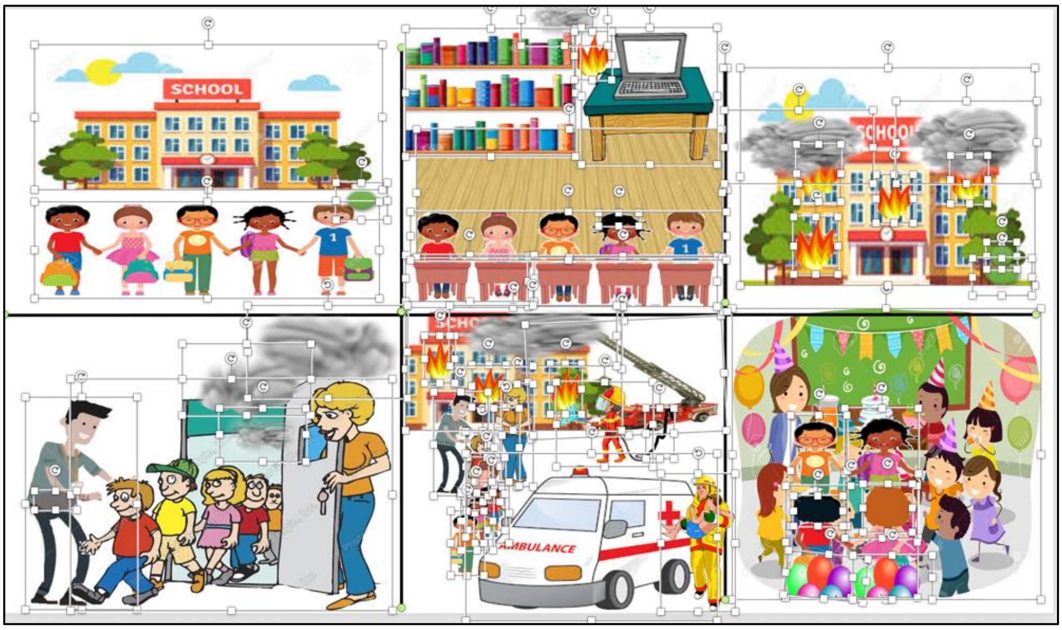


Figure 32. Designing the Fire

B-The Ship

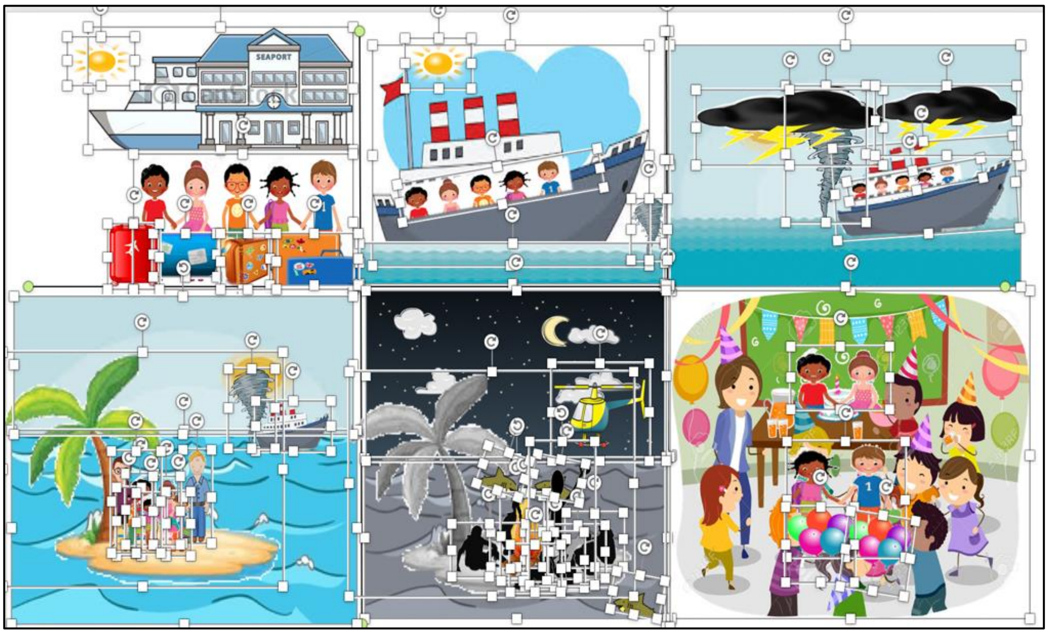


Figure 33. Designing the Ship

Appendix 5: Picture prompts in the pilot study

A-The Ship

<u>Instructions:</u>

-Describe the picture prompt to a friend who cannot see it

-Explain how students and their teacher happen to be on the island and how they were rescued

