An examination of the impacts of variations in task cognitive complexity and task input format on L2 oral fluency, measured at the middle and end of clauses

Stewart John Cooper

PhD

Teaching English to Speakers of Other Languages

University of York

Education

September 2020

Abstract

Second language (L2) learners often state a desire to produce speech as fluently as first language (L1) speakers. However, it is still not completely clear how L2 oral fluency is understood, and neither is it consistent in how it is measured. The study investigates the differences in L1 and L2 speakers' oral fluency across levels of task complexity and task input formats. Research has established that higher levels of task complexity return lower levels of spoken fluency. PSTM was also investigated to examine for any correlations between task design, task complexity, and oral fluency. Furthermore, investigating oral fluency at the middle and end of clauses will hopefully enable a better understanding of L2 speech production and the attendant mental processes.

For this research project 80 participants' performance on two narrative tasks were recorded: 40 English first language speakers and 40 Korean participants who speak English as a second language. Each participant performed two tasks differentiated by the level of cognitive complexity (operationalised through the chronological/random order of information) and the input format (pictorial or written). It was hypothesised that pauses occurring 'within a clause' (Skehan et al., 2016) are evidence of difficulties at the formulation stage of speech production (Levelt, 1999), while pauses occurring 'between clauses' provide evidence of problems during conceptualization. A phonological short-term memory (PSTM) test was used to examine the role it plays in mediating the effects of task complexity and task design on L2 fluency.

The findings reveal that L2 speakers show evidence of significant increases in the frequency of pauses at the mid-clause when performing the cognitively more complex tasks. A strong negative correlation was evidenced between PSTM and the frequency of mid clause pauses. These findings are explained in terms of cognitive fluency, automaticity, and the processes of L2 speech production.

Table of Contents

Appendices 209

List of Tables

List of Figures

Acknowledgements

I would like to take this opportunity to thank my supervisor Dr Nadia Mifka-Profozic for her expert guidance and support throughout the PhD process. I appreciate all the chats we had and the friendly advice you gave me. I would also like to thank Dr Zoe Handley for her excellent advice as my TAP member, and also for taking the time to teach me how to properly use PRAAT to code audio files correctly. This was an excellent learning opportunity and something which really helped throughout my studies, as well as providing me with valuable first-hand knowledge of research work. I would also like to offer thanks to Dr David O'Reilly and to Dr Norbert Vanek for their kind and helpful advice on using RStudio to perform analysis on my data.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1 Introduction

1.1 Oral Fluency

Many second language learners state their ultimate aim is to become ‘fluent’ in the target language. Fluent meaning, they would like to be able to speak with confidence, quickly and easily, communicating comfortably with native speakers in the language they are studying. However, despite years of study, practice, and effort many second language learners remain unable to meet their goal of producing fluent speech. These same learners may have excellent knowledge of the linguistic structures and grammatical forms of a language, but this does not result in comparable levels of oral fluency. The language learners then can be said to ‘know’ the language, but not be able to ‘use’ the language, at least not in the way which they desire. The ability to ‘use’ a language fluently often represents the greatest challenge requiring the implementation of linguistic knowledge, what they ‘know’, immediately to communicate, instantly and successfully face to face.

The ability to transfer what we know of a language into the ability to use it in speech efficiently is necessary for the production of fluent speech. If we pause too much, or too often, or if we speak too slowly then we are said not to be fluent. We can consider that in a first language this process is effortless; we do not need to consciously select which words to say, or which order to put them in. However, this is not true in a second language, particularly for less proficient speakers. Considerable mental effort is required to use what we know of a second language in real-time whilst speaking. When a language learner no longer needs to think about what they ‘know’ of words and linguistic structures in one language, and can simply ‘use’ the target language, they can communicate easily and are considered to be able to speak fluently in that language.

Often when ‘fluency’ is the goal institutions, educators, and learners themselves will place emphasis on ‘native-like’ speech and pronunciation, with the stated goal of becoming ‘as fluent as an L1 speaker’. However, with the increasing prevalence of English globally, with more English L2 speakers than English L1 speakers, and the often ubiquitous position of English as a lingua franca in many domains, it is important to note that speech produced in an L2 is not positioned as automatically deficient. Cook (2002) cautions against this ‘bias’ inherent in much of the SLA research, and warns against the danger of adopting the L1 speaker or monolingual as the norm because it fails to capture the difference in bilingual mental processing. The goal of a second language instructor should be to create students who have multi-competence (a compound state of mind with two grammars), rather than attempt to create L2 speakers in the image of L1 speakers Cook (2002). With this in mind, the present study seeks to compare speech in an L1 and L2, not to champion L1 speech as an idealistic norm, or as a target goal to achieve, but rather to provide a better understanding of how speech production varies, and what this can tell us about the underlying mental processes involved in learning and producing a second language.

Achieving a better understanding of fluency is important not just to better meet learner goals, but fluency has also been shown to be an important factor in language development (Kormos, 2011; O'Brien et al., 2007), and as a predictor of overall language proficiency (Housen et al., 2012). Less fluent speakers may have less chance to interact and practice the target language, due to L1 speakers finding them problematic to communicate with, and potentially even avoiding speaking with them altogether (Derwing, 2017). This is an important factor as it reduces the amount of practice available to learners, something which is a vital requirement necessary to improve speaking proficiency and ultimately oral fluency (Segalowitz, 2010). Finally, and maybe most importantly to learners themselves, fluency is an important component of many current English language proficiency tests, such as TOEIC, TOEFL, and IELTS, as well countless university entrance exams, job interviews, and even some promotion and advancement opportunities. These tests often act as gateways, or barriers, to achieving tangible real world success, providing direct intrinsic motivation for learners to increase their grades by improving their oral fluency.

To help students realise this goal several challenges are immediately evident, they range across the entire spectrum of the language education industry. The task facing language instructors is to select and deliver those activities which give learners the most opportunities to improve their oral fluency in the classroom. Course designers must create syllabi which allow for the gradual and continual improvement of fluency over an extended period of time. Language testers need to implement reliable and valid assessments which accurately measure learner's levels of spoken fluency. However, before these myriad demands can be met researchers must first provide empirical evidence, and a clear definition of what fluency actually means, and how fluent speech is produced.

For the reasons stated above, the interest in increasing understanding and more accurately defining oral fluency has become a focus of much research in the area of second language acquisition in recent years. Areas of particular advancement in research include techniques which employ technology to measure fluency (De Jong & Wempe, 2009), understanding how L1 and L2 fluency and speech production differs (Peltonen, 2018; Kormos, 2006; Segalowitz, 2010; Tavakoli, 2011), how fluency can be conceptualised (Segalowitz, 2010; Skehan et al., 2016), the role of fluency in language teaching pedagogy (Tavakoli & Hunter, 2018), how tasks may be sequenced and organised to improve fluency (Tavakoli, 2015), perceptions of fluency (Kahng, 2018, 2020; Préfontaine, 2013; Préfontaine et al., 2016), working memory and its impact on fluency (Wright, 2012), and how fluency relates to the underlying cognitive speech processes of language production and acquisition (Robinson, 2001, 2007; Skehan, 1998, 2009; Skehan et al., 2016).

1.2 Task Complexity

In this research project oral fluency is assessed through participant's performance on language learning tasks, which are part of the task based language teaching (TBLT) methodological approach to language learning. One of the central tenets of TBLT is that a communicative approach to language learning facilitates the ability of students to communicate in real-time and in authentic situations. It could be argued that the aim of improving communicative ability in TBLT creates more of a focus on oral fluency, rather than the more traditional teaching methodologies, which provide more emphasis on the complexity of language structures and the accuracy of production. As a result of the need to better understand how tasks impact language acquisition and performance, as well as the most appropriate way to sequence them in a syllabus, tasks themselves have become an integral part of much research surrounding TBLT, as well as in second language acquisition (SLA). In his general framework for TBLT Long (1985) proposed that tasks should be arranged in order of ascending complexity, from less to more complex, determined by their difficulty for learners to complete. The goal being to create pedagogic tasks which mirror authentic real-world communicative situations. This is an area which has been the subject of much research in SLA. To successfully categorise and define those characteristics of a task which impact complexity, in a reliable and replicable way, is no small undertaking. The aim of this research project is to isolate and examine those elements of task design which impact complexity, in this case the order of the information and the format of the task input, and then to understand how learner performance is affected by the manipulation of these factors. With the ultimate aim of being able to create and organise tasks along the lines of complexity, and task input format, to better suit the needs of second language learners (Long, 2015).

Two major theories have been hypothesised in the area of TBLT, both of which investigate the information-processing perspective of SLA and focus on the cognitive factors impacting learner performance. The aim of research here is to better understand how task performance directs the focus of attentional mental resources, and the extent to which these limited cognitive resources are determined by the characteristics of a task. The first of these major theories was put forward by Skehan (1998, 2009) in the Limited Capacity Model. The model describes how there is a trade-off between the attentional mental resources in linguistic production, meaning that increases in one area will result in corresponding decreases in another area. This competition for cognitive resources means a speaker is able to focus on form, or fluency, but not both. In line with this perspective an increase in task complexity would see a corresponding increase in the required attentional mental resources to transact the task, but would not benefit areas of performance simultaneously, rather the remaining mental resources would be directed towards the specific speech demands of the task. The second major theory is Robinson's Cognition Hypothesis (2001, 2007) which takes the view that attentional mental resources can be allocated simultaneously, and do not compete with each other, allowing for a focus on both linguistic form and meaning at the same time.

Many studies have investigated these competing theories, with much research supporting one theory, and roughly equal numbers supporting the other, but without any real conclusion over which theory is more accurate in describing how mental resources compete during task transaction. As both of these hypotheses, and their more recent updates, are more than ten years old, and in an attempt to explain the inconclusive results obtained so far, research has now shifted towards exploring alternative approaches to investigation of the variances in performance of L2 speakers on tasks. One such approach in regards to measuring and conceptualising oral fluency is that of Skehan et al., (2016) who propose that performance on tasks is better understood by a distinction between 'clause level' and 'discourse level' fluency. The distinction drawn between pauses occurring in the middle of clauses, and those occurring at the end of clauses, Skehan et al., (2016) argue is indicative of the way in which cognitive resources, and mental attention are allocated during speech production and highlights the psycholinguistic nature of the differences between L1 and L2 speech production. Relating increases in the cognitive demands of tasks to particular elements of Levelt's model of speech production (1989, 1999) Skehan et al., (2016) provide explanations for why, as well as where, particular breakdowns in oral fluency occur. It is hypothesised that pauses occurring at the boundaries of clauses, between ideas, are as a result of difficulties in conceptualising speech, and with macroplanning. On the other hand, those pauses which occur in the middle of clauses, between words, are related to formulator difficulties in speech production and provide evidence of problems with microplanning (Skehan et al., 2016).

The present research project intends to investigate the role that task complexity and task input format play in the frequency of pauses in speech, measured at the mid and end-clause position, for both L1 and L2 speakers. Linking increased task complexity, and the resultant losses of oral fluency to specific phases of speech production will provide a better understanding of the cognitive process involved in L2 speech production, and how mental resources are allocated during task performance. This understanding will be facilitated through contrasting and comparing the performance of L1 and L2 speakers' oral fluency measures on the same tasks. The tasks employed in the present study are differentiated by the mental difficulty required to transact them, not the linguistic difficulty. In other words, the same level of language is required to perform the task, but the cognitive processing required and the need to allocate more mental resources for task organisation will increase the level of task difficulty. It is thought that by increasing the mental load on L2 speakers through the increased cognitive complexity of tasks, vulnerabilities in speech production will be evidenced, and the locations of any pauses and breakdowns in speech will provid insight in to which elements of speech production are impacted.

1.3 Task Design and Task Input Format

The present study also seeks to explore the role which the task input format (e.g., written, pictorial, aural, video, etc.) plays in task complexity. Much research has investigated how task design can influence learner performance, but research has mainly focused on operationalisations of the inherent structure of a task (De Jong & Vercelotti, 2016: Kormos & Trebits, 2015; Awwad, Tavakoli, & Wright, 2017) and/or how they relate to directly Robinson's Cognition Hypothesis (2001, 2007). Few studies have so far investigated the role that the input format of tasks plays in language performance, speech production, and the production of fluent L2 speech. This is an area which has been neglected so far in research, and little is known about how the variation in task input format interacts with the cognitive process in the production of fluent speech in a second language (Long, 2015; Préfontaine & Kormos, 2015; Tavakoli & Foster, 2008).

1.4 Phonological Short Term Memory

This research also further explores the relationship between L2 speaker levels of phonological short term memory (PSTM) and oral fluency performance on tasks of varying design. The differences in individual learner's levels of PSTM and working memory (WM) are further areas of research interest in SLA. Research has attempted to explain the variation in L2 performance by linking the cognitive processing benefits of greater access and use of both short and long term memory stores. The two areas which have seen the most research interest are L2 proficiency (Awwad & Tavakoli, 2019: De Jong et al., 2015; Malicka & Levkina, 2012), and working memory (Baddeley & Logie, 1999; Kim et al., 2015; Kormos & Trebits, 2011), defined here broadly as the ability to manipulate and store information relevant to a specific task (Martin & Ellis, 2012). One particular component of working memory is PSTM, defined as the ability to recall and recognise, in order, phonological information for a limited amount of time (O'Brien et al., 2007). Research has linked PSTM to language acquisition (Cheung, 1996; Martin & Ellis, 2012), the development of oral fluency (O'Brien et al., 2007), learning vocabulary (Service et al., 2004), language aptitude (Bolibaugh & Foster, 2013), and has long been linked to the development of speech production in L1 children (Baddeley et al., 1975, 1998; Adams & Gathercole, 1995, 2000). Despite the large amount of research conducted into PSTM, few studies have so far looked at how levels of PSTM can facilitate more fluent L2 speech in adults during the performance of tasks of varying design. The present study seeks to address this gap in the literature and discern to what extent higher levels of PSTM assist in the cognitively demanding requirements of L2 speech production, across a variety of task complexities and input formats.

1.5 Aims and Research Questions

The current research project then seeks to examine the effects of both task design; cognitive complexity, task input mode, and levels of PSTM on L2 fluency. These effects will be examined through a comparison of L1 and L2 speaker oral fluency, as evidenced by pauses at both the mid-clause and end-clause position. The questions that guided this research project were:

RQ 1: What is the impact of a variation in task cognitive complexity on the performance of L1 and L2 speakers' oral fluency measured at the *mid*-clause?

RQ 2: What is the impact of a variation in task cognitive complexity on the performance of L1 and L2 speakers' oral fluency measured at the *end*-clause?

RQ 3: What is the impact of a variation in task input format (written or pictorial) on the performance of L1 and L2 speakers' oral fluency at the mid-clause and end-clause?

RQ 4: What is the relation between L2 speaker's levels of Phonological Short Term Memory (PSTM) and their measures of oral fluency at the mid-clause and end-clause?

1.6 Structure of the Thesis

This introduction will now turn to a description of the layout of the thesis and provide a brief introduction to the content of each chapter. This thesis has been organised into eight different chapters, and chapter two begins with an exploration of the concept of fluency, firstly by looking at the speech production process, then by looking at models of L2 speech production. The stages of conceptualisation, formulation and articulation are detailed, as is the monitoring phase. Definitions of fluency are then proposed from the literature and the three domains of fluency: cognitive, utterance, and perceived. Skills based approaches to fluency are next detailed, before moving on to describe how fluency has been measured in previous research. The chapter concludes with a brief account of a recent study by Skehan et al., (2016) which proposes a new distinction in the way in which breakdowns in fluency are understood in terms of oral fluency and speech production.