-You have one minute to plan

Figure 34. Picture prompt (the Ship) with instructions in Arabic

B-The Fire

<u>Instructions:</u>

-Describe the picture prompt to a friend who cannot see it

-Explain how the fire starts and how the students and their teacher were rescued

-You have one minute to plan

Figure 35. Picture Prompt (the Fire) with instructions in Arabic

Appendix 6: Designing slides for the secondary task

A-The Storm

Figure 36. Designing slides for the secondary task

B-The trial picture prompt

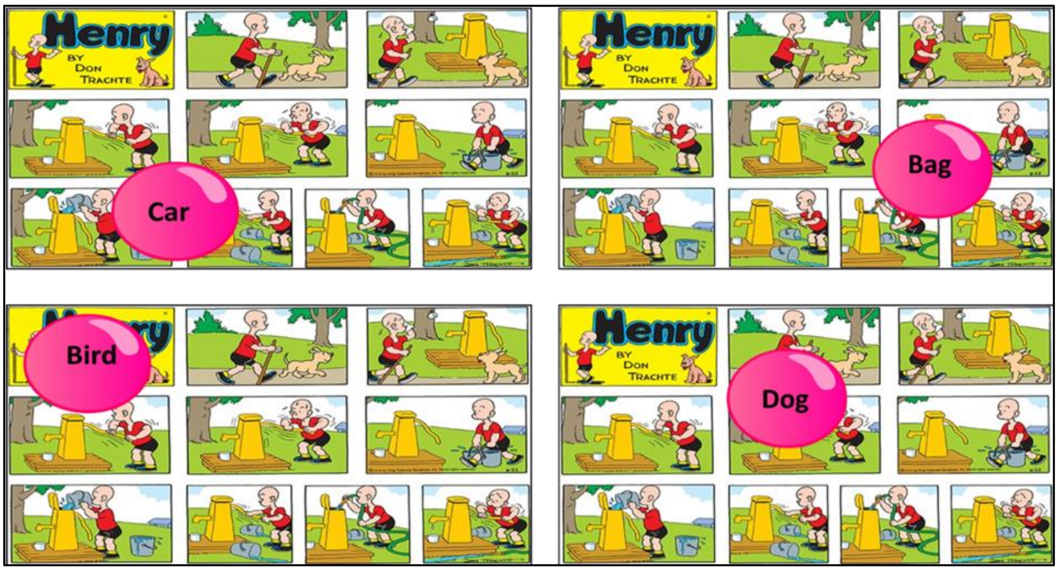


Figure 37. Designing slides for practice trials of the secondary task

Appendix 7: Survey Monkey: (Comparability of picture prompts)

A-Degree of story complexity

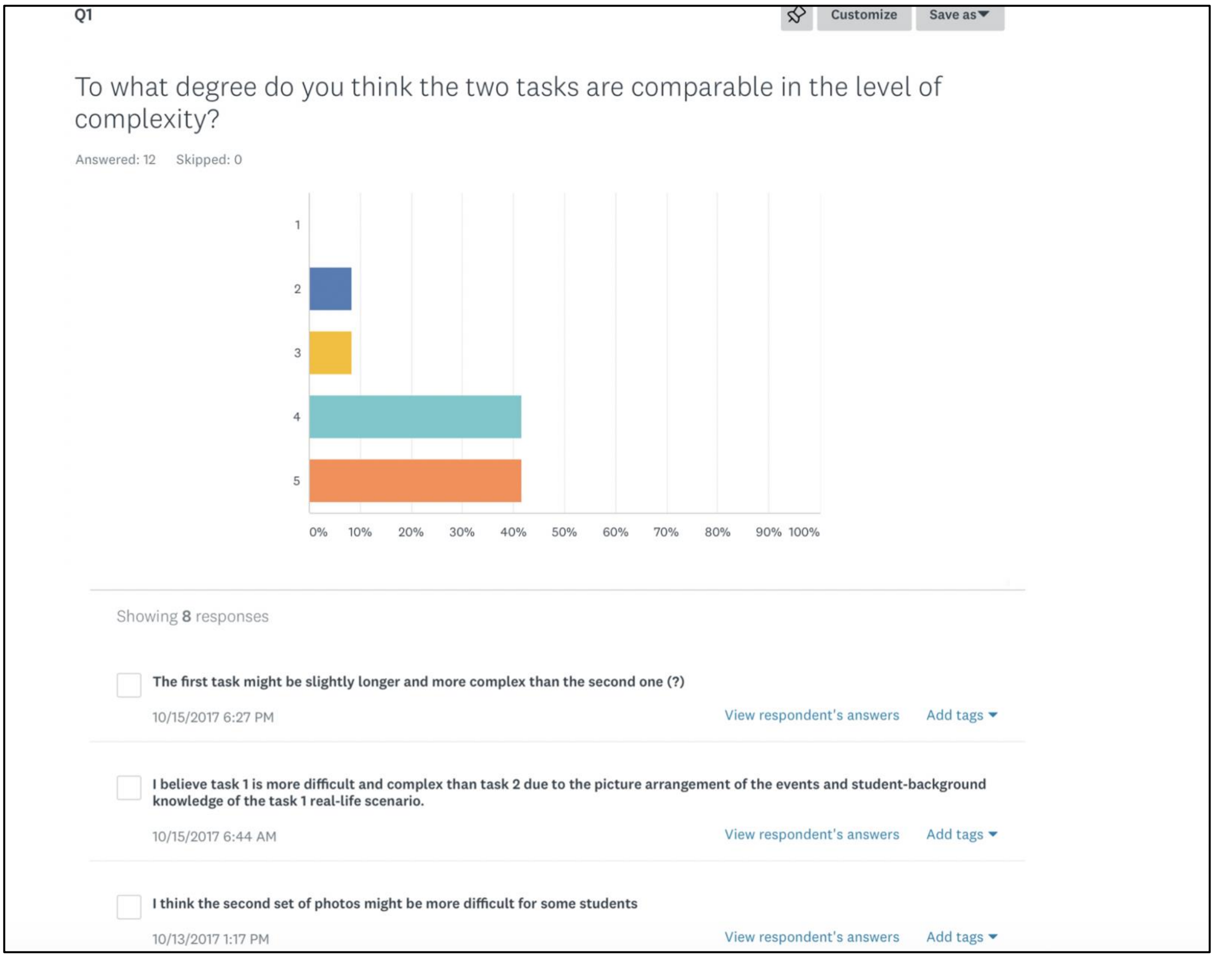


Figure 38. Survey Monkey: Degree of story complexity

B- Vocabulary comparability

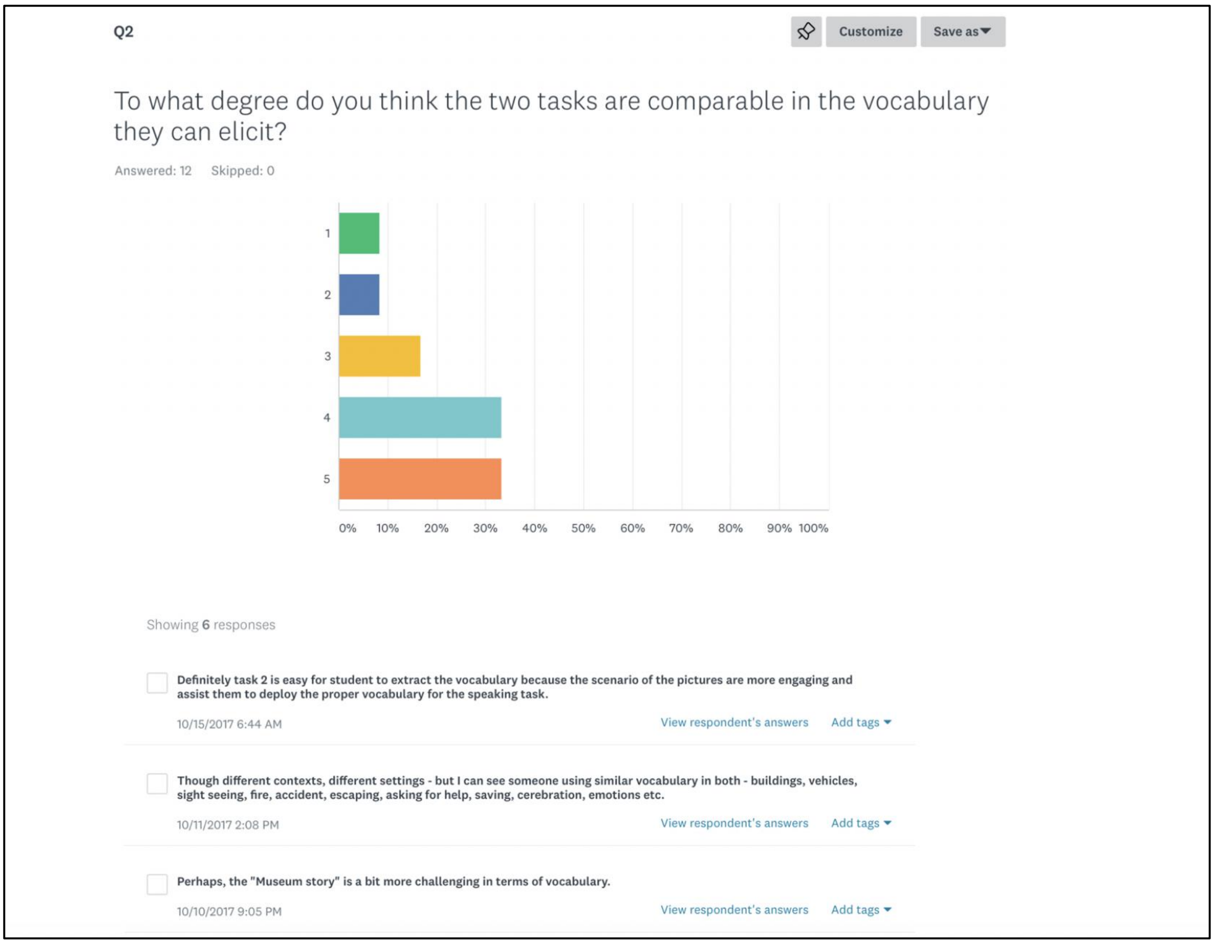


Figure 39. Survey Monkey: Vocabulary comparability

Appendix 8: Picture prompts in the main study

A-The Museum

Instructions:

- Describe pictures and narrate the story to a friend who cannot see what's happening
- Explain how characters feel in each picture and why they may feel in that way.
- Explain why people's feeling may change in each picture.

التعليمات

❖ قومي بوصف الصور وحكاية أحداث القصة لصديقة لاتستطيع مشاهدة مايحدث.
❖ اشرحي كيف يشعر الأشخاص في كل صورة ولماذا قد يشعروا بهذه الطريقة.
❖ قدمي تفسيرا لتغير مشاعر الشخصيات في كل صورة.