Chapter three focuses on TBLT, tasks, task complexity, task input format, and task design. The chapter begins with detail of Willis' (1996) framework and the division of tasks in to three constituent parts: pre-task, task cycle, and language focus. A definition of what makes up a task is then provided with reference to a task supported approach to TBLT. Known issues with implementing TBLT are then discussed, in particular within the East Asian context, as this is the location of the study and all of the L2 participants are Korean L1 speakers. The next section of that chapter explores task design, and the way in which a variation of task design can impact SLA and speech production in learners. The chapter then details how tasks have been employed in research to date. Task complexity is then described with specific attention paid to the competing theories of the Cognition Hypothesis (Robinson, 2001, 2007) and the Limited Capacity Hypothesis (Skehan, 1998, 2009) and how they explain the psychological processes involved in task performance. Studies which explore the impacts of task complexity are then described, with a focus on conceptualising, task design (input format), and PSTM, and how they impact the cognitive processes involved in the fluent production of L2 speech. The section again concludes with a further detailed report of Skehan et al., (2016) with details of how increased task complexity impacts speech production in L1 and L2 speakers, and what the differentiation between pauses occurring at the mid and end-clause position suggests about the cognitive processes involved.

Chapter four looks at the constructs of working memory (WM) and phonological short term memory (PSTM), and provides a definition for both. The section begins with a description of Baddeley and Hitch's Multi-Component model (1974) and how it is comprised of three elements: the phonological loop, the visuospatial sketchpad, and the central executive. Research into the effects of WM and PSTM are then discussed in regards to relevant studies, and a focus on how PSTM has been operationalised and measured. Links are drawn between levels of PSTM and L2 performance variation, occurring as a result of a differentiation in task design. Theoretical explanations are then suggested, in relation to speech models and learning models explored earlier.

Chapter five revisits the research questions proposed by the literature review and the corresponding predictions that the research questions generated. The chapter then provides a detailed description of the methodology employed to answer the research questions, with a detailed description of the instruments employed, the data collection methods, and information on the analysis performed. Details of the measures employed and the justification for the implementation are also covered. The chapter continues with a look at how statistical software was implemented to create models for comparative analysis. The chapter concludes with details of the pilot study and how it informed the research procedures and data collection.

Chapter six explains how the data was collected and coded using PRAAT software (Boersma & Weenink, 2018) and utilised scripts written by De Jong and Wempe (2009), as well as Handley and Haiping (2018). The benefits to the reliability and validity of using automatic measures of fluency is then explained. Inter-rater reliability was then detailed and the high degree of consistency reported.

Chapter seven first details the results obtained from the qualitative and quantitative data analysis used to establish task complexity. It is then divided into four main sections, one to detail the results in relation to each of the research question.

Chapter eight concludes with detailed analysis of how the findings of the research project answer the original research questions posed, and how they relate to the predictions made. Again, each research question is explored in a separate section. The chapter concludes with information on the theoretical, methodological, and pedagogical implications of the study, with the limitations of the current project, and implications of the study for future research.

2 Speech Production and Oral Fluency

2.1 Introduction

The construct of L2 fluency is theoretically complex and multi-faceted, but there are important benefits to be gained from an increased understanding of it, not just for learners and teachers in the language classroom, but also for testers, syllabus creators, and researchers. This chapter will begin with an introduction to, and an overview of, the theoretical research that has so far been conducted into both L1 and L2 speech production. Beginning with an explanation of a psycholinguistic model of L1 speech production, and how this has helped shape the current understanding of L2 speech production, through an examination of bilingual and fluency based production models. The chapter will then move on to a discussion of the varying ways 'fluency' has been defined in literature in qualitative terms, continuing with an examination of current research in the field and a particular focus on the notion of clause and discourse fluency. The chapter concludes with a description of how L2 oral fluency has been measured and operationalised in research, including a discussion of the limitations of these measures.

2.2 Speech Production

The most influential and comprehensive psycholinguistic model of speech production is Levelt's blueprint of a speaker (1989, 1999). Levelt examined the native speaking process of speech production through a psycholinguistic approach and sought to explain how L1 speech is produced. Levelt's model describes the four processes of speech production; how a speaker must first conceive of or 'conceptualise' their message, then 'formulate' it by accessing their language stores, and then 'articulate' the message as overt speech. These three basic steps are widely held to be accurate representations of how speech is produced and are incorporated in connectionist models of speech production, for example those of Dell, Schwartz, Martin, Saffran, and Gagnon (1997), and Nozari, Dell, and Schwartz (2011). However, an important distinction between Levelt's models (1989, 1999) and connectionist models is in the detection and correction of errors. Levelt argues that the production cycle is continuously 'monitored' for errors, with both overt speech, and as yet unvoiced internal speech scrutinised. In connectionist models on the other hand error detection is performed through a process of 'conflict monitoring', carried out by a domain-general executive monitor (Felker et al., 2019).

It is important to point out that in both models of speech production all of these processes are taking place simultaneously, and the ability to produce speech fluently is dependent on them operating efficiently together. As this research project seeks to further understanding of the impacts of task complexity and task input format on the cognitive process employed in speech production by both L1 and L2 speakers, it will draw on the psycholinguistic approach employed by Levelt (1989, 1999) and the studies in SLA which are based on those models.

2.2.1 Conceptualisation

The first stage of Levelt's model is 'conceptualisation' and is initially concerned with the speaker deciding what they want to say; generating the communicative intention or goal of the message, choosing the relevant information for the purpose, and ordering the expression, before concluding with the production of the 'preverbal message' (Levelt, 1989). The conceptualisation stage can be further divided into two stages: macroplanning and microplanning. Macroplanning is a process which involves the speaker deciding on what to say and what order to say it in, it also elaborates on and divides the communicative goal further into smaller subgoals; planning what to say based on the speaker's surroundings, as well as assumptions about the person they are talking to, for example where the speaker selects the appropriate register; whether to speak in a polite, formal, casual, or friendly way. Microplanning on the other hand is focused on the conceptual planning of the message and shapes the information, the topic, its focus, and how it will attract and guide the listener's attention (Levelt, 1989) it also takes into account the speaker's attitude and position. Microplanning is also concerned with the decision-making process for the selection of the relevant lexis which must be retrieved to complete the preverbal plan.

The pre-verbal plan is created as a result of both the macroplanning and microplanning stages of conceptualisation and is held in the speaker's working memory (Baddeley, 1986). Working memory is a complex and well-studied phenomenon which will be explored in greater detail in relation to the current study in chapter 4. For now, it will suffice to understand working memory as a term which is used to define a space available to the speaker for a limited time, and contains the information the speaker currently has access to. In other words, a speaker's working memory contains the currently attended information that can be covered by message generating or monitoring procedures, dealing with the comprehending and producing of utterances. Finally, the preverbal plan is 'sent' or delivered to the 'formulator' for conversion into language.

2.2.2 Formulation

The formulator takes the output from the conceptual stage and translates it into a linguistic structure through the selection of the desired lexical, grammatical, and phonological structures. This is achieved through two core systems of encoding: grammatical and phonological. The first step is the grammatical encoding of the message and is achieved through accessing the mental lexicon. The mental lexicon is described by Levelt (1989) as a repository of all the information about different words in a speaker's language which can be found in their long-term memory, in other words it is a speaker's mental store containing all of their vocabulary. This is where lemmas are accessed and syntactic building procedures are followed (Levelt, 1989). Lemmas consisting of semantic information and syntactic information, as well as lexemes which consist of phonological information and morphological information, are the two types of lexical items selected for grammatical and phonological encoding. When the desired lemma has been accessed and the syntactic building procedures are complete, they produce a 'surface structure' (Levelt, 1989). This surface structure is a string of lemmas in order, of various grouped phrases and sub-phrases and is comprised of the appropriate syntactic pattern for the message. The framework of the message is now prepared and the surface structure enters phonological encoding.

The second stage of coding, which takes place during the formulation phase is phonological encoding. After a speaker has selected the desired phonological and morphological forms for the lemmas in the surface structure, they are then able to add a phonetic or articulatory plan for their speech (Levelt, 1989). This plan is based on grammatical encoding and is where the message is syllabified and intonation patterns are established. This results in the creation of a 'phonological score' where each of the syllables is linked to an appropriate articulatory gesture (Levelt, 1989, 1999). The formulation of planned internal speech is then complete and results in an articulatory plan, containing the phonetic plan and internal speech, being sent to the articulation stage of production where it will be turned into overt speech.

2.2.3 Articulation

The articulation stage of speech production is responsible for the physical process of actually speaking; using the mouth, tongue, lips, velum, glottis, teeth, palate, and breath to create sound (Bygate, 2001a). The articulation stage executes the plan created in the formulation stage, it involves pronunciation and intonation of speech. The generation of internal, unvoiced speech is often faster than the physical process of articulation necessitating the storage of the phonetic plan in a "device" called the articulatory buffer (Levelt, 1989). Chunks of internal speech are retrieved from the articulatory buffer and then in turn are used to produce overt speech.

2.2.4 Self-monitoring

Throughout the three aforementioned stages of speech production a process of self-monitoring simultaneously takes place, allowing a speaker to control their speech either before or after utterance. All of the speech information, overt as well as internal, is stored in the speaker's working memory where it is checked for errors by a self-monitoring system which contains three 'loops' (Levelt, 1989). The first loop assesses the compatibility of the preverbal plan to the initial intended outcome of the speaker, to see if it matches. The second loop checks the phonological score or internal speech to ensure the message is correctly encoded before it is articulated. The third loop takes place as the speaker hears their own speech and monitors it for suitability, as well as checking for errors of pronunciation and the clarity of the speech sound.

2.3 Controlled Processing and Automatic Processing

The stages of speech production in Levelt's models (1989, 1999) can be divided between those which are controlled; those which a speaker is consciously aware of, and those which are carried out automatically. The conceptualisation stage is highly controlled as there are many different ways a speaker can choose to convey their message, depending on the context, thus necessitating the need for constant attention to convey the message as intended. Self-monitoring is also a conscious and controlled process requiring the speaker to make corrections to any errors in their speech. However, the processes of formulation and articulation are carried out unconsciously and automatically, allowing speech to be produced rapidly and accurately without delay by L1 speakers. Several studies have argued that the distinction between controlled and automatic processes is an important difference between language production in L1 and L2 speakers (Kormos, 2006; Segalowitz, 2010) and one which will be examined in more detail later in the section on L2 speech production (2.5).

2.4 Serial and Parallel Production

As detailed above speech production takes place through several phases of production. It is important to note however that these stages consist of both processes that are operating at the same time, as well as those that are performed in a series. The serial process of speech production sees each utterance pass through the conceptualisation, formulisation, and articulation stages in order. The three stages however also operate independently of each other as they are responsible for different and separate parts of speech production. These production processes are important to understand when exploring the similarities, and more importantly the differences, between L1 and L2 speech production, which the next section will seek to address.

2.5 L2 Speech Production

Levelt's model has been revised and adapted several times by different authors to try to better explain L2 speech production (De Bot, 1992; Segalowitz, 2010; Kormos, 2011). All of these adapted models agree that L2 speech production shares many similarities with L1 production, and follows the same serial process of speech production; from conceptualisation through formulation to articulation, as well as containing self-monitoring (De Bot, 1992; Kormos, 2006; Segalowitz, 2010). The key area of difference is assumed to be mainly in the formulation stage, with De Bot (1992) arguing that L2 production requires conscious effort and lacks the automaticity of L1 formulation. This is especially true of lower proficiency L2 speakers, who require extra time to grammatically encode their speech and meet their communicative intentions in formulation. This is due to what Hilton (2008) describes as a distinction between 'higher order' processes, which are related to meaning and deal with discourse planning and are conceptual in nature, and 'lower order' processes which involve linguistic coding through lexical selection and phonological encoding. For an L1 speaker these lower order processes are almost entirely automatic and require little or no attention from the speaker (Levelt, 1999; Hilton, 2008). It is not necessary to consider and ‘think about’ where to place an auxiliary verb in a sentence, or how to properly conjugate a verb in our first language. This process is automatic for an L1 speaker, but this is not true for L2 speakers as both higher and lower order process can require an L2 speaker to consciously attend to the cognitive processes inherent in speech production, resulting in pauses and breakdowns in speech.

2.5.1 Updated Speech Production Models: Kormos (2006); Segalowitz (2010)

Kormos’ (2006) bilingual model of speech production, and Segalowitz’s (2010) L2 speech production model, are both based on Levelt’s (1989, 1999) L1 speaker models, but they take into account the unique challenges of language production in an L2, with Segalowitz’s model being of particular importance to this study as it focuses on how speech production relates to fluency. The following section reviews a current and updated view of Levelt’s model (1989), in light of Kormos’ (2006) and Segalowitz’s (2010) adaptations, and also contains Levelt’s own updates to his model (1999) .

Kormos (2006) and Segalowitz (2010) detail how each of the stages of L2 production matches those of L1 production in Levelt’s model (1989) beginning with conceptualisation, moving through formulation, and then to articulation, with all of these processes continuing while self-monitoring is taking place. Conceptualisation is described by Kormos’ (2006) as when the required concepts to be encoded are activated. The speaker plans the goal of the message through macroplanning, and also which language perspectives to use to express that goal through microplanning (Kormos, 2011). A speaker using an L2 is thought to choose the language they wish to use at this point, in a similar way to how they would select the register they wish to employ (De Bot, 1992). This process produces the preverbal plan, but the message is still not linguistic in form and Segalowitz (2010, p. 10) highlights that it is still not yet “organised in language specific terms.” Up until this point in the process of speech production there is little difference between conceptualisation in a speaker’s L1 or L2 as the preverbal plan generated so far holds no grammatical or phonological information. The conceptualisation in this phase is not language specific and could contain lexis that an L2 speaker does not know (De Bot, 1992). For example an English L1 speaker might wish to talk in a friendly way, in Korean, about their lunch, they have made all of these choices, but the choice of the specific word '식사' (the Korean word for meal, pronounced /ʃɪksɑː/ is still to be decided upon. The English L1 speaker then may create a preverbal plan to talk about a subject in their L2; Korean, without having the prerequisite vocabulary, and not knowing that the Korean word for meal is '식사'. This highlights the fact that the preverbal plan is still just an idea when speech production is at conceptualisation, and may even lead the speaker to a plan which contains words which the speaker does not know in their second language. This provides evidence which shows clearly that the conceptualisation stage is exactly that, dealing with concepts and ideas, but is not language specific, and still does not include any phonological or morphosyntactic data.

Formulation is the next stage in speech production where the preverbal plan is grammatically, lexically and phonologically encoded. The L2 speaker's mental lexicon is accessed to select the necessary lexical items to meet the requirements of the preverbal plan. The appropriate lemma is activated first, as the message is syntactically coded, then is built up through the incorporation of phrases and then clauses. The first major difference Kormos (2006) draws between L1 and L2 speech production is in this the formulator stage with the processing of grammatical and phonological rules. As was pointed out earlier in this chapter these processes are automatic for L1 speakers, but Kormos notes that with L2 speech production these linguistic rules are not always automatized, rather it is possible that they may be stored in the form of declarative knowledge (Kormos, 2011). Declarative knowledge is knowledge that is static and must be consciously accessed. In the case of a foreign language, it will take the form of grammatical pedagogic rules and other linguistic structures that form the basis of a L2 speaker's understanding of the target language. It is Kormos' (2006) contention that L2 speakers must consciously access this declarative knowledge, which is stored in their long-term memory, resulting in greater processing demands and requiring more cognitive resources of L2 speech production at the formulation stage.