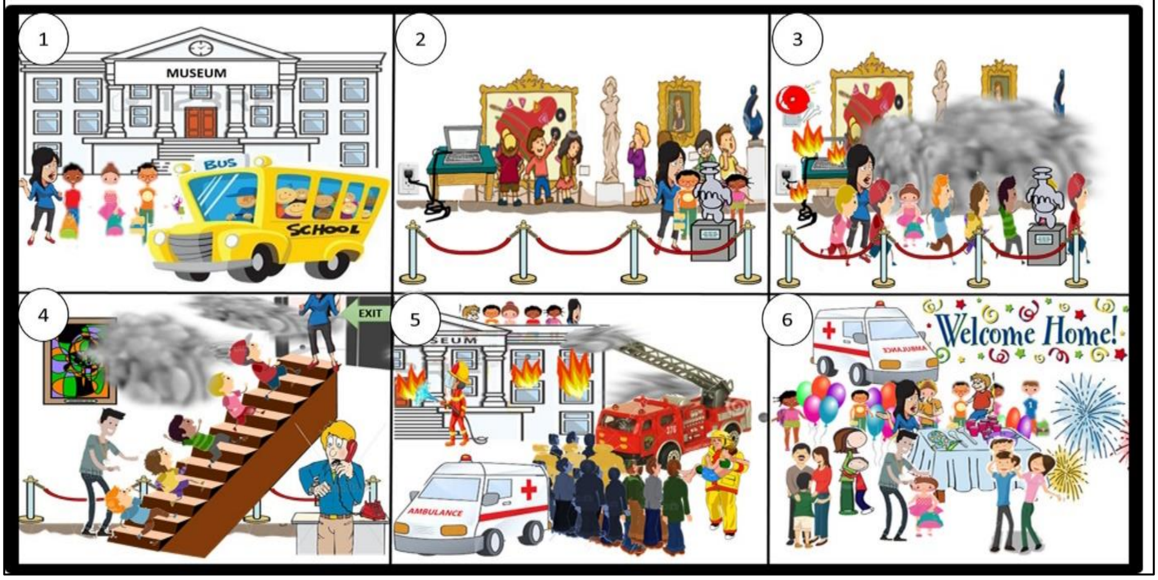


Figure 40. Picture Prompt (The Storm)

B-The Strom

Instructions:

- Describe pictures and narrate the story to a friend who cannot see what's happening.
- Explain how characters feel in each picture and why they may feel in that way.
- Explain why people's feeling may change in each picture.

التعليمات

- ❖ قومي بوصف الصور وحكاية أحداث القصة لصديقة لاتستطيع مشاهدة مايحدث.
- ❖ اشرحي كيف يشعر الأشخاص في كل صورة ولماذا قد يشعروا بهذه الطريقة.
- ❖ قدمي تفسيرا لتغير مشاعر الشخصيات في كل صورة.

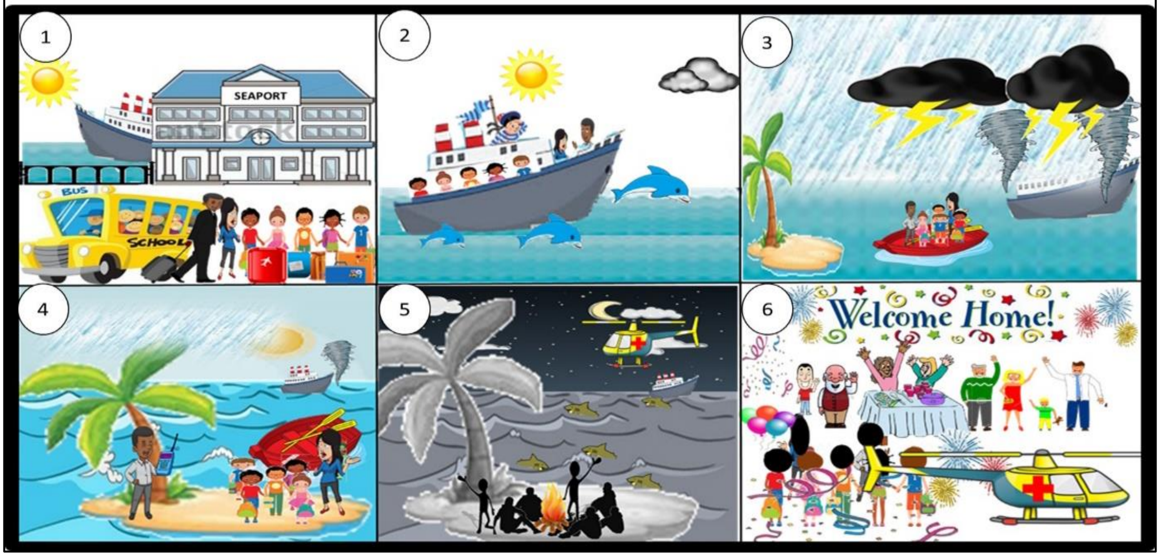


Figure 41. Picture Prompt (The Museum)

Appendix 9: E-Prime structure

A –Experiment explorer (E-Prime)

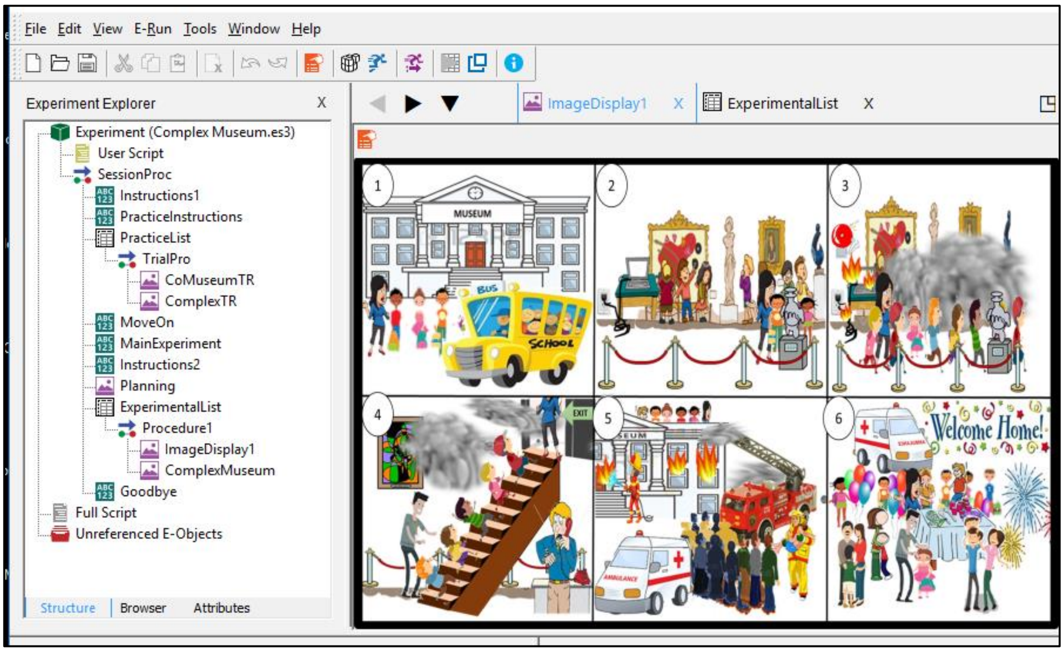


Figure 42. Experiment explorer (E-prime)

B – Experimental list (E-Prime)

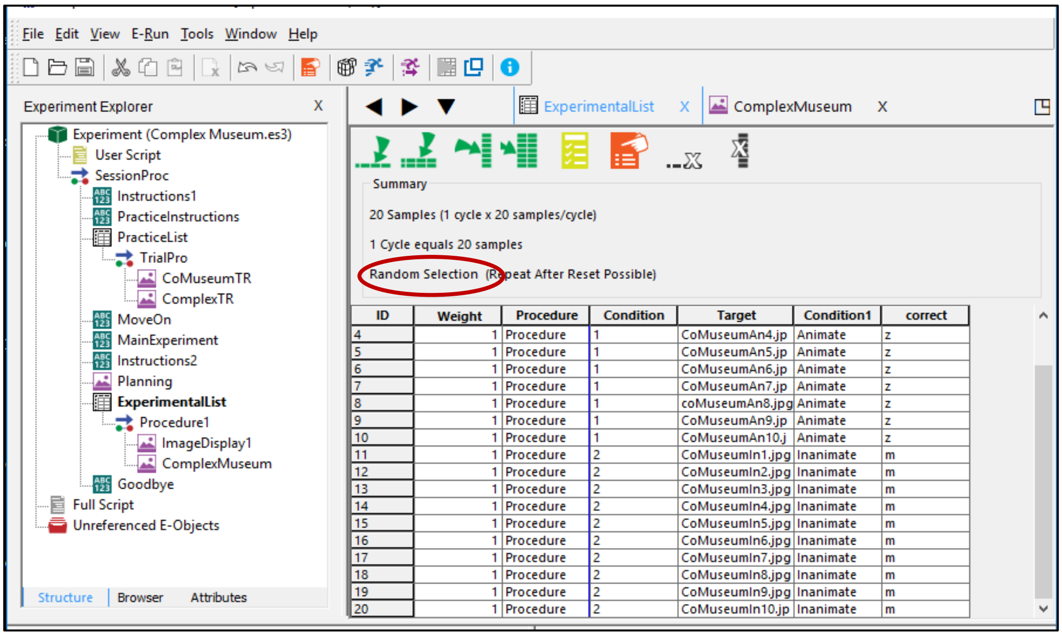


Figure 43. Experimental list (E-Prime)

C- Word lists in the secondary task

Table 42. Word lists in the secondary task

No.	Animate	Inanimate	Order of presentation
1	Dog	Paper	Random
2	Bee	Pencil	Random
3	Horse	Glass	Random
4	Fish	Car	Random
5	Camel	House	Random
6	Tiger	Door	Random
7	Sheep	Key	Random
8	Rabbit	Book	Random
9	Bird	Bag	Random
10	Monkey	Chair	Random

Appendix 10: C-test

English Language Ability Test (C-test)

*Name: ____________________ *Student number: __________________

*Age: ____________________

Task 1: Please fill in the missing letters. You have 25 minutes in total. Roughly the second half of the word is deleted. Be aware of the inflection such as third person singular, plural, past tense, etc. Only words with correct spelling can score. There are 100 points available with the whole test, with 20 points available from each passage. Thanks for your interest in this test and your contribution will be highly appreciated.