Segalowitz (2010) explains that the formulation stage is where the 'linguistic shape' is applied to the preverbal message through grammatical encoding, and where the mental lexicon is accessed to complete the message. This is where words are ordered and their relationship to each other is assigned, and the speaker must access the mental lexicon. At this point of the formulation stage the message is now becoming language specific, but Segalowitz details how similar lexical items may be stored close together, and L2 lexis need not have a distinct and separate store. Rather Segalowitz (2010) explains that in a similar way to how closely related words and phrases in the L1 are stored in proximity to each other, so too are L2 items of vocabulary. Bilingual speakers do not necessarily have a different or separate mental storage, with the target language words kept in locations closely related to the same vocabulary in the L1 (Segalowitz, 2010). Lemma retrieval then is another important area of difference in the speech produced in an L2. The question raised by Segalowitz is whether a speaker using their L2 retrieves a lemma directly in the L2, or if they first need to access the L1 lemma to retrieve it. In other words; is L2 vocabulary only accessible through L1 vocabulary? Segalowitz (2010) suggests that all of the words related to 'meal' for example, may be stored together regardless of language. Much debate has surrounded this area and research has investigated whether there are discrete co-existing language systems; the 'separatist' view (Grosjean, 1989), or whether two or more languages combine to form a total language system; the 'wholistic' (Grosjean, 1989) view. Recent research has tended to support the later position, and provides evidence that bilingual speakers have access to a 'unified lexicon' and draw vocabulary from a store which is not language specific. Cook (2002) however cautions against an over simplification of language systems and lemma storage, and instead argues that there is much evidence to support the notion that language systems are a combination of these two views, with some areas of overlap and distinction, rather than being diametrically opposed. Whichever view is taken here much anecdotal evidence can be seen for lexical items sharing the same storage space, or at the very least being linked to the activation of similar lexical items. One common example is when a proficient L2 speaker, communicating in their L2, suddenly uses some L1 vocabulary. The word is unconsciously spoken and was not part of the 'planned' speech, rather it just 'pops-up'. This does seem to suggest that lemma activation is not language specific, or at least that lemmas are stored based on their meaning, rather than separately for language. In the context of the current study, and the exploration of L2 speech production and oral fluency, it is interesting to note that if direct lemma retrieval does not occur, then it would impose a further step on L2 speech production. This would ultimately increase the mental load on L2 speech production at the formulation stage requiring more cognitive resources, as evidenced by pauses and losses of fluency. Segalowitz (2010) also questions if L2 lemma retrieval could be linked to a speaker's proficiency, or if it is dependent on the specific lemma in question. Whichever it is research (De Bot, 1992; Kormos, 2006; Segalowitz, 2010) agrees that lemma retrieval is an important area of research for L2 speech production and one where disfluency can occur.

As was described earlier speech production implements both serial and parallel processing to allow L1 speakers to create utterances without undue delays (Levelt, 1989) and it is important to remember that each separate stage of production operates in parallel. However, L2 speech lacks the automatic production of speech in formulation that L1 speech provides, and so requires controlled access to grammatical and phonological information from their long-term memory. This information is in the form of declarative knowledge and requires conscious effort to access, so L2 speakers are unable to encode messages as quickly as L1 speakers. This delay is further exacerbated by unfamiliar or complex linguistic structures (Samuda & Bygate, 2005). This distinction in access is partly due to the way in which the vocabulary and grammar are learned; L1 grammar rules are often leant implicitly and as part of larger language structures to form part of a rich and varied mental lexicon, resulting in L1 speakers automatically storing a word in many variations and grammatical forms. This is contrasted with L2 speakers who will have separate knowledge of the word lemma and the grammar, resulting from the fact that they have often learned the grammar and vocabulary explicitly. This is especially true of L2 speakers at lower proficiency levels, who will often require considerable mental effort to put the word lemma and the grammatical function together when using a specifically desired term.

This is an important distinction between the way in which knowledge is acquired, as knowledge which is learned explicitly is classified as declarative knowledge and is not automatically accessed, instead requiring mental attention. With L1 speech grammatical and phonological rules are automatised and form part of the formulator process of speech production (Kormos, 2011). L2 speech however does not benefit from this automaticity because rules are not automatically applied and are instead stored as declarative knowledge (Kormos, 2011). As a result, some required vocabulary and the resultant grammatical structures required for L2 speech necessitates conscious attention to access in the formulation stage of production. This results in accessing declarative knowledge, the explicitly learned factual knowledge, requiring more time and attention, especially for less proficient speakers (Kormos, 2006). It is only possible for a speaker to pay conscious attention to a limited number of things so parallel processing of conceptualisation, formulation, and articulation may not be a realistic possibility for all but the most advanced L2 speakers (Kormos, 2006). L2 speakers then would need to pay attention to each separate stage of speech production individually and in turn, negating the possibility of parallel processing and resulting instead in 'serial encoding' (Kormos, 2011). This is obviously impractical for the real time back and forth of a conversation and may force L2 speakers to focus their attentional resources more either on conceptualisation; *what* they want to say, or formulation; *how* they want to say it. This prioritisation of focus could result in breakdowns in communication resulting in a loss of fluency, with more frequent pauses and restatements, as well as errors in the target language. When an L2 speaker has to produce speech under the regular time constraints required by a typical conversation it is of course possible, and even likely, that disruptions to oral fluency will occur due to other factors as well. A lack of familiarity with the context, or language structures necessary to complete the utterance correctly may lead to disfluencies, but more importantly within the context of this study, are those cognitive processes inherent in L2 speech production which lead to a loss of fluency. Less automatised speech production and the lack of familiarity with employing chunks of preconstructed language in the speech production process can cause major difficulties in L2 formulation, resulting in difficulties producing fast flowing and fluent speech. However, the idea that a specific focus or prioritisation of attentional mental resources can sometimes be required in L2 speech production resulting in serial processing, does offer an interesting insight in to how speech production differs in an L2 in terms of how declarative knowledge is employed.

Articulation in the next stage of production and this sees the planned internal speech processed into an utterance. After lemma selection has been completed the L2 speaker must then transform the surface structure into overt speech. This part of production involves information about the phonology of a lemma, and making the correct choice based on how it sounds. The lexeme or phonological code that each lemma has is used to identify it and enable the creation of a phonological score (Levelt, 1999). This may cause difficulties in L2 speech as identifying the appropriate lexeme or phonological code will likely be more difficult, as this process again lacks the automaticity of speech produced in an L1 (Segalowitz, 2010). The phonological score then needs to be converted into an articulatory score, which necessitates a speaker accessing information from the syllabary, a further information store which provides the data on creating sounds or gestural scores. The gestural score is where the rules and forms for the physical production of sound are held, as well as how to make them, and how they combine. It is possible that the difficulty between selecting the usual L1 pronunciation and the, newer L2 sounds may lead to difficulty, or at least slow down the process (Segalowitz, 2010). This is an especially pertinent factor for L1 Korean speakers of English as the Korean language contains many English loan words, or hybridisations of English words - *Konglish*, which use significantly differing pronunciation in Korean and English. This theory is reliant on the premise that L1 and L2 gestural scores are not separately stored for speakers of more than one language, but as mentioned above are accessed in similar areas of mental storage.

The final phase of speech production is self-monitoring which takes place throughout the three previously mentioned stages and is responsible for checking for errors in language production across the entire process. Theories of self-monitoring are often problematic as it is hard to link evidence of what processes are occurring with the difficulty arising from the silent, or internal, nature of the self-monitoring in the first two loops, as it occurs before articulation. It is obviously simpler to test theories of monitoring in the third loop as evidence is provided with speech, which can be heard and recorded. Attempts have been made to navigate this problem by interviewing speakers and asking them how they felt as they were correcting errors, but this is not an entirely straightforward process (Kormos, 2003) and it is an area that requires further research. Monitoring is also believed to play a key part in a speaker learning a second language (Kormos, 2006). Apart from the obvious practice and practical skills acquired from speaking in the L2, monitoring allows the speaker: to become aware of common errors in their utterances, to attempt new and creative language solutions, to increase their language resources, and to notice any ‘gaps’ in their interlanguage and highlight what they do not yet know (Kormos, 2006). It is also believed that as a language learner’s proficiency increases so too does their ability to monitor their speech production (Lennon, 1990; Kormos, 2006). This improvement is thought to be largely due to the increased attentional resources available as a result of the increase and development of automaticity in language production (Kormos, 2006). These self-corrections and the repairing of errors in articulation demonstrate that automaticity is still developing, and it is possible that they can be used to provide evidence and gauge the levels of automaticity in L2 speech production (Kormos, 2006). Evidence has been recorded that the type of repairs change, but that the quantity remains similar as L2 speaker proficiency increases (Dörnyei & Kormos, 1998).

2.6 Formulaic Language

The important role which formulaic language plays in the conceptualisation stage of speech production is a further important area which Kormos (2011) points to as being a significant difference between L1 and L2 speech production. Formulaic language is language that is not creatively constructed, rather it constitutes combinations of remembered expressions, sentences, and phrases (Kormos, 2011). Formulaic language then refers to those expressions that are fixed in meaning, and have become to some extent 'fossilised' in the mental lexicon. These formulaic expressions are typically used for greetings: 'You alright?', 'How d'you do?', making requests: 'I'll have a …', 'Can I get …', or to satisfy a plethora of other basic communicative functions and are activated in the conceptualisation stage of speech production. These language groupings cover many varied concepts which are in turn linked to linguistic data stored in the lexicon as a single lemma. Individual words are stored separately, but can also be activated and retrieved as a unit that combines these words. For example, 'fine' and 'thanks' are stored separately, but also together as one unit, or piece as 'fine thanks.' The item can then be activated when necessary to achieve the desired communicative goal. The advantage these groupings of words provide, both to L1 and L2 speakers, is enabling faster speech production with less conscious effort in comparison to when a speaker is constructing original, creative utterances (Kormos, 2011). This formulaic language provides a large amount of language in ready to use preconstructed groups, which are available to a speaker as a single choice, despite being compiled of single words (Sinclair, 1991). The need for these formulaic items may be necessitated by the pressures of real time language production in face-to-face conversation, but they may also be a reflection of the recurring nature of the language that is most commonly used in daily interactions. Fixed expressions and chunks of language are often ‘fossilised’ by usage as everyday conversations, and are almost always lexicalised i.e., used as one word, e.g., “how d’you do”, “what’s up”, etc. These language chunks provide an illustration of how the memory stores work to reduce effort in lemma retrieval and word selection (Sinclair, 1991), with formulaic sequences represented in the mental lexicon and are thus able to be processed faster than normal language (Conklin & Schmitt, 2012). These formulaic language sequences can be of great benefit to L2 speech production as they allow speakers to speak more ‘fluently’, as they reduce the amount of mental processing and conscious attention required during the conceptualiser and formulator stages. This means those students who have been taught using TBLT, a communicative approach to language learning, where for example verbs are learnt with the accompanying prepositions, will be able to produce and employ language more efficiently and more ‘fluently’. This is because lemma retrieval is eased in the formulation stage of processing, by the preconstruction of these formulaic groupings. Even for intermediate level learners the benefits of employing formulaic language and using chunks of preconstructed speech will have benefits to speech production and oral fluency, especially on topics which they are familiar with and have most practice using, though perhaps not for all topics, or for those which they are unfamiliar with.

2.7 Summary of L2 Speech Production

In both L1 and L2 speech we can see that speech production processes follow a similar pattern, and occur in the same order, as detailed by Levelt’s (1989) four main stages; conceptualisation, formulation, articulation, and monitoring. The conceptualiser is responsible for the content planned, the formulator provides the linguistic detail, the articulator produces the physical speech, and the monitoring stage checks for any errors. A distinction can be drawn between the stages of conceptualisation and monitoring, which are regarded as being under the conscious control of the speaker, and the formulation and articulation stages which are carried out automatically. In L1 speech production messages are completed incrementally which each stage simultaneously working on different parts of the same message. Less proficient L2 speech production however may lack this parallel processing and processes are ordered in a serial manner, there is little automatic processing, and L2 production often requires the speaker to access a declarative memory store.

These differences between L1 and L2 production are found at ‘vulnerable points’ (Segalowitz, 2010) in the process where L2 speech requires greater effort due to increased processing demands, which in turn leads to delays and results in the slowing down of speech production. As the formulation stage demands greater attention so the articulation and monitoring phases suffer in contrast, there are no longer enough attentional resources to complete these three stages at the same time, resulting in losses of fluency. This necessitates what Kormos (2006) calls ‘serial processing’; where the speaker must complete one stage before moving on to the next, which eventually results in greater demands when trying to restart the parallel processing to allow for the flow of fluent speech.

This is not the case in L1 speech production however, as it is believed to be utilising automatic and parallel processing, allowing the production of speech more easily and efficiently. The ease with which speech is produced in an L1 allows for faster and smoother, or more ‘fluent’ speech. The rate at which a person speaks and the lack of pauses and errors in speech are often used as key indicators of L2 proficiency and oral fluency, and in the next section I will explore oral fluency and the various ways in which it can be defined and measured. The chapter will now shift in focus from speech production to an examination of oral fluency in the section below.

2.8 Fluency

Fluency is notoriously difficult to define due to it being a multifaceted and multidimensional construct (Housen et al., 2012). Research findings from empirical studies on oral fluency have shown varying results, often due to widely differing definitions of and operationalisations of fluency and its attendant variables (Préfontaine & Kormos, 2015). One of the original investigations into fluency and its definition was conducted by Fillmore (1979) who details four aspects in a definition of spoken fluency; not pausing whilst speaking a lot, speaking in a reasoned way using lots of complex language, speaking about lots of different issues with lots of different people, and to be witty by making jokes and expressing original ideas. While Fillmore's (1979) definition provides a starting point and a broad understanding of fluency; detailing the importance of such elements as how quickly someone speaks; how much they say; as well as the amount of pauses they make, as a definition it is rather vague and is hard to quantify and measure consistently. More importantly in relation to this study relying solely on Fillmore's definition would be problematic in the creation of operationalisations of oral fluency. Pawley and Syder (1983) undertook one of the first studies into L2 speaker fluency and they described fluency as the ability to produce fluent stretches of speech. A far narrower and more focused perspective than Fillmore's description, and one which provided a starting point for many further studies seeking to define fluency in more fine grained terms.

2.8.1 Defining Fluency

A widely used definition which provides a commonly agreed upon basis for many further studies into the understanding of fluency for L2 speakers is that of Lennon (2000, p. 26) who describes fluency as the "rapid, smooth, accurate, lucid, and efficient translation of … thought or communicative intention under the temporal constraints of online processing." In other words, the ability to say what you are thinking and to be understood, quickly and efficiently without too much pause or delay. Lennon's (2000) definition takes into account not just the rate at which speech is produced including terms such as 'rapid' and 'smooth', but it also acknowledges mistakes, saying that fluent speech should be both 'accurate' and 'lucid', as well as concerning the psycholinguistic nature of the speech production process incorporating terms such as 'translation', 'communicative intention', and 'online processing'. A similar definition to Lennon's (2000) states that speaking fluently is being able to use a language in real time, with emphasis on meaning, as well as *possibly* being able to incorporate using more lexicalized systems (Skehan & Foster, 1999). This definition again focuses on the 'real time' aspect of speech in relation to speed, while also having an 'emphasis on meaning', acknowledging that the communicative goal of speech must be met, but also adds that language form, is of less importance than the communicative goal by stating that greater lexicalisation is only a 'possibility' and not a requirement (Skehan and Foster, 1999).

Lennon's (2000) definition focuses more on accuracy as an indicator of fluency, than Skehan and Foster's (1999), but both share the view that fluent speech should, at the very least, involve spontaneous communication, without delay, and be focused on achieving the desired communicative goal.