Here are some examples to help your understanding:

1. I li__ to go with you.
   Answer: like
2. I wo____ love to go with you.
   Answer: would
3. They are teen_____.
   Answer: teenagers

A. Learning to write

I was four when I started to learn to write. My grandfather started to teach me before I went to school. I remember th____ I always fo____ the cap______ letters mu___ easier t__ write th___ the sm___ letters. I reme____ that on___ we sta____ writing i___ school w__ were n__ allowed t__ use pe___, we h___ to u___ pencils un___ we bec___ really go__ at writing. I can write a few characters in Chinese now, but not very many.

B. Teenagers

It is clear that not all teenagers respond in the same way to peer group pressure. For exa_________, young peo_________ in t_______ early ye________ of seco__________ school a_______ more lik_________ to fe________ under pres__________ to we________ the sa________ clothes a_______ listen t_______ the sa________ music a_______ the re________ of t_______ peer gr________. By t________ time th________ reach middle or late adolescence, however, young people are more able to stand up against such influences.

C. Sleep room

One in three Japanese suffer from sleep disorders, which has prompted technologists to build a sound- insulated capsule – the Suimin (Sleep) Room. Users start by sit_______ up i_______ bed i_______ front o_______ a screen sho_______ a river win_______ through for_______. Soft mu________ plays i_______ the backg________, along wi______ the so_______ of wa_______ and bird________. After a f_________ relaxing min________, the lig_______ dim, t________ screen go________ blank, t___________ music fades and the bed reclines into a sleeping position.

D. Geography

The UK is located on a group of islands known as the British Isles, which lie between the Atlantic Ocean and the North Sea, northwest of France. At i_________ widest t_________ UK i_________ 300 mi_________ across a_________ 600 mi________ from No_________ to So_________. It sha_________ a sin_________ land bor_________ with the Irish Repu_________. Despite i_________ relatively sm_________ size t_________ UK boa_________ incredibly var_________ and of_________ very beau_________ scenery, fr_________ the mountains and valleys of the North and West to the rolling landscape of the South, and from downland and heath to Fens and marshland.

E. Record employment

Latest employment figures show that there are 28.2 million people in work. Work & Pensions Secretary said this showed the UK labour market has coped well so far with the current international economic uncertainty. He said: "Employment cont_________ to ri_________, with th_________ month's fig_________ showing a rec_________ 28.2 mil_________ people i_________ work. Th_________ are 65,000 mo________ people i_________ work th_________ last qua_________ and 252,000 mo_________ than la_________ year. Alth__________ both meas_________ of unempl_________ have ri_________ slightly, th_________ are st_________ significantly lower than they were a year ago."

C-test Keys

A. that, found, capital, much, to, than, small, remember, once, started, in, we, not, to, pens, had, use, until, became, good

B. example, people, the, years, secondary, are, likely, feel, pressure, wear, same, and, to, same, as, rest, the, group, the, they

C. sitting, in, in, of, showing, winding, forests, music, in, background, with, sound, water, birdsong, few, minutes, lights, the, goes, the.

D. its, the, is, miles, and, miles, north, south, shares, single, border, republic, its, small, the, boasts, varied, often, beautiful, from.

E. continues, rise, this, figures, record, million, in, there, more, in, than, quarter, more, last, although, measures, unemployment, risen, they, still.

Appendix 11: Elicited Imitation Test (EIT)

A-Sentences in the EIT adopted from Wu & Ortega (2013

Table 43. Sentences in the EIT

No.	Sentences	Syllables
1	I have to get a haircut	7
2	The red book is on the table	8
3	He takes a shower every morning	9
4	What did you say you were doing today?	10
5	After dinner I had a long, peaceful nap	11
6	It is possible that it will rain tomorrow	12
7	The little boy whose kitten died yesterday is sad	13
8	That restaurant is supposed to have very good food	13
9	I want a nice, big house in which my animals can live	14
10	You really enjoy listening to country music, don't you	14
11	Cross the street at the light and then just continue straight ahead	15
12	The person I'm meeting has a wonderful sense of humour	15
13	I hope it will get warmer sooner this year than it did last year	16
14	A good friend of mine always takes care of my neighbour’s three children	16

15	The terrible thief whom the police caught was very tall and thin	17
16	Would you be so kind as to hand me the book which is on the table?	17
17	I don't know if the 11:30 train has left the station yet	18
18	The exam wasn't nearly as difficult as you told me it would be	18
19	There are a lot of people who don’t eat anything at all in the morning	19

B - Scoring Criteria

4 = Perfect repetition

3 = Accurate content repetition

2 = Changes in content or in form that affect meaning

1 = Repetition of half of the stimulus or less

0 = Failure to repeat anything

Appendix 12: Working memory tests (Arabic & English)

A- Arabic working memory test

إختبار الذاكرة العاملة (العد العكسي)

هذا الإختبار مصـــمم لقياس مدى الذاكرة العاملة من ناحية التخزين والمعالجة لدى المشـــاركين في البحث. يطلب من المشاركين الإستماع إلى مجموعات من الأرقام وإعادتها بشكل عكسي. عدد الأرقام لكل مجموعة سوف تزداد في كل مرة وســـوف تقدم من خلال ملف صـــوتي وبفاصـــل زمني ثانية واحدة بين كل رقم. مدى الذاكرة العاملة يتحدد بناءً على اخر مجموعة من الأرقام تم إعادتها بنجاح مرتين.