Definitions of fluency specifically for an L2 language user can be considered in two senses the: 'broad' and the 'narrow' (Lennon, 1990). The 'broad' sense refers to total oral proficiency and means that a fluent speaker has a strong command of the language e.g.: error-free grammar, a large vocabulary, and native speaker like pronunciation. In the 'narrow' sense fluency refers to a single component of oral proficiency such as might be recorded in an oral language assessment in a speaking test (Kormos, 2010) and can be described as the impression created for the listener that the speaker can plan, produce, and deliver speech easily and smoothly (Lennon, 1990). The narrow definition of fluency is frequently used alongside accuracy and complexity to measure and asses a speaker's speech performance and language proficiency. This view of fluency as either 'broad' or 'narrow' can be considered overly simplistic and can lack the required nuance necessary to accurately describe fluency in an L2 speaker (Tavakoli & Hunter, 2018). It may be more useful to define fluency not as opposite ends of a spectrum, but rather on a continuum, from very broad, i.e.: when a language student says they wish to be 'fluent' in English or 'master' the language, to the very narrow i.e.: fluency as the frequency of pauses made in a given speech sample (Tavakoli & Hunter, 2018). This definition arises from research Tavakoli and Hunter (2018) conducted into language teacher's perceptions of fluency, as opposed to the more widely regarded views of language researchers, and seeks to bridge the gap between the two to provide a more accurate and agreeable definition for both groups. While this is certainly a worthwhile effort and helps to inject some practicality in to the debate about how to best define fluency, it is concerned primarily with perceptions of oral fluency, which is just one of several domains of fluency. These different domains of oral fluency will be explored in the following section, but it is clear from the above examples and definitions that an agreed upon definition of fluency is problematic, and that there are many different ways of conceptualising fluency. For the purposes of the present study, it is the more narrowly focused and detailed view of fluency which this research seeks to explore.

2.9 The Three Domains of Fluency (Segalowitz, 2010)

Segalowitz (2010) details how the narrower view of fluency can be further broken down into three distinct domains. The first is *cognitive fluency*, which refers to the underlying mental processes that are responsible for the efficient production of speech. The second is *Utterance fluency*, which are those parts of speech that can be measured acoustically and display the speakers' cognitive fluency. Thirdly is *perceived fluency*, which refers to inferences made by listeners about the speaker's cognitive fluency based on their perceptions of their utterance fluency (Tavakoli, 2016).

2.9.1 Cognitive Fluency

This following section will explain Segalowitz's (2010) three domains of fluency. The first of these is cognitive fluency and it refers to the mental capacity to deal with the varied processes which constitute the speech production process. Cognitive fluency allows a speaker to produce speech efficiently and smoothly, without interrupting the delivery of an utterance. Lemma selection and phonological encoding are examples of two of the processes which must be carried out efficiently, if a speaker is to produce cognitively fluent speech. Cognitive fluency is thought to be related to how much speech processes have become automatic (Kormos, 2006; Segalowitz, 2010). As was detailed above, the amount of conscious effort required to undertake the mental processes involved in speech production, such as lemma retrieval, affect the extent to which parallel processing can take place. If these processes become proceduralised and automatic, no longer requiring direct attention and conscious effort, then speech production can become smoother and more 'fluent'. However, the concept of 'automaticity' is one which has also seen much discussion with often varied definitions; both for how the process is defined as well as how a process can become automatic in the first place. The following section details a brief outline of the development of automaticity in speech production.

2.9.1.1 The Development of Cognitive Skills and Automaticity

Automaticity in second language acquisition has been widely researched and as a result has a generated a variety of different definitions (Levelt, 1989; DeKeyser, 2001; Segalowitz & Hulstjin, 2005). However, most theories are derived from research published by Anderson (1982, 1983) on the acquisition of cognitive skills. The acquisition of a skill first begins through the "interpretive application of declarative knowledge"

(Anderson, 1982, p. 403) resulting in the formation of a procedure. In other words, knowledge is first acquired then as this knowledge is used it can move towards becoming proceduralised. This procedural form of the knowledge or skill then gradually and continually undergoes refinement, often through the conditions in which it is applied, this results in increases in the speed at which it can be accessed and deployed. Declarative knowledge takes more time to utilise as it requires retrieval from long term memory storage, and this retrieval results in a heavy burden on the speaker's working memory (Anderson, 1982). This burden on working memory is directly linked to the errors and slowness of L2 speech (Anderson, 1982), as accessing declarative knowledge stores takes cognitive resources away from speech production.

Speed plays an important role when seeking to define the process of automaticity, this is because as a process becomes more automatic it becomes easier to deploy and is faster to use (Segalowitz, 2010; Anderson, 1983). However, the speed of processing cannot be the only factor when determining if a process is automatic or not, there must be something else happening (Segalowitz, 2010). As a process becomes automatic it must also become more reliable and is ultimately refined and becomes more efficient. Anderson (1982, p. 381) agrees and says that: "By building up procedures to preform specific [language] tasks … a great deal of efficiency is achieved both in terms of time and working memory demands." Automaticity in language production then can be said to involve to some extent the speedier and more efficient utilisation of skills through a compilation of processes, moving from the interpretive application of declarative knowledge, to the direct application through proceduralisation (Anderson, 1982).

While this helps us to understand what automaticity actually is, it still does not help us to understand how it can be achieved. Theories regarding automatization of speech production can be broadly divided into two distinct groups; those which are 'rule based' and those which are 'item based' (Kormos, 2006). The rule-based theories are similar to theories of skills acquisition and follow the process that once declarative knowledge becomes procedural knowledge, it is then a habitual process and can ultimately become an automatic process. Declarative knowledge enables learners to utilise the target knowledge, and by continually and repeatedly employing it form procedural knowledge, with the repetition then forming a habit, and the resultant habit becoming automatised through yet more repetition (DeKeyser, 2017). Item based theories on the other hand differ, suggesting that linguistic forms and phrases can be remembered as easily accessible chunks, which do not need to be analysed.

Rule based theories of automaticity start from the principle that language learners first learn the pedagogic rules and grammar of a language, which is defined as declarative knowledge, and that this knowledge becomes procedural, and then eventually habituated and automatic. This notion of second language acquisition is still widely regarded as accurate in much of the literature, despite the increase in communicative language teaching methods and the increased focus on communication in language teaching (DeKeyser, 2017). In the context of the current study, which explores Korean L1 speaker fluency in English as an L2, there is evidence of reliance on more traditional forms of language learning (PPP, Grammar Translation), and reluctance to fully embrace communicate language teaching (CLT) methods. This may be because of a stronger support for, and belief in, rules based theories and can perhaps be linked to the state of Korean English language learning, which often prioritises the learning of grammatical rules and vocabulary, at the expense of speaking and practicing communicating in authentic situations. This may well be a factor in making Koreans English L2 speakers good at tests of grammatical knowledge, but poor at communication, as they have not had the chance to practice and proceduralise language production skills, and therefore may well be less fluent speakers.

As was explained before when a speaker draws on declarative knowledge during speech production it requires a conscious effort and can lead to dysfluencies and breakdowns in the communication process, due to a lack of attentional mental resources. However, through practice and a process of trial and error, doing it faster, and doing it more accurately and more often they first develop procedural knowledge which in turn then becomes automatic (DeKeyser, 2017). If we follow the example used by Anderson (1982) of solving a complex mathematical problem for the first time, this is obviously something which is quite daunting initially, and requires much conscious effort and practice; we have to follow a set of rules and often find it difficult to proceed from one step to the next. However, after receiving the declarative knowledge, in the form of the mathematical procedure necessary to successfully solve the problem, and with enough repetition and implementation of the procedure required to practice using this knowledge, the vast majority of students are able to rapidly and easily complete problems that were once daunting. Anderson (1982) explains this process comes about through a process called the ‘compilation of knowledge’, which is where skills move from declarative to procedural. The process involves the proceduralisation, composition, and the collapsing of production sequences into single productions, all of which combine to embed the declarative knowledge into productions and skills (Anderson, 1982). In the above example of a mathematical theorem the sequential steps followed to solve the problem i.e.: ‘square the value of A’, then ‘square the value of B’, then ‘sum the two squares to calculate the length of X’ etc., would be collapsed into a single procedure i.e.: ‘calculate the value of x’. The example used here of a mathematical formula is an oversimplification, but it is useful for illustrative purposes, and to explain how a procedure can gradually become a habitual skill, with each constituent step requiring less, if any, conscious effort, and proceduralisation resulting from the repeated application of the process.

More mundane examples of skills which the majority of people have which have become habituated, include things such as: tying a neck tie, fastening our shoe laces, riding a bike, recalling a specific phone number. At first these skills required a great deal of focus and conscious mental attention to perform well, with instructions built up incrementally and sequentially enabling us to complete the tasks. However, they have now become proceduralised and automatic processes; we can all put on a tie or tie our shoe laces whilst having a conversation or watching the television. The rules we had learnt explicitly can now be applied automatically and do not require conscious effort on our part.

Segalowitz and Hulstjin (2005) believe that learning a foreign language follows the same principles, and with enough practice in both productive and receptive skills the mental processes involved in speech production can lead to proceduralisation, automatisation, and ultimately more fluent speech. Kormos (2006) agrees that language learner’s start with declarative knowledge of pedagogic rules and linguistic structures and with sufficient practice this knowledge becomes procedural, and eventually through continued use the speaker will be able to follow the grammatical or linguistic conventions without needing to directly think about it. Interestingly and with implications for second language learning and teaching this declarative knowledge can become 'lost' with the initial underlying knowledge becoming inaccessible as the process becomes automatic (Kormos, 2006). This is as a result of L1 grammar being acquired implicitly and this phenomenon is evidenced by 'native speaker intuition', where by an L1 speaker has knowledge of the correct grammatical form, but may be unable to explain the knowledge behind it. This is because either they learnt it as an infant and have now forgotten it, or never learnt it explicitly in the first place, either way they are still able to distinguish between correct and incorrect implementation of the grammar. This highlights the importance for language teachers to study grammar, because knowing something implicitly and being able to produce it automatically does not make one able to explain it explicitly. The same may also be true of advanced L2 speakers, who may well have forgotten the grammar rules they can now use with ease, through a process of proceduralisation in speech production.

The importance of, and distinction between explicit and implicit knowledge and the role they play in L2 learning has been an area of much research in SLA since Krashen (1977) first detailed their differences relating to language ability development (Han & Finneran, 2014). Explicit knowledge has been linked with metalinguistic knowledge and is subserved by declarative memory, whereas implicit knowledge has been linked to neurolinguistic knowledge and is subserved by procedural memory (Ellis, 2005). As was mentioned previously explicit knowledge is open to introspection and its use is deliberate and intentional, whereas implicit knowledge is not open to introspection and its application is effortless and not intentional (Hulstjin, 2005). To return to the previous example, the ability of a native speaker to tell when something is grammatically incorrect, but unable to explain the reason why, is an example of implicit knowledge and how it is stored in procedural memory. If we explore this in terms of language learning we can consider the difference between the way in which an L1 is largely first learned *implicitly*, and stored in *procedural* memory, and the way in which an L2 is learned *explicitly* and stored in *declarative* memory.

Much of the understanding of explicit and implicit knowledge enjoys a consensus in research, however a far more divisive issue is the extent and manner that these two forms of knowledge interact, if indeed at all. A short description of the major positions follows. The extent to which they connect or overlap in referred to in the literature as ‘interface’, and is reflected by three competing positions; the non-interface, the weak interface, and the strong interface. The strong interface position is exemplified by skill acquisition theories (Anderson, 1982) and suggests language learning is akin to skill development and starts with a declarative stage (knowledge that), before moving on to a procedural stage (knowledge how), and finally a stage of automatization, whereby the procedural knowledge becomes effortless, and ‘fluent’ (Han & Finneran, 2014). The weak interface hypothesis, which contains several variants all lying somewhere in-between the strong and no interface positions, holds that both explicit and implicit knowledge are possible outcomes in SLA, with Ellis (2005) arguing that explicit knowledge can become implicit under certain conditions, such as the learners current state of learning and proficiency. These weak interface theories are more closely tied to skill acquisition theories such as Anderson’s (1982), as they strongly emphasise the importance of consciousness in the learning process, and the indirect role of explicit knowledge in developing an implicit knowledge of the language. The noninterface position holds that it is not possible for learners to learn everything, and so believe that explicit learning has far less importance (Krashen, 1981). A distinction is drawn here between knowledge that has been learned, and knowledge that has been acquired, and that the two are exclusive, dissimilar and cannot replace each other (Han & Finneran, 2014). More recently research has attempted to reconcile these positions, with Han and Finneran (2014) claiming that L2 speakers will have both explicit and implicit knowledge in their interlanguage and with all three types of interface working between them.

Whichever interface position is espoused it has important ramifications for language learning and teaching. In relation to the current research project, we may notice the difference between the way in which knowledge is accessed during speech production; if it is explicit knowledge it is accessed deliberately and will require conscious effort and precious mental resources, however it the knowledge held is implicit and is stored in procedural memory it will allow for faster access and require less attentional mental resources to activate. We can draw the conclusion that creating and utilising implicit knowledge is important for the production of fluent L2 speech.

Item based theories of automatization offer another explanation for how speech processes become automatic. Research has demonstrated that declarative knowledge leading to proceduralisation is not the only way speech can become automatised. Kormos (2006) explains how ‘chunks’ of language that have been built by learners through the considered application of grammatical and phonological rules become saved as a single phrase in the memory. This is a distinction between formulaic language, as formulaic language is not considered based on grammatical or phonological rules. They can then be accessed later as a single ‘unit’ of speech, even without overt awareness of the declarative rules at the time of performance. These language chunks can be deployed rapidly and as a complete ‘unit’ by the speaker, who may not be fully aware of the individual rules or meanings of each word, but rather understands the whole for its communicative effect (Kormos, 2006).

Cognitive fluency then is reliant on the ability of a speaker to produce speech automatically, either through the process of declarative knowledge developing toward a procedural process and ultimately becoming automatic, or through the storage in the speaker’s memory of chunks of language units which can be accessed and utilised complete without requiring individual linguistic analysis.

2.9.2 Perceived Fluency

Perceived fluency is the next of Segalowitz’s (2010) domains of fluency and is used to refer to inferences that listeners make about the fluency of a speaker. The listener’s perceptions are based on inferences about the cognitive fluency of a speaker, how quickly and smoothly they speak; their utterance fluency (Segalowitz, 2010).

Throughout research and in the literature, there are a range of different perspectives of what is considered to accurately define perceived fluency. There is a contrast between Segalowitz’s (2010) tightly focused view of perceptions of fluency which refers only to those based solely on a speaker’s cognitive fluency, with the far broader interpretation of each individual’s views. The individual’s perspective may take into account several different aspects of a speaker’s fluency based on a wide range of different criteria, inevitably varying between individuals. The differing views and varied criteria make this broad view of perceived fluency extremely hard to define and means it will not have much in common with several other definitions used in L2 speech production research (Brown, Iwashita, & McNamara, 2008).

A further issue with perceptions of fluency is the extent to which communicative strategies allow language learners to avoid certain topics or grammatical structures which may be problematic for them, thus allowing them to 'hide' their lack of fluency. A language learner who is proficient in an L2 will have also become practiced in the use of several communication strategies allowing them to 'paper over the cracks' in their procedural knowledge through the utilisation of their declarative knowledge (Dekeyser, 2017). This is possible without any loss of fluency as higher levels of oral fluency will provide the speaker with enough mental resources to detect possible areas of difficulty ahead of time and avoid them through the use of 'chunks' and phrases requiring only a little declarative knowledge (Dekeyser, 2017).

2.9.3 Utterance Fluency

Utterance fluency is the third aspect detailed by Segalowitz (2010) it also has a narrow and a broad perspective. A narrow-focused view of utterance fluency refers just to those features of speech which are attributable to cognitive fluency (Segalowitz, 2010). However, a much broader and open to interpretation definition is often used which includes any single part of speech which could be a reflection of cognitive fluency of a speaker. This distinction is complicated by the fact that there has been little research in this area, and there is still very little evidence to match utterance fluency measures with cognitive fluency measures. This may be because of the inherent difficulty in tapping into specific cognitive fluency measures reliably, whilst at the same time relating them to aspects of utterance fluency, which are often highly susceptible to a number of confounding variables within individual speech patterns.