تعلميات الإختبار

سوف تستمع لمجموعات مختلفة من الأرقام. سوف أذكر الأرقام وأنت تكررها بشكل عكسي. عدد الأرقام في كل مجموعة سوف يزداد بشكل مطرد. سوف نبدأ بمجموعة من ثلاث أرقام. عندما تحقق محاولتين ناجحتين ننتقل للمجموعة التالية (4 أرقام) وهكذا بشكل متزايد حتى 9 أرقام كحد أعلى. ينتهي الإختبار عندما لا تنجح مرتين في تكرار أي من المجموعات.

مثال:

عندما أقول: **5 4 3**

أنت تقول: **3 4 5**

أخبرني عندما تكون مستعداً.

√ / ×	**المحاولة الثالثة**	√ / ×	**المحاولة الثانية**	√ / ×	**المحاولة الأولى**	**المدى**
	715		936		524	**ثلاثة**
	9762		5146		7913	**أربعة**
	41539		64951		41527	**خمسة**
	269721		197249		639514	**ستة**
	4962413		9172631		3915372	**سبعة**
	62916473		53719231		72529416	**ثمانية**

	316497625		971492564		173956431	تسعة
						إسم الطالب
						مدى الذاكرة العاملة

B- English working memory test

Backward Digits Span Test (English Language)

This auditory task is designed to test learners' complex working memory capacity (storage and processing). Participants are required to listen to sets of increased digits and repeat them backwards. Numbers are recorded at one digit per second. Each learners' working memory span is determined based on the last digits set he/she has repeated successfully twice.

Instructions:

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

For example:

When I say: "**4 5 6**"

You say: "**6 5 4**"

Let me know when you are ready.

Span	First trial	√ / ×	Second trial	√ / ×	Third trial	√ / ×
Three	582		395		627	
Four	3915		4826		1973	

Five	68471		73169		25184	
Six	592834		469172		358261	
Seven	7452846		8361957		5913728	
Eight	92518753		16829374		81492573	
Nine	483261759		692748315		751936845	
Student name						
L2 Backward Digits Span result						

Appendix 13: Coding

A-The coding symbols

| = the end of AS-Unit **::** = the end of a clause

errfr = error-free clause
***** = error (lexical, grammatical, phonological)

^ clauses that have errors

(repair) type of repair

[ ] duration of temporal phases of repair

~ repetition

(?) disagreement

B- Sample of the coded data (1)

Student	**Speaking time**	**Prof. level**	**Task condition**	**Picture**	**Order**
P11 Rawan	**60**	**Elementary**	**Dual**	**Strom**	**1**

| In the photo* I can see uh the first picture 0.31 there *1.04 boy and girl ^ :: they~ uh [1.67] they uh 0.25 will go to uh 0.60 journey 0.67 with her[**0.93**] **0.28** with his*[**0.41**] uh 0.67[**0.85**] (lexical repair) teacher 2.68^| they[**0.40**] uh 0.30 uh[**0.85**] all of this at* the morning [**1.48**] (D-repair) | and they~ uh 3.44 they get uh 1.24 boat 0.26 in the morning ^| and uh 0.49 they was* uh 0.75 very happy^ :: and uh 0.46 see the* dolphins 0.31 around them 0.92^| then suddenly 1.30 they uh see uh 0.48 dark uh 0.32 cloud and storm 1.47 **errfr**| they feel tired 1.30 **errfr** | th there* almost uh arrive to uh 0.79 the island* 5.18 ^| then they~ uh 0.29 they feel scared **errfr** |and uh her uh 0.98 teacher 0.33 give* uh ^|

No.	Breakdown	Per occurrence	Per minute
1	FLP no.	21 (pauses/total time x 60)	21
2	UFLP INT no.	22 (total silence/total time x 60)	22
3	UFLP EXT no.	6 (total silence/total time x 60)	6
4	Repetition	3 (total repetition/total time x 60)	3
Repair			
5	**D-repair**	**1**	Divided by number of repair
6	Error-to-cut-off	0.40	0.40
7	Cut-off-to-repair	0.85	0.85

8	Repair	1.48	1.48
9	**E-repair**		
10	Lexical-repair	1	Divided by number of repair
11	Error-to-cut-off	0.93	0.93
12	Cut-off-to-repair	0.28	0.28
13	Repair	0.41	0.41
14	**Accuracy**		
15	No. of clauses (::)	10	
16	Number of errors (*)	9	
17	Error-free clauses (errfr)	3	
18	Erroneous clause (^)	7	

B- Sample of the coded data (2)