L2 speakers can mitigate their lack of cognitive fluency by employing a number of strategies; pausing, using repetitions, asking for agreement, etc. in any given instance in which it is impossible to talk at the required speed necessitated by the time demands of face-to-face oral communication. Pauses in the flow of speech may be most commonly associate with difficulties in word selection, or in applying the correct grammatical rule, but pauses come in various forms, and meet various needs. Kormos (2006) describes how L2 speakers can react to the time pressures of communication either by making use of *stalling mechanisms*; these are techniques which 'buy time' for the speaker and allow them focus more on meeting the cognitive demands of formulation, without the added demands of articulation. Examples of stalling mechanisms include; pauses, which are further divided into filled (sounds, such as 'er', 'um', etc) and unfilled (silence); sound lengthening, where a sound, commonly a vowel is elongated ('Aaaanyway'); using filler words ('like') or phrases ('Do you know what I mean'); repetitions of their own words, or repeating something the interlocutor said (Dörnyei & Kormos, 1998), or they can alter their message by simplifying it, replacing some sections with more simplistic language; or by giving up on what they were saying completely (Segalowitz, 2010). These pauses and stalling mechanisms often apparent in L2 speech have formed the basis for much of the evidence and indications for measures of oral fluency, a topic which the next section will explore in more detail.

2.10 Measuring L2 Oral Fluency

The task of accurately and consistently operationalising and measuring oral fluency has proven to be even more problematic and contentious than that of defining it (Préfontaine & Kormos, 2015). This is primarily because of the multifaceted nature of fluency, but is also as a result of the plethora of measures which have been employed in research to date (Ellis, 2020). However, despite the problematic nature it remains an clearly important task to accurately measure not just oral fluency, but also learner performance, and improvement in an L2, as this will have great benefit to not just language learners and teachers, but also to examiners, researchers, content designers, and policy-makers. To this end Skehan (1996) devised a three-way model of L2 proficiency; which views fluency, as well as accuracy, and complexity, as the core components of the complexity, accuracy, fluency (CAF) framework. Complexity is defined as the ability of a speaker to use more advanced language, which may not always be controlled effectively. Accuracy is explained to be the process of avoiding errors in performance, this includes making 'easier' rather than 'harder' language choices and steering clear of any areas which the speaker knows may be problematic for them (Ellis, 2009). The final component from Skehan's framework (1996); fluency is defined as the ability to emphasize meanings and use language in real time (Ellis, 2009).

CAF measures have become ubiquitous as a way of measuring L2 proficiency and SLA, as well as providing the basis of ratings for language learner's performance on many important L2 examinations both written and spoken examples include: TOEFL, TOEIC, and IELTS. However, despite the constant inclusion of CAF measures and constructs for examinations of L2 performance they are often operationalised in different ways, across differing research environments. This problem with a lack of reliability is further compounded by the CAF framework's lack of accuracy when exploring L2 improvement and development. It is often difficult to distinctly differentiate the three constructs from each other and to say definitively where the influence of other variables and outside influences lie, problematising the isolation and analysis of individual effects (Housen & Kuiken, 2009). The importance then of clearly defining the key constructs in a study, as well as how and why the CAF framework will be implemented in a particular study is made evident. It is often important to acknowledge the influence the other CAF factors have on each other and their influence on fluency, as well as on overall L2 performance (Tavakoli & Skehan, 2005). For the purposes of the present study however, considering the measures selected and the aims of the research in investigating a specific measure of oral fluency, the focus will be solely on fluency.

2.11 Speed, Breakdown, and Repair Fluency (Skehan, 2003)

Oral fluency has most commonly been operationalized into three distinct fluency domains coined by Skehan (2003) those of 'speed', 'breakdown', and 'repair'. Speed fluency is concerned with the linguistic density of speech and how quickly a learner speaks, it is often measured by counting the number of syllables produced. Frequently employed measures include articulation rate and speech rate. Breakdown fluency looks at the continuous flow of speech and typically measures the number and length of pauses as well as the 'phonation time' i.e.: the amount of time spent speaking, as opposed to pausing. Repair fluency seeks to identify the frequency of errors in language and how they are repaired with repetitions, false starts, and corrections accounting for the most commonly used measures. These constructs and the attendant measures have been used widely in research, and are among the most prevalent for measuring fluency in L2 oral speech. Recent studies employing these measures include: Lahmann et al., (2015); De

Jong et al., (2012); De Jong and Perfetti, (2011); Préfontaine and Kormos, 2015; Segalowitz et al., (2017); Skehan and Shum, (2014); Suzuki and Kormos, (2020).

Revisions and refinements of the three domains of fluency have often taken place, with Tavakoli and Skehan (2005) updating these domains of oral fluency into two broader groups. The first of group is 'temporal fluency' and details measures which seek to reflect the time sensitive components of L2 speech. Examples include; the speed at which a speaker talks, as well as the frequency of pauses. The second updated group is 'repair fluency'; which seeks to address the online monitoring aspect of speech production and the tendency to correct errors in speech, for example counting the number of self-corrections, restatements, and false starts. However, in a further revision Kormos (2006) describes *four* key divisions of oral fluency measures in the study of L2 learner's speech. They are those concerned with 'temporal aspects', 'temporal aspects combined with turn-taking', 'phonological aspects', and finally aspects of 'formulaic speech'. From a theoretical stand point, we can see that important features of speech production are related to both cognitive fluency, as well as the automaticity of speech production. As automaticity increases and cognitive fluency improves, we can expect to see that L2 speakers are able to speak faster, and with fewer pauses. However, as detailed above, this can be problematic as more advanced L2 speakers are also able to employ communicative strategies which mitigate their lack of cognitive fluency. The links between cognitive processing and how it relates to oral performance have been relatively under researched, one recent study De Jong et al., (2015) seeks to expand the knowledge in this domain, but utterance fluency, and the way in which it reflects and reports on cognitive processes during L2 speech production remains an area of uncertainty, and one which is in need of further research and exploration. This study aims to shed more light on this issue and investigate the manner in which cognitive processing interacts with oral fluency.

2.12 L2 Fluency Measures

A great deal of research has been conducted into the oral fluency of second language speakers, however much of this research has employed a wide variety of different operationalisations and measures of oral fluency. This means undertaking a replication study is often not possible, due to the specific context not being widely generalisable, or the research measures employed often vary and are not measuring the ‘same’ aspect of fluency. All of this underscores the importance of isolating and identifying reliable and valid measures of oral fluency, and those which target specific second language learners as well as particular areas of oral fluency. Much research has been conducted into which measures are the most valid and reliable in terms of assessing L2 speaker fluency, in what particular context and under which given conditions (Segalowitz et al., 2017). However, as Kormos (2006) points out many of the measures of fluency used are still problematic as the majority of studies which employed them had very small sample sizes, with relatively few participants, and unlike the present study did not use any computer-based technology for the increased reliable identification of pauses. The benefits of computerised measures of fluency are not limited to greater reliability and consistency, but also provide an increase in the objectivity of those measures collected.

Studies have investigated comparisons of L1 and L2 speaker’s oral fluency through pausing (Foster, 2001), though these studies have tended to focus on how these pauses reflect perceptions of fluency (Kahng, 2018), rather than the underlying cognitive processes of speech production. This research however, does provide insight into the reasons pauses occur in an L2 and how they are related to cognitive fluency through comparisons with L1 speaker performance. Skehan and Foster (2007) as well as De Jong (2016) found that pauses from L2 speakers were more common within AS units. Tavakoli (2011) conducted a study of L1 and L2 speakers pause patterns and the findings showed that L2 speakers have longer and more frequent pauses, the study also provided evidence to suggest that L2 speakers pause more within clauses, which this study refers to as *mid-clause pauses*. De Jong’s (2016) study also found that pause location was a difference among the pausing patters of the L1 and L2 speakers, with L2 speakers again found to pause more at the mid-clause level. Studies have also been conducted into comparisons of L1 and L2 speaker task performance on a language learning style task, in order to find measures of oral fluency which are valid and consistent (De Jong, 2016; Tavakoli, 2011). The idea for research in this area is that L1 and L2 speech production is different because of the underlying cognitive processes, and how developed, or undeveloped they are. So, the utterance fluency is seen to be a reflection or representation of the cognitive process of speech production. More recent research has provided evidence to support the claim that perceived fluency is strongly associated with fewer mid-clause pauses (Suzuki and

Kormos, 2019; Kahng, 2018), and with direct relevance to the present study, that mid-clause pauses are related to L2-specefic cognitive measures (Kahng, 2020).

In an effort to bring unity to the research in oral fluency measures, and in an attempt to create some consensus of measures Kormos (2010) conducted an analysis and review of several recent studies into oral fluency. The review found that the most reliable and the most valid predictors of L2 speaker's oral fluency were speed fluency and breakdown fluency factors namely: *speech rate*; the number or syllables articulated per minute, the *mean length of runs*; the average number of syllables spoken between pauses of 0.25 seconds or greater, and to a lesser extent *phonation time ratio*; the amount of time spent speaking as a percentage of the total time taken to produce the speech sample. The study reported mixed results in relation to mid clause pausing, with individual learner variation offered as a possible confounding variable.

In a more recent effort to address this continuing swiftly increasing number of and use of myriad measures of oral fluency Segalowitz, et al., (2017) sought to address the issue by conducting a logical and systematic review of the most frequently utilised operationalisations of 'utterance fluency'. The importance of defining what we mean by fluency, and in which specific context, is a necessary first step essential to any evaluation of fluency measures, as problems are not only created by *how* we measure fluency, but also *what* we measure and what we consider fluency to *be*. *Utterance fluency* (Segalowitz, 2010) is used to clearly define what is meant by fluency in their study; "the temporal and hesitation phenomena that characterize the fluidity of speech delivery" (Segalowitz, et al., 2017, p. 92). Speech fluency in the study then is the narrower view of the fluidity of speech, this is as opposed to the broader understanding of fluency which can include, but is not limited to: speech proficiency, size of vocabulary, depth of vocabulary, syntactic knowledge, formulaic language, etc. This narrower and more clearly defined and focused view of fluency allows for a more accurate assessment of what is to be measured, as well as how best to measure it. A corpus-based study was employed by Segalowitz et al., (2017) to evaluate speech measures used in previous studies, and to ascertain which the better indicators of speech fluidity were i.e., fluency, and which could be used to show gains in learner fluency over time. An original approach was taken in that the study first identified the measures of oral fluency operationally and logically, before moving on to explore where improvements and developments in fluency were made through the practical application of those previously identified measures. This process avoids the ‘circularity’ of many studies, and removes the over reliance on those fluency measures which show improvement in a given study (Segalowitz et al., 2017). This unique approach utilised to define measures of fluency; essentially operationalizing L2 oral fluency separate from, and before the analysis of any fluency gains, allowed for the creation of a more theoretically robust and logical study and one which could be more readily replicated across multiple contexts and learning environments. The study reported that mean length of run, defined by Segalowitz et al., as the number of syllables between silent pauses, and the mean length of silent pauses were those features which best represented L2 fluency when investigating fluency development or other fluency issues.

In summary much research has been conducted into the various measures of speed fluency, with findings suggesting that there are several measures which are largely agreed upon and reasonably accurate reflectors of cognitive fluency among L2 speakers. The role of pause location is still a relatively new area of study, and one which is still seeing development in both the way in which it is defined and implemented in research.

However, there is of course still much room for improvement in understanding and accurately measuring L2 oral fluency. More focused and specific measures of speed fluency in particular are required if cognitive fluency is to be more accurately represented and better understood. Perhaps the most obvious place to start would be to measure the performance of a particular speaker transacting a given task in their first language, and then compare their performance with a similar task completed in a second language. If the measures of speed fluency shown to be more reliable in the literature, namely; phonation time ratio, mean length of run, and speech rate (Kormos, 2010; Bosker et al., 2013) were compared it may help to reduce the variations in fluency measures inherent with individual speaking and pausing styles. This comparison may enable a more nuanced exploration of the differences in speech production in general, and more specifically the attendant mental processes involved in producing more fluent speech in a second language.

In relation to the current study, Skehan et al., (2016) believe that their new perspective on speed fluency; the contrast of measures at a ‘clause internal’ and ‘clause external’ position, will allow a more accurate view of cognitive fluency. The next section provides a detailed analysis of Skehan et al., (2016) and how it may enable a new perspective on L2 oral fluency.

2.13 Clause and Discourse Fluency (Skehan et al., 2016)

Recent attempts to better explain oral fluency and its impact on SLA have focused on varying different aspects and definitions of fluency. One such example is the recent article by Skehan et al. (2016) who seek to better understand fluency by drawing a distinction between fluency at the ‘clause’ level and the ‘discourse’ level. Skehan et al. (2016) define fluency, based on Levelt’s model (1999) and explanation of speech production as comprising four conditions: *knowing what you want to say, having the means to say it, not changing your mind,* and *anticipating problems effectively*. This new definition is explored in comparison to Levelt’s updated model of speaking (1999) to better locate the theory and to provide a frame of reference; with the first stage being conceptualization. Conceptualization for Skehan et al. (2016), involves *knowing what you want to say* i.e., selecting and organizing concepts into the pre-verbal message. The *knowing what you want to say* stage creates the input for the second stage which is *having the means to say it*; Levelt’s formulation stage, where lemmas are retrieved from the mental lexicon to create the phonological and syntactic framework that are then converted into speech in the final stage; articulation. If these stages operate together smoothly, without error, then ideas are translated into spoken language resulting in fluent speech (Skehan et al., 2016). The third stage of *not changing your mind* is concerned with making decisive linguistic choices in real time during the entire speech process while it is being reviewed; this is Levelt’s monitoring stage. Nonverbal signals from the listener (a puzzled facial expression, a shake of the head, etc.) may force the speaker to reformulate or clarify their utterance, resulting in disruption of fluency. The final stage Skehan et al., (2016) detail is *anticipating problems effectively* and it is also concerned with the maintenance and monitoring of fluency, and involves correct lemma retrieval in the formulator stage of Levelt’s models of speech production (1989, 1999).

Skehan et al. (2016) argue that second language speaker's fluency; or rather dysfluency, can be better explored through the distinction between breakdowns in fluency at the clause/AS level and breakdowns at the discourse level. A clause or AS-unit here is defined by Skehan et al., (2016) as 'an independent clause, or sub-clausal unit, together with any dependent clauses associated with either' (Foster et al., 2000). The term 'sentence' is not used in the definition as it is fails to capture the fragmentary nature of spoken language produced by L2 speakers (Skehan et al., 2016). It is proposed that mental processing at the clause/AS level takes place within a clause and is concerned with the formulator and articulator stages of production, processing is focused here on lexical choices, *having the means to say it*, including lemma selection and is when micro-planning occurs. Processing at the discourse level takes place above the clause/AS level and is governed by the conceptualizer stage; *knowing what you want to stay*, dealing with problems arising in constructing a group of units and organizing them into a coherent argument and is evidence of macro-planning.

It is this distinction Skehan et al. (2016) draws between macroplanning and microplanning (explained in more detail on page 104) which they believe helps to explain the differences between L1 speakers, and their increased levels of spoken fluency, and L2 speakers and their lack of fluency. Every speaker pauses during their speech and patterns of speech obviously vary greatly between speakers, both for L1 and L2; however, Skehan et al, believe that L2 speakers will pause more for micro-planning. In other words; the second language learner's lack of spoken fluency can be evidenced by their greater number of pauses and breakdowns in speech at the clause/AS level. This lack of fluency is put down to L2 speakers drawing upon a mental lexicon which takes longer to access, is smaller, less well organised, and contains less formulaic language (Bolibaugh & Foster, 2013). Parallel processing as detailed before is much harder for L2 speakers due to the increased demands on attentional resources at the formulator stage (Kormos, 2006). This necessitates a switch to serial processing, and results in problems at each stage needing to be solved before it is possible to move on to the next stage of speech production (Skehan, 2014).