Student	Speaking time	Prof. level	Task condition	Picture	Order
P21 Bashayer	**60**	**Elementary**	**Dual**	**Storm**	**1**

|the* schools uh have uh children ^ :: to want* uh travel in the sea 1.73 ^| uh[**0.41**] after[**0.41**] uh[**0.22]** the weather **is**[**0.53**] (D-repair) was [0.59] uh[**0.18**] (Grammatical) sunniest [**1.31**][**0.67**] **0.57** full sunniest*[**0.75**] (A-repair)^:: and they was* uh happy 0.46^ | uh[**0.30**]afteri they[**0.88**] **1.02** after they[**0.85**] (Phonological)*uh in the sea**[1.97] 0.55** ^:: the~ 1.50 uh the* thunder storm will begin 0.97 ^| and they will~ 0.31 they will stop in island*^ | uh and uh[**0.37**] night will begin 0.59 uh 0.64 agai[**3.59**] uh[**0.30**] will^ begin[**0.59**] **0.47** (A-repair)^| uh they say* the sharks 1.34 ^:: and need to help[**1.24**] **0.79** and he was need to help*[**1.14**] (Grammatical)**0.34** ^| then they call uh the hil uh 2.45 uh they call uh 0.51 helicopters (?) to help 1.13 **errfr** | and they will uh 0.86 comes to[**2.69**] 1.33 uh 0.55 uh 4.02[**7.41**] they comes to uh the childrens**[1.84]) (Grammatical**)|

No.	Breakdown	Per occurrence	Per minute
1	FLP INT no.	23 (total pauses/total time x 60)	23
2	UFLP INT no.	13 (total silence/total time x 60)	13
3	UFLP EXT no.	8 (total silence/total time)	8
4	Repetition	3 (total repetition/total time x 60)	3
Repair			
5	**A-repair**	**2**	Divided by number of repair

6	Error-to-cut-off	4.9	2.45
7	Cut-off-to-repair	.97	.49
8	Repair	1.34	.67
9	**D-repair**	1	Divided by number of repair
10	Error-to-cut-off	0.41	.41
11	Cut-off-to-repair	0.22	.22
12	Repair	0.53	.53
13	**E-repair**		4
14	**Grammatical repair**	3	Divided by number of repair
15	Error-to-cut-off	4.46	1.50
16	Cut-off-to-repair	8.99	3
17	Repair	3.57	1.20
18	**Phonological**	1	Divided by number of repair
19	Error-to-cut-off	0.88	0.88
20	Cut-off-to-repair	1.02	1.02
21	Repair	0.85	0.85
22	**Accuracy**		
23	No. of clauses (::)	12	

24	Number of errors (*)	9
25	Error-free clauses (errfr)	1
26	Erroneous clause (^)	11

C- Sample of the coded data (3)

Student	Speaking time	Prof. level	Task condition	Picture	Order
P4 Mariam TH	**60**	**Intermediate**	**Single**	**Museum**	**2**

|Uh there is school trip* to the museum^ 0.40 | the kids all* were excited^ 0.26 :: to see the art **Errfr** 0.51 | they went to the mu uh museum **Errfr** |they were amazed 0.47 by the art and uh 0.91 uh paitings* ^ 0.58 |uh 0.60 then suddenly the ala**[0.47]** **0.15** fire alarm**[0.68]** (**A-repair**) to wear on* 0.90^ |the kids were terrified and afraid 0.37 because of the smoke and the fire **Errfr** **1.35** | the[**0.43**] **0.38** one of the persons**[1.35]** uh**[0.62]** one of them**[0.50]*** (**A-repair**) who work in the museum 0.31^ :: help* the kids ^ :: to 0.46 uh 0.26 take the stairs ^ 0.39 |so they're running 0.51 up the stairs **Errfr** :: to get to to the exit **Errfr** 1.04 | uh 0.91 eh they were 0.25 screaming **Errfr** | and uh 0.62 they were afraid 0.27 because of the smoke and the fire **Errfr** 0.89 | thu**[0.34]** **0.14** so[**0.21**] (**Phonological**) they were 0.32 quickly ran* uh up the stairs^ :: to get to the exit **Errfr** 0.60 | uh 0.57[**0.90**] they[**0.59**] **0.19** manager of the museum***[1.42]** (**A-repair**) called uh 0.39 for a help or the 0.84 ambulance 0.42 and uh 0.33[**0.63**] fire[**0.44]** uh 0.39[**0.82]** firefighter**[0.67]** **(lexical) 0.69** ^|

No.	Breakdown	Per occurrence	Per minute
1	FLP INT no.	15 (total pauses/total time x 60)	15
2	UFLP INT no.	18 (total silence/total time x 60)	18
3	UFLP EXT no.	13 (total silence/total time)	13
4	Repetition	1 (total repetition/total time x 60)	1

Repair			
5	**A-repair**	**3**	Divided by number of repair
6	Error-to-cut-off	1.49	.50
7	Cut-off-to-repair	0.72	.24
8	Repair	2.6	.87
9	**E-repair**		2
10	**Lexical**	1	Divided by number of repair
11	Error-to-cut-off	.44	.44
12	Cut-off-to-repair	.82	.82
13	Repair	.67	.67
14	**Phonological**	1	Divided by number of repair
15	Error-to-cut-off	0.34	0.34
16	Cut-off-to-repair	0.14	0.14
17	Repair	0.21	0.21
18	**Accuracy**		
19	No. of clauses (::)	18	
20	Number of errors (*)	8	

21	Error-free clauses (errfr)	8
22	Erroneous clause (^)	9