While Skehan et al., (2016) offer an interesting new perspective on measures of dysfluency, and how they can be linked to speech production problems with their interpretation are apparent. The most obvious being the repositioning of microplanning from the conceptualiser to the formulator. In Levelt's model (1989), both microplanning and macroplanning are said to take place in the conceptualisation stage of speech production. Levelt's model is the basis for much research into speech production and is a commonly agreed upon starting point for many theories in L2 speech and SLA. Skehan et al., (2016) offer little explanation for the repositioning and it is unclear how they justify this reinterpretation. While it is true that microplanning relates to lower order process, it has not typically been tied to involvement in lemma selection or the syntactic planning of utterances. It does appear that the notion may have been retroactively applied, with the results supporting the idea those pauses in the mid clause position are related to microplanning and 'clause level decisions'.

A further problematic element with Skehan et al., (2016) is the lack of a detailed explanation in their description of the final stage in 'anticipating problems effectively'. It is not clear if this refers to problems in regards to formulation, or with problems with the discourse message. The problems mentioned may also refer to problems inherent in making lexical choices or with difficulties overcoming pragmatic gaps i.e., strategic competence (Canale & Swain, 2017). The distinction here is an important one as it directly relates to the breakdowns in L2 oral fluency, and the location of the pause is said to be linked to the specific mental process employed at the time of the breakdown. This could be better explained if framed within a view of the distinction between linguistic knowledge and linguistic skills, an area of importance in the distinction of L2 oral fluency, and is discussed in greater detail below in section 2.14.3.

While the suggestions put forward by Skehan et al., (2016) are certainly intriguing ones, and offer possibly significant advances in the understanding of L2 speech processing and the role of pauses in measuring oral fluency, the ideas at present lack a robust and comprehensive theoretical justification, and are requiring of more investigation. Something which the present study seeks to address.

2.14 Known Issues with L2 Oral Fluency

As the suggestions and analysis of Skehan et al., (2016) exemplifies there are several issues and conflicts which remain in the area of L2 fluency, a discussion of the most relevant to this study follows.

2.14.1 Difficulties Measuring Oral Fluency

Perhaps the most obvious difficulty with measuring fluency, as detailed above, is the fact that so many different measures have been used to assess fluency in research (Ellis, 2009; Segalowitz, 2010). This ultimately makes the comparison of results from these studies problematic, as the measures used are often employed in quite significantly different ways. Some measures employed in research studies are given the same name, for example ‘speech rate’, but refer to something different or are employed and calculated in a subtly different way, leading to confusion and difficulty of comparison. This underlines the importance of providing highly detailed and easily replicable instructions in research, to allow for greater consistency of both implementation and calculation. Some of these measures utilised are also of questionable reliability and validity, with many researchers trying to find the ‘best’ measure, and as a result studies are constantly introducing new measures which further complicates the issue (Michel, 2017). This introduction of a novel measure of oral fluency is something which the present study could fairly be accused of, but it is hoped that the implementation of the distinction in measures employed by Skehan et al., (2016) will allow for greater understanding of which measures of fluency reflect which aspects of speech production in an L2.

A further issue with measuring fluency is concerned with the increasingly common use of technology to analyse and examine elements of fluency. This can make comparisons of older studies problematic as they no longer share a similar methodology, but as Segalowitz (2010) points out the use of software, such as PRAAT (Boersma & Weenink, 2005) to automatically analyse fluency makes the process much more objective and precise, with the added bonus that much larger data sets can be processed due to the increased efficiency of the software. The software can be utilised to automatically detect silences in speech samples and to count syllables. This is something which would have previously been done by hand, and as a result would have been far more time consuming and less reliable. The software is an especially good fit for analysing speed fluency because of the exacting temporal nature of utterance fluency measures, and the accuracy which the new technology can now provide. One other area of concern with the automatic analysis is that it does require very clear speech samples; with even just a little background noise or slight a variation in the volume of a participant's voice causing significant problems for the automatic detection of both pauses and syllables.

One final concern for the accurate measurement of L2 oral fluency is the extent to which other factors of speech performance; accuracy and complexity, interact or affect each other. Investigations into the impacts of CAF and the 'trade-off' between them, details how CAF compete for attentional resources (Skehan, 1998). A focus on accuracy and complexity, or *form*, is often sacrificed for an improvement in fluency, or *meaning* (Skehan, 1998). Michel (2017) also adds that if fluency is studied separately then we are failing to capture the whole picture of oral performance, and the way in which CAF factors interact with each other. The call for standardisation of the theory and constructs which are used to investigate fluency is at odds with the knowledge that there is a large amount of variation and interaction between CAF in L2 speech (Michel, 2017).

2.14.2 Formulaic Language and Creative Language

To what extent oral fluency relies upon the employment and implementation of formulaic language, as opposed to creative language use, is a contested area of research in speech production, and one with obvious relevance to the current research project. The previous section on formulaic language (2.6), explained how it is an area of significant difference between L1 and L2 speech production, and can help ease production in the conceptualiser and formulator stages. The following section will detail issues surrounding formulaic language and discuss the implications for oral fluency in relation to creative language use.

Firstly, to clearly define which language is classified as formulaic can be problematic, Wray (2002, p. 9) offers a description of formulaic language as "A sequence, continuous or discontinuous, or words or other elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar". This definition could be viewed as rather vague, but confronts the complexity of clearly defining formulaic language. The definition is founded in the constructs of pattern-based language systems, and connectionist theories of SLA (Ellis 1998; 2003). Language is viewed here as a system of complex data embedded in society. This data is represented and stored in chunks, and when a language learner becomes more proficient, and has greater exposure to the language, the data patterns and chunks are reinforced and solidified in the developing language system (Wray, 2002).

The extent to which theses chunks of preconstructed language can be identified, and said to be separate and distinct from those utterances generated through the application of grammatical rules is also problematic. Myles, et al., (1999) question if it is even possible to determine if a particular construct has been retrieved by the learner as an unanalysed whole, or whether it has been derived creatively from a rule, or if it is indeed some combination of the two processes working simultaneously. However, there is compelling evidence to support the idea that the ability to deploy formulaic sequences, from long term memory stores, bypasses the need to compose them online through word selection and grammatical sequencing in capacity limited working memory (Conklin & Schmitt, 2012). This is achieved through the easy retrieval of these formulaic sequences, and does not necessitate the online production through grammatical sequencing or word selection (Pawley & Syder, 1983). When speaking in an L1, it is possible to ease demands on the cognitive load through formulaic language, however, the literature so far provides little conclusive evidence to prove to what extent L2 speech production benefits from the implementation of formulaic speech. Whether or not any but the most fluent L2 speaker can employ, or even has the opportunity to acquire formulaic language, as effectively as an L1 speaker is the subject of debate. An L2 learner, in a foreign language setting with limited exposure to the target language (2-3 hours a week) would have little opportunity to develop real language use, and thus the use of formulaic language would not appear to significantly impact leaner performance in terms of fluency (Mora, 2006).

It may be problematic for researchers to provide clear empirical evidence of L2 speakers using formulaic language, however anecdotally language teachers will attest to hearing lots of chunks of preconstructed language, deployed by even the lowest proficiency levels of learner. A ubiquitous example for all levels of language learners in South Korea is the response: “Fine, thanks and you?” when asked how they are doing. It is clear that this response is produced without much on-line cognitive processing, and rises almost unbidden, and subconsciously to the speaker lips. This response is delivered smoothly and rapidly, and meets the criteria discussed above for fluent speech.

The ability to deploy any number of preconstructed chunks of language would certainly appear to ease the cognitive load in formulation during speech production, and allow for the production of more fluent speech. This conclusion might lead us to question whether or not well-remembered linguistic chunks, deployed in speech as formulaic language, to swiftly communicate in real time, is evidence of ‘real’ fluency. Can an L2 speaker be said to be more or less fluent in a language if they are using these short cuts, by avoiding the cognitively resource heavy necessity of complex grammatical encoding and careful lemma retrieval from the mental lexicon? We may view fluency as either; the ability to recall and quickly deploy chunks of preformed language, or to create and construct original, grammatically correct clauses, or rather some combination of the two. Whichever view is taken, it is clear that formulaic language can play an important role in the cognitive processing in L2 oral fluency, and is one which learners, teachers, and researchers should be aware of. Ellis (1996) states that language learning is essentially the learning of linguistic sequences, and it is clear that deploying higher quantities of formulaic language can result in more fluent speech.

In relation to the current research project and the investigation of task design, it might be possible to see the influence of formulaic language deployed in response to the varying task input formats. With a written input task there are specific language elements embedded in the task which the participant must employ in task completion, will this encourage the deployment of formulaic language by the participants, or will the picture task encourage participants to employ chunks or formulaic language to meet the demands of task completion, as no linguistic constraints are placed on them, they may choose whichever utterances they think are best. An increase in task complexity may result in less fluency, as the increased cognitive load increases on L2 speech production, however it is possible that the deployment of language chunks, and formulaic language may ease the processing load and allow for the production of more fluent speech, with less pauses occurring at the mid clause position.

2.14.3 Linguistic Knowledge and Processing Skills

A further area of contention in the development and understanding of L2 oral fluency, is the extent to which fluency occurs either as; a result of knowledge of the language and its structures, or the ability to process and use communicative skills to achieve the desired goal. This is reflected in the section detailing definitions of fluency (2.7), with Lennon's (2000) definition of 'online processing' and the importance of accuracy, while Skehan and Foster (1997) place more importance on achieving the communicative goal. The difference in definitions also echoes the distinction between the 'narrow' and 'board' views of fluency (Kormos, 2010). In other words, fluency can be said to be the result of *knowing* a language, or having the ability to *use* a language. This is of course an oversimplification, and research has provided compelling evidence for both of these dimensions (Dörnyei & Thurell, 1991; Hilton, 2008), as well as for some combination of the two (Hilton, 2008). However, the exploration of the importance of knowledge and processing in the production of fluent speech does help highlight the fact that there is more than one way to understand L2 fluency, and more than one way to produce fluent language.

This distinction has obvious relevance for SLA, TBLT, and task design. To encourage and promote learners using more fluent language in their L2, should tasks be designed which encourage learners to deploy communicative skills? such as those supported by communicative language teaching approaches, including task based language teaching (see section 3.2 for description). Or are those tasks which provide a focus on improving and activating language processing more suited to the production of fluent speech? Those more traditional language teaching methods which provide a focus on forms, i.e., presentation, practice, production (PPP) and audiolingualism. If the former view is held then a communicative-based methodology may be implemented, to help less proficient learners overcome losses in fluency by encouraging the use of communicative strategies such as: circumlocution, word avoidance, word coinage, etc. It has been argued that the underdevelopment of strategic competence is a large part of the reason for a lack of oral fluency and conversational skills (Dörnyei & Thurrell, 1991). The use of communicative strategies can certainly help learners produce more fluent language through the avoidance of problematic terms and missing lexical items, which invariably lead to break downs and pauses in the flow of speech. However, the advantages of a well-developed knowledge of the target language, and the ability to select the correct term efficiently and easily, is argued by Hilton (2008), to have a far greater positive impact on oral fluency. The advantages of a well-developed mental lexicon, with readily available and accessible language, provides a great benefit to the on-line mental processing in speech production (Holton, 2008).

This issue can be related to the distinction between two differing teaching approaches. Firstly, those that provide a *focus on forms*; which is the more traditional approach to language teaching and uses explicit instructions to teach grammatical forms specifically and explicitly, authenticity and communicative competency are largely ancillary, and are often ignored completely (Long, 2015). This is contrasted with a *focus on form* however, which emphasises the form meaning connection and teaches grammar located within contexts and through the performance of communicative tasks (Ellis, 2005). These two different approaches to language learning can obviously be related to the type of fluency development they support, with a focus on forms lending itself towards a knowledge based route toward fluency, and a focus on form favouring a skill based approach.

Communicative strategies employed in a focus on from then may be useful for less proficient learners, or for encouraging language development, and aid in the *use* of language, but the advantages to *knowing* the language, with a focus on forms, are also important in the production of oral fluency, especially in relation to the mental processing required. Communicative strategies can be resource intensive and require conscious attention to perform, choosing a word you know from a well-developed mental lexicon is far less taxing (Hilton, 2008). Comparisons are evident between L1 and L2 speech production as seen in the explanations provided by Segalowitz (2010) and Kormos (2006) in section 2.5.1. L1 speech is far less cognitively demanding at the formulation stage due to the automaticity of language produced, however this is not the case for all but the most proficient L2 speakers. A rich and well developed mental lexicon and knowledge of grammar structures will help to reduce the cognitive demands of formulation, and should result in the production of higher levels of oral fluency.

In relation to the current research project, which is located in a South Korean educational context, we are met with evidence of learners who have relatively high levels of linguistic knowledge, but comparatively low levels of oral fluency. This provides anecdotal evidence that knowledge of a language and the linguistic structures is not solely enough to promote fluent speech. Speaking fluently in an L2 is obviously a skill that needs practice as well as knowledge, how much oral fluency is about processing skill and how much is it about linguistic knowledge is a relevant question, and one which is still requiring of further research. In a study on the impact of linguistic knowledge and processing skills on oral fluency De Jong et al., (2013) found that oral fluency measures were found to correlate with both linguistic knowledge and processing skills.

The production of fluent L2 speech then is related to and requiring of both linguistic knowledge and linguistic processing skills. We may be able to conclude, that to achieve higher levels of L2 fluency a balanced approach is required; necessitating both a detailed knowledge of linguistic structures, as well as sufficient and focused practice processing in the language. To that end, it is important to discern which elements of task design are related to which specific components of speech production so that teachers, learners, and researchers can better understand and achieve L2 oral fluency.

2.15 Summary of L2 Fluency

Throughout this chapter I have attempted to provide an overview of the major ideas which govern the research, and underpin the understanding, of speech production and oral fluency in L2 speakers. To briefly summarise; speech production is a complex and difficult construct to measure, but most of what we know is related to Levelt’s (1989) model of speaking. Models of L2 speech production; Kormos (2006), as well as Segalowitz (2010), have adapted Levelt’s model specifically for L2 speech (and fluency in Segalowitz’s case). Segalowitz’s (2010) three dimensions of fluency; cognitive fluency, perceived fluency, and utterance fluency explain how L2 fluency is affected by the mental processing which takes place during speech production. ‘Automaticity’, where linguistic rules are more readily accessed, in L2 speakers takes place through the procedure of moving knowledge from declarative to procedural, as it becomes practiced and habituated, and ultimately automatised. The accurate measurement of fluency in L2 speakers is difficult and an area of some disagreement, but fluency can be better understood with the aid of the CAF framework and the place fluency has within it. Fluency measures have typically been divided into one of three broad groups; speed, repair, and breakdown fluency. Several studies have employed various different measures to assess oral fluency in L2 speakers, with varying approaches and results. Speed fluency has been found to employ some of the most reliable and valid indicators of L2 oral fluency through research. Skehan et al., (2016) believe a distinction between fluency measures at the clause level and the discourse level can enable a more accurate view of L2 oral fluency. They believe that speech in an L2 will see more pauses than in an L1 at the microplanning stage, which they link to formulation, due to the lack of mental processing capacity available, and the inability to automatize speech production as effectively as in an L1. There are still issues identifying the most accurate measures of L2 cognitive fluency, but the perspective put forward by Skehan et al., (2016) on measuring fluency may offer insight into the cognitive processes active during speech production. This new perspective could provide a more detailed understanding of how any why L1 and L2 oral fluency, and speech production in general differs. This was then discussed in terms of formulaic and creative knowledge use, as well as the distinction between fluency derived from knowledge and from communicative skills. These concepts were then explained in terms of teaching approaches and the distinction between a focus on forms, and a focus on form.

3 Tasks, Task Complexity, and Task Format

3.1 Introduction

In this chapter I will first explore what a task-based approach to language teaching (TBLT) is and how it fits into a communicative approach to second language learning. This is followed by a definition of what precisely constitutes a task with a discussion of the definition of tasks themselves, as well as an exploration of how the tasks have been used in research to date. The chapter will next detail the framework employed in TBLT, with a look at work by Willis (1996) and the framework he provides. The chapter continues with an examination of the variation in task design and input, with reference to relevant studies, and how this can impact the effectiveness of tasks. The focus will then move on to the issues surrounding the extent to which TBLT has been implemented in classrooms, with an explanation of a task supported approach to task based language learning. The following section will outline some of the limitations and known issues which have arisen with TBLT and its implementation in a variety of learning environments, with an exploration of the research conducted into a TBLT approach to language learning. The next segment will introduce the concept of task complexity, with a description of the two major theories proposed by Skehan (1998, 2009), and Robinson (2001, 2007) regarding the cognitive complexity of tasks and how it has been operationalised. The chapter will next provide analysis and a discussion of the research conducted so far into task complexity and its implications, with a particular focus on the relevance towards the present study and the focus on speech production and spoken fluency. The following sections detail the different ways to view and produce fluency. The chapter concludes with an examination of recent work by Skehan et al. (2016) on how differentiating between the location of pauses, as either within a clause or at the boundary of a clause, may allow for a better understanding of the impacts of task cognitive complexity on L2 oral fluency and what this can tell us about the cognitive functions of speakers as they produce L2 speech.

3.2 Task-Based Language Teaching

Task based language learning (TBLT) has been derived through the implementation of a communicative language teaching (CLT) approach, which strives to provide opportunities for language learners to improve their language abilities through ‘real communication’ in the target language (Ellis, 2003), that which Tavakoli and Hunter (2018) refer to as ‘very broad fluency’. Communicative approaches to teaching language and TBLT were introduced as a response to more traditional methods of language instruction such as PPP (presentation, practice, production), grammar translation, and audiolingualism. These traditional approaches to language teaching focused primarily on the study of language structures and a *focus on forms* (detailed in Knowledge and Skills section 2.14.3), and are often considered to be lacking in the ability to provide learners with practical real-world communicative skills. Indeed, these teaching methodologies were largely concerned with literature and frequently focused on teaching reading and writing skills, in the case of the grammar translation methodology, or were heavily involved in rote learning and the memorisation of words and phrases, as seen in audiolingualism. Both methodologies provide little emphasis or opportunity for *real* and authentic communication; where learners express their own ideas to achieve a desired communicative goal in the target language. CLT approaches to language teaching were implemented to address this gap, and sought to *focus on form* (2.14.3) and meaning, specifically the ability of language learners to achieve a communicative goal in the target language. Approaches to CLT are often been divided into *strong* and *weak* forms (Ellis, 2003). In CLT a ‘weaker’ version would primarily focus on language functions to teach the target language, while the ‘stronger’ version of CLT affords learners a communicative approach to language acquisition through interaction in the target language. TBLT is an example of a strong version of CLT and uses tasks as the primary focus to facilitate second language acquisition through communication. The strongest proponents of TBLT believe that the tasks themselves are entirely sufficient for language learning to take place (Ellis, 2003). In other words, the tasks are said to contain sufficient input, provide ample communicative practice, and facilitate the production of sufficient output to learn a language. However, this belief is not universal and there are a number of different understandings and implementations of TBLT that are employed to help develop language learners L2 competency (Ellis, 2003). The following section will provide a brief overview of the main concepts informing TBLT and then move on to an explanation of the tasks themselves.

The TBLT approach to language learning, and the pedagogic tasks they employ to improve performance in an L2, have been the focus of considerable amounts of study in recent years (Ellis, 2009). According to Long (2015) a TBLT approach has to satisfy many facets: it must be psycholinguistically plausible; based on research into how second and foreign languages are learned, it should be grounded in the philosophy of education, it must be accountable, relevant, avoid apparent problems with current approaches, and be learner-centred, as well as functional. A long list of requirements, but one of the great benefits of TBLT is that it is flexible and adaptable to many different teaching and learning environments. Through the implementation of a simple needs analysis tasks can be chosen which are directly relevant to the learners, and as a result a specific syllabus can be designed with unique learner goals, as well as their past and present communicative needs, in mind (Long, 2015).

3.3 Definition of Tasks

So far, I have explained what a TBLT approach is, but the following section will move on to detail what a task actually is, and how they can be identified and defined. At the centre of any implementation of a TBLT approach to L2 learning is the accurate selection of the appropriate tasks for the specific learners (Skehan, 2006; Skehan & Foster, 1999). This has resulted in an increased interest in what exactly makes a task, as well as a proliferation of definitions as to what constitutes a pedagogic task. Tasks are described as activities where specific target language is employed to meet a communicative need and achieve an outcome (Willis, 1996). Tasks are similarly defined by Skehan (1998, p., 95) as any "activity which: places meaning as primary, contains a communication problem, has some comparable real-world activity, places the completion of the task as a priority, and is assessed in terms of outcome". Bygate et al.'s (2001) definition follows a compatible description stating; tasks are those activities which require the use of language to complete an objective and contain an emphasis on meaning. Most research agrees then that a task can be defined as: something real and authentic, which allows learners to use their L2 to communicate with each other, and through some negotiation or interaction to meet and aim or achieve a purpose or goal. This remains a quite broad definition, but the importance of the communication element cannot be understated and is integral to the understanding of tasks (Skehan, 1998; Willis, 1996;

Bygate et al., 2001). The communicative nature of tasks is one which has grown out of the fact that SLA is best served by what Skehan referred to above as a *real-world* activity. This authentic communication is based on the theory that the primary focus should not be on linguistic accuracy, but rather should be on communicating a message in similar conditions to those which exist in real-life situations (Ellis, 2003). The importance of the authenticity of tasks is paramount to TBLT and tasks should not be created simply to enable the practice of specific language forms or structures, rather they should reflect the learner's own meaning (Willis, 1996).

To better define what makes a task a 'task' and not an 'activity' or an 'exercise' Breen (1989) draws a distinction between the 'task-as-workplan' and the 'task-as-process'. The task-as-workplan consists of the teaching materials that constitute the task; input comprised of instructions and a rubric specifying the intended learner outcomes and what they are being asked to achieve (Ellis, 2017b). In an example task described here called the 'Colonist Selection Task', information is provided about several candidates who have applied to form part of a colony on a distant planet, and the learners must decide who the most viable candidates are. The task-as-process is the actual activity resulting from the input. In the above example it would involve the learners in groups discussing the information about each of the candidates and evaluating their suitability, reaching a consensus and making a decision on who should be selected, and then finally giving the reasons for their choices. In line with the previous definitions explored above recent work by Ellis and Shintani (2013) defines a workplan as constituting a task when: it focuses primarily on meaning, has some form of gap (the need to convey information, to express an opinion, or to infer meaning), requires learners to rely on their own resources, and has a clearly defined outcome or goal. According to this criterion we can see that the Colonist Selection Task is clearly a task. An activity or exercise which asks learners to focus on forms, such as one which employs blank filling pre prepared sentences, does not satisfy these criteria. The primary focus is on language structures and forms, not communication, there is no gap to bridge, the learner has most of the information supplied already and has no need to draw on their own resources, and the completion of the task has no end product or goal, other than the completion of the exercise (Ellis, 2017b). Some further examples that could be identified as pedagogic tasks are students working together to solve a puzzle or a problem, sorting or ranking things, telling a story, students negotiating the best advice to give on an ethical or legal problem, and having a student run a particular errand for them as they are unable to do it e.g., going to a friend's home to feed their cat or turn off their oven, etc. (Skehan, 1998; Willis & Willis, 2013).

To briefly summarise, tasks are primarily communicative in focus and see learners deal with multiple aspects of language at the same time, they involve using language to convey information in a way which language is normally used, engaging language use to achieve a non-linguistic goal, while meeting a linguistic challenge, with the overall aim of promoting language learning (Samuda & Bygate, 2008). The *non-linguistic goal* of Samuda and Bygate's (2008) definition refers to a task's goal not being language focused, but rather necessitating the use of language for its successful completion. It should be noted that this is a general definition and is representative of a broad consensus, but it is true that many researchers still disagree to a certain extent as to what constitutes the make-up of tasks. Willis and Willis (2013) for example also include the requirement for a task to be relatable to a real work activity, matching a real life task which is authentic to the given situation (Ellis, 2017). This would obviously be quite demanding in many teaching environments, which may lack the time, resources, and knowledge required to meet these exacting criteria.

Differing approaches to TBLT, and how tasks have been implemented in the classroom, have used a variety of teaching frameworks and approaches, they include: Prabhu (1987), Skehan (1996), Willis (1996), and Lee (2000), however they all have the same basic structure and follow a similar pattern using three 'principal phases' (Ellis 2003). The three phases approach is by far the most commonly adopted, by both TBLT instructors and researchers, as it is practical to implement in the classroom and reasonably straightforward to understand; involving a pre-task, during task, and post task element. TBLT is of particular relevance to teaching and researching oral fluency as it requires learners to actually speak during both the task and the post-task phases, therefore necessitating the use and practice of oral communication in the target language. The following section will detail the Willis' three task phases in his (1996) framework.

3.4 Three Task Phases; Willis' (1996) Framework

Willis' (1996) TBLT framework is based on three distinct stages, and it emphasises the importance of the sequence of these phases describing how this is necessary for the success of TBLT. Willis' (1996) three stages are: *Pre-task*; introducing the topic and preparation, *Task Cycle*; performing the task including planning and reporting, and *Language Focus*; language analysis and practice.

3.4.1 Pre-Task Stage

The pre-task phase concerns the preparation of learners; making sure they understand the learning objectives, outcomes, requirements, and whether or not time is allocated for planning the performance of the task. The purpose of the pre-task stage in Willis' (1996) framework is the activation of the learners' prior knowledge and linguistic resources. It is here in the pre-task phase where the focus of a task is established. As learner's attentional capacity is limited, the task can focus on either the general communicative skills required to complete the task, or an emphasis can be placed on the language forms used (Skehan, 1998). It is in this initial phase of task design that teachers help to "create the conditions that will make tasks work for acquisition" (Ellis 2003, p. 249). The amount of time allotted for pre task planning and preparation is an important component of any task, and the impacts this planning time has on task performance as well as target language acquisition has been the subject of considerable research. Several studies have found that allowing students more time to plan and prepare for a task has enabled a greater focus on accuracy in task performance (Yuan & Ellis, 2003; Ortega, 1999; Wigglesworth, 1997), and reducing the time available to plan pre-task has been linked to an increase in a focus on fluency (Long, 2015). These results and the conclusions drawn have not always been replicated across different teaching environments, and several studies have reported contradictory results (Awwad & Tavakoli, 2019; Iwashita et al., 2008). What is clear however, is that the amount of time allocated for pre-task planning is an important component in the composition of a task, and can have considerable implications for the performance and output of a task (Long, 2015).

3.4.2 Task-Cycle Stage

The next part of the task is the task-cycle which takes place during the task, this is where learners focus on undertaking or performing the task itself in groups, pairs, or individually. Students are encouraged to work with whatever linguistic knowledge they have to complete the tasks and interact with each other, while the teacher observes and only provides assistance if necessary. The idea here is that learners collectively negotiate what they want to say and how they want to say it and overcome any difficulties communicatively as a group (Ellis, 2003). This stage allows for a wide range of methodological options for instruction. As with the pre-task phase time limits for task completion can be employed; with many studies investigating the impacts of time limitations on task performance as evidenced by the variations recorded in language output and performance in regards to the CAF framework.

A key variable which can impact fluency in task performance, and one which is of specific interest to this study, is the provision of materials that learners have used to prepare for the task during the task itself. Examples may include maps, timetables, pictures, adverts, etc. Allowing the learners to use these materials can benefit the production of oral fluency in a number of obvious ways; by allowing increased familiarity with the topic, easing performative anxiety, reducing reliance on memory. However, access to materials during the task can also help by 'priming'; a phenomenon where prior exposure to language influences subsequent language processing in recall and speech production (McDonough & Trofimovich, 2009).

In relation to the effects of priming and the current study, it is important to point out that research has shown priming may occur with a variety of task designs and inputs, and is not specifically linked to the provision of linguistic items (Segalowitz et al., 2011). However, it may be that linguistic content in the form of sentences, words, or bullet points provides learners with greater levels of priming, due to the easing of lemma selection with linguistic elements provided. However, when compared to images or pictures, which are devoid of linguistic content, the impact of priming may not be as substantial. Regardless of this distinction in task design, if a task input provides some form of priming then we may expect to see an improvement in L2 fluency. Priming acts to facilitate the recollection, or ease of access to vocabulary or linguistic structures, but also spread activation by linking to similar concepts, to further facilitate recall (Trofimovich & McDonough, 2011). Priming can be linked to a reduction in the cognitive load, by facilitating grammatical and phonological encoding in the formulator, and may well improve performance (McNamara, 2005). It may be expected then that priming will lead to a reduction in the frequency of pauses at the mid clause, due to easing of the cognitive load and enable the production of faster and smoother L2 speech.

Another area where fluency can be impacted by task design, with the provision of linguistic material, is through ‘pushed output’. This is a phenomenon where by language learners are pushed, or required, to use input language from the tasks to form comprehensible output in task transaction (Ellis, 2005). The notion is grounded in Swain’s (1985) Output Hypothesis, which assert that producing language is a key part of the process of second language acquisition. As learners attempt to speak in the target language, they become aware, or notice, that they lack the resources to linguistically encode and express their desired meaning (Pannell et al., 2017). Swain (2005) describes how producing language may act as a trigger for ‘noticing’, and this directs learner attention to the language or expressions required to meet their communicative goal. In relation to speech production, pushed output in the task design may facilitate the accessing of linguistic information by requiring language necessary for the completion of the task. However, with a variety of input formats; written, and pictorial, tasks may not provide the same amount of support for the cognitive process in speech production. Also, it remains to be seen if requiring participants to use specific language will help them in speech formulation, and result in more fluency, or constrain them and lead to reductions in fluency.

The design and input format of a task can also facilitate the ‘borrowing’ of target language vocabulary and structures. This is where a learner utilises a new verbal formulation provided by the task input to express their own meaning, rather than simply relying on language they already know and have used before (Prabhu, 1987). For example, a learner may use vocabulary they have not used before and are unsure of, but may employ it during the task performance phase as it is presented in the input as they can be sure of its correct use. This is obviously an excellent opportunity for learners to acquire new linguistic elements and reinforce their learning through implementation, and is one of the strengths of a TBLT approach. An unplanned element may also be introduced to the task in the performance phase, although some studies (Foster & Skehan, 1997) have shown that this does not have a significant effect on learner performance. However, it may serve to extend student talk time and increase their intrinsic interest and participation in the task (Ellis, 2003), an important aspect in classroom based task design.

Motivation and interest are important factors in task design, and to maintain them both requires that tasks should be relevant to a specific learner, or group of learner's actual language learning needs. Intrinsic interest can be defined as motivation which occurs when a task is inherently enjoyable or interesting (Deci & Ryan, 2000), it is something which the participant actively wants to transact for their own satisfaction. Higher levels of intrinsic motivation have been shown to lead to higher levels of creativity and learning (Deci & Ryan, 2000). The importance of designing and implementing tasks which are specifically tailored to the interests and motivations of the language learner is clear. Links between motivation, oral fluency, and specific components of task design are scarce in research, but are also difficult to extrapolate from, as motivation in notoriously fickle and subject to interference from myriad design factors, as well as learner individual differences. However, anecdotally it may be safe to assume that learners who have higher levels of intrinsic motivation for a specific task will be inclined to be more engaged in the task, and ideally produce more speech.

3.4.3 Post-Task Stage

The task is concluded with the language focus stage, a post-task stage focusing on the language used and following up on task performance. The learners are encouraged to notice new forms of language and practice them after completing the task. It can see the students report back to the class, build self-confidence and enforce new learning, orally or in a written report, with the teacher able to provide feedback on language and performance (Long, 2015). The addition of a public performance element in the final phase has been shown to motivate learners to produce more fluent language (Willis, 1987), as well as providing a comparative element to facilitate the implementation of peer review. The flexibility of the implementation of the three stages in a general TBLT framework enables opportunities to emphasise the focus on any one, or all three of Skehan's (1998) CAF performance factors. The post task phase can take a variety of forms; it may be written, spoken, involve a performative element, or may contain a dialogic or monologic oral aspect. This adaptability of TBLT provides a teacher ample opportunity to tailor a lesson to the desired learning outcomes appropriate for their class, while also maximizing the promotion of language learning (Shehadeh, 2005).

3.5 A Task Supported Approach to TBLT

Since the adoption and implementation of pedagogic tasks a distinction has been drawn between those who argue that tasks should form the basis and foundation of language instruction and those who use tasks to simply supplement language teaching. Long (1985) and Prabhu (1987) were the first to argue that an entire teaching method could be based on tasks and task performance alone. The use of tasks is prevalent in many current language teaching classrooms, but they do not always closely follow the TBLT suggested pedagogic recommendations (Long, 2015). As Van den Branden (2016, p. 123) states: "a task based approach to LT (language teaching), and the use of pedagogic tasks is often integrated in an eclectic manner and usually employs a myriad of varying approaches, often comprising a combination of traditional teaching methods with the implementation of tasks on an ad-hoc basis." This has led to a *task-supported* (Ellis, 2003) approach to language teaching (TSLT), where tasks are only when they are thought necessary or useful and are often combined with more traditional teaching methods.

The distinction between TBLT and TSLT is an important one and helps to provide a more clearly defined outline for what actually comprises a TBLT approach. The most obvious distinction between the two approaches is how they handle attention to form (Ellis, 2017). In TBLT a focus on form may result from a problem with task completion, or when the teacher draws attention to a specific linguistic form. The focus on form then occurs as the learners are transacting the task in the during-task phase, and is secondary to the primary concern of communicating in order to facilitate task completion, which remains the primary focus of TBLT (Ellis, 2017). The focus on form here is as a result of attempting to complete the task. A TSLT approach on the other hand emphasises the focus on form from the outset and attention is directed specifically towards a particular target form in the initial stages of the lesson. To summarise the primary focus of a TBLT lesson is the communicative use of language, whereas in TSLT the focus is on accurate use of the target form (Ellis, 2017).

TSLT it seems then is based more on the notion of skill based learning theories such as Anderson's (2000), than a focus on form typical of TBLT, and claims that learning comes about through the declarative representation of a linguistic form that transforms through practice in to procedural knowledge. This process is accompanied by the move from controlled processing to automatic processing (Ellis, 2017). The importance of the tasks in TSLT is that they provide genuine real world operating conditions (DeKeyser, 1998) to achieve automaticity of processing through the creation of the conditions required to use a target language feature in a communicative task. This process is enabled after the declarative knowledge of the target language structure has been established through the use of explicit instruction (Ellis, 2017). TBLT however, is derived from a more varied group of theories in second language research such as 'usage-based language learning' (Ellis, 2005), but can also be implemented with the 'direct methods' or more general approaches to SLA an suggested by the 'interaction approach' (Gass & Mackey, 2007). These differences in the central tenants of the two approaches are important as they highlight the fact that TBLT does not aim to teach learners declarative knowledge of specific features prior to task performance (Ellis, 2017). Rather the interaction approach, drawing on the theories in Long's interaction hypothesis (1996), believes that learners can "acquire the kind of linguistic knowledge they need to engage in communication" (Ellis, 2017., p. 111). This is achieved through learner interaction making input comprehensible, the employment of feedback on learners' attempts to use target language, and with learners modifying their own output to meet the target usage. TBLT is designed to assist in the 'natural' process of language acquisition, and helps learners to develop and improve the skills they need communicate in the target language.

These two theoretical approaches provide fundamentally different perspectives on how second languages are learned, which is represented by the way in which they are utilised in the classroom. TSLT bears similarities with more traditional language teaching methods such as the 'presentation, practice, production' approach, and provides a teacher centred, top-down approach to learning. Language is deployed in smaller more manageable pieces, before it is passed to learners in the form of easy to understand and employable pieces. TBLT on the other hand provides a more student focused, bottom up approach, with learning facilitated through active participation and the creation of authentic opportunities for language use (Samuda & Bygate, 2008).

3.6 Known Issues With TBLT

There are some known issues and challenges with the successful implementation of a TBLT approach, especially in an East Asian context. Long (2015, p. 6) cautions that "without sufficient emphasis and without learners' conscious focus on form language development may not be balanced". This is especially important when examined from the perspective of a university setting in South Korea, the local of the L2 participants in the present study, as there is often a reluctance to universally adopt a TBLT based approach to LT (Jeon & Hahn, 2006). This reticence often comes from a lack of a clear understanding of how to best implement a TBLT approach, but also arises from uncertainty by teachers as to how tasks facilitate language learning in the classroom (Jeon & Hahn, 2006). The adaptability of TBLT can work against an argument for its implementation, this is often due to the perceived lack of a clearly defined focus on the linguistic elements of a language.

This conceptual uncertainty about a TBLT approach, and with tasks more generally, has affected TBL's implementation in many East Asian countries (Littlewood, 2007). This uncertainty raises questions about the viability of TBLT as a teaching methodology in specific relation to the South Korean context (Jeon & Hahn, 2006). Ellis (2003, p., 252) summarizes the size of the issue perfectly when he says: "task-based teaching calls for the classroom participants to forget where they are and why they are there and to act in the belief that they can learn the language indirectly through communicating in it rather than directly through studying it". Learners may also need to be convinced of more 'experimental' teaching methods and "may not initially see the value of a task" (Ellis 2003, p. 244) or a task based approach to language learning. Some critics (Burrows, 2008) have even argued that TBLT is not suitable for East Asian learners who are accustomed to more traditional classroom learning styles, as it is not what these learners are accustomed to or what they have come to expect from a language class. Language learners in South Korea are often asked to disregard previous teaching styles and methods, many of which they have successfully employed, and instead place their faith in a system they have less experience with and understanding of. As TBLT gains more prominence and becomes more widely implemented in language learning environments in South Korea, it is increasingly coming into conflict with more traditional teaching and evaluation styles employed effectively in the learning and achievement in most other subjects (Jeon & Hahn, 2006).

We can observe that as Korean freshmen are often grouped by major in required English classes, so subject specific tasks may be implemented, greatly improving the relevance, usefulness and learner-centeredness of LT as Long (2015, p., 14) explains "…students respond immediately and positively to materials and teaching that treat them like adults and have clearly been designed to cater specifically to their communicative needs." As an example, if a freshman group of language learners are majoring in Law they may be given a task which involves deciding on the best advice to give to a client or how to most appropriately interpret legal or ethical guidelines, but more importantly the aim of a TBLT system must be to prepare learners for how they will ultimately use English in the future. The matching of students' ultimate language needs with their classroom instruction, the "targeted selection of tasks" Skehan (1998, p., 98), not only makes TBLT more relevant, but can also achieve a better balance between the development of the different areas of proficiency within their L2. The accurate selection, design, and implementation of tasks to create a specifically relevant, appropriate, and intrinsically interesting learning environment is paramount and can lead to greater opportunities for language learning to occur (Foster, 1999).

Willis' (1996) framework has also been criticised for lacking support from empirical research into its effectiveness as a teaching methodology, and has rarely if ever been implemented in its entirety (Samuda & Bygate, 2008). Also, despite the perceived benefits of Willis' (1996) framework limitations are still evident when implementing a task based approach to language teaching in a real classroom setting. The initial task phase may require specialized English discussion and/or written group work skills which may be too difficult for some students, so a simplified or teacher led task phase may be required for less proficient students. Difficulties are also inherent in sustaining the attention of students for the duration of all of the various stages, and this is often exacerbated by large class sizes. These and other factors can negatively impact language learning, primarily through classroom management difficulties with teachers being unable to monitor all of the L2 student interaction, resulting in some learners reverting back to or relying on their L1 (Sakui, 2007). To mitigate this issue, it may help to keep stages shorter and provide more learner feedback to individuals, each group, or even together as a class, this may help keep students involved and create a more comfortable learning environment. It is important to present tasks to students in a motivating manner, especially with learners from more traditional learning backgrounds (Dörnyei, 2001).

It has also been argued that TBLT is unsuitable for low-level learners (Swan, 2005; Bruton, 2002). Less able learners may lack the linguistic skills in the target language to construct meaningful sentences and form a dialog, therefore greatly reducing the ‘communicative’ nature that is central to the TBLT approach to SLA. A difference in learner ability within a group of learners can also result in a higher degree of variation in pupil involvement in task performance (Moser et al., 2012), more proficient students may dominate, while other less proficient speakers are less involved and adopt a more passive role. The passive students are communicating less and have fewer opportunities to practice and communicate, leading to fewer opportunities to practice and acquire the TL.

The long and complex nature of Willis’ framework (1996) often means it is also problematic to implement over a class which lasts longer than an hour, especially for instructors who are new to TBLT or lack the required teaching experience (Ellis, 2009). It is also true that a communicative approach to language learning often fails to take into account ‘why’ students are learning a second language (Moser et al., 2012). Often learners are preparing for a grammar focused test, such as TOEIC, IELTS or TOEFL, and may be demotivated by learning without a focus on grammar. TBLT may not be meeting the fundamental language learning needs of the students, a major failing in any teaching context, and one which TBLT in particular is often implemented to address. In relation to the current study, it is clear that this focus on form may becoming at the expense of specific knowledge of the target language structures and forms. In relation to fluency, it is also interesting to consider the conflict between a focus on form and a focus on language use, which can result in a lack of oral fluency in language learners. As was detailed earlier (2.14.3) a more balanced approach to language teaching, through gaining detailed linguistic knowledge, while at the same time practicing communicative speaking skills may result in the greatest gains in L2 oral fluency.

If we explore TBLT in relation to skill building models of learning (2.9.1.1) and focus on fluency we can see another possible area of conflict. Skill based theories begin with the movement of declarative knowledge to procedural, with declarative knowledge being factual in nature, and involves the explicit learning of knowledge about grammatical rules and linguistic functions (Ellis, 2003). Once this declarative knowledge has become procedural knowledge and 'automatized' the language learner will have the ability to use the language without paying consciousness effort to it. This will result in the production of more fluent speech, but if TBLT does not result in the formation of declarative knowledge then this may impact on the ability to produce fluent speech. This means skill based theories, such as Anderson's (2000) would appear to have more in common with traditional classroom based language learning techniques, such as the presentation practice production (PPP) approach. The focus in PPP classes is on the presentation of explicit linguistic rules and forms, supported by the facilitation of ample practice using them, with the goal being to ultimately enable automatization (Ellis, 2003). It has been argued that TBLT and communicative approaches to language learning do not place enough emphasis on focused practice and the conversion of declarative knowledge to procedural knowledge (Swan, 2005). This uncertainty has implications for TBLT, and task design, and is something the present study seeks to address.

Therefore, possible problems exist with the suitability of the TBLT framework and its implementation in differing teaching contexts. However, with much of this criticism it is important to point out that there is often a fundamental misunderstanding of what exactly a TBLT approach to SLA involves, and the extent to which it can be modified and combined with more traditional teaching approaches (Ellis, 2009). Through the adaptation of Willis' (1996) model these issues with TBLT and others can be addressed through the modification of the framework for a specific language-learning group (Long & Norris, 2009). Adapting the TBLT methodology to better meet student expectations, and to more readily take into account their language learning needs and proficiency, can produce a more effective and successful learning experience. The flexibility and adaptability of TBLT are some of its strengths and make it suitable for most language learning contexts, with careful application and modification for the specific setting. Ellis (2009), Nunan (2006), and Long (2015) all offer detailed explanations of how a TBLT approach can be focused to meet the specific goals of a particular learner group. In the case of fluency, it the goal of the learner is to improve and produce more fluent speech then tasks can be designed and implemented specifically to achieve this. Research needs to explore to what extent is it possible for L2 speakers to produce more fluent speech, through the manipulation of task design and specifically task complexity. A better understanding of how tasks design impacts oral fluency can lead to better learner outcomes for language learners, and it can also help improve our understanding of the cognitive processes at work during language production.

3.7 Tasks in Research

It is important to remember that TBLT is the only 'teaching approach', distinct here from a 'teaching method', which is supported by research into SLA. The impacts of TBLT and its effects on SLA have been explored through considerable research, with tasks being central to much of this work. Tasks have been used as the basis for considerable amounts of research as they are the tools used to facilitate language learning, and as a result they are of interest to not just language teachers, but also to language researchers (Ellis, 2003). The interest in tasks is based on their communicative nature and the way in which they generate language production, which is very similar to how natural language production occurs. This allows researchers to record language extracts produced during tasks and analyse them for insights into how learners acquire a second language, as well as the impact of the myriad factors affecting this. Tasks have been employed in a plethora of different studies into SLA and explored a wide range of varied research perspectives (Robinson, 2001). However, the tasks used in SLA research are not always the same as those tasks which are described in the literature defining pedagogic tasks (Nunan, 2006). The tasks may not meet all of the criteria of a task (see above definitions from Skehan, Willis etc.) and often contain key differences which could impact the ability of these studies to draw conclusions about how SLA occurs (Samuda & Bygate, 2008). So how and why tasks can focus in on performative elements of speech production elements, such as fluency, is still the subject of much research.

There are other key differences between 'authentic' tasks used in the language classroom to facilitate SLA and those conducted for research purposes. These differences include the setting of the task; whether it is in a language classroom or a laboratory setting, as well as the format and implementation of the task; whether or not it is part of a task cycle, or pedagogic sequence which would include pre and post task activities (Bygate, 1998). These differences could impact the validity of any conclusions drawn from such research and must be considered when exploring the data collected. In relation to the current study, it is important that the tasks are authentic and as close as possible to tasks the learners would encounter in their own language learning classrooms. This authenticity will enable a clearer view of the extent to which the impact of task design, here in the format, and the complexity of the tasks, has on the learner's ability to produce fluent speech in the L2.

3.8 Task Complexity

The complexity of tasks and how they can be operationalised and defined have become central constructs in much recent research in TBLT examples include: Awwad, Tavakoli, and Wright (2017), Declerck and Kormos (2012), Norris and Ortega (2009), Robinson (2005), and Ssayama, (2016). The recent research into task complexity is primarily based around two notable theories of task complexity those of Robinson (2001, 2007), and Skehan (1998, 2009). The two theories are discussed in more detail in the following section (3.9). The importance of task complexity is derived from the theory that certain task characteristics can encourage language development and acquisition (Préfontaine & Kormos, 2015; Awwad & Tavakoli, 2019). If a task is not difficult enough to challenge a learner, then they may not have the chance to notice the 'gap' in their knowledge and could miss the opportunity to acquire the target language. Conversely, if the task is so difficult that it becomes impossible to transact then the learner will not only be discouraged, but again be denied the chance to learn. A learner's cognitive abilities enable them to acquire L2 knowledge, but can also influence how they use this knowledge to perform different types of task (Kormos & Trebits, 2012). An understanding of task complexity and how this drives SLA is obviously central to the design of the tasks themselves, and also to TBLT tests and language assessment, as well as to the sequencing of tasks in a TBLT syllabus (Awwad & Tavakoli, 2019). It is also important in relation to developing fluency, and other specific targeted aspects of language performance.

While task complexity is a widely researched area it is still one which requires further investigation and study. Awwad and Tavakoli (2019) highlight two areas of importance for further research: firstly, the varying aspects of task complexity, for example the reasoning demands required, and the way in which these variables are operationalised. The present study seeks to address this through the investigation of the impacts of tasks cognitive complexity, operationalised through the +/- order of information, to establish how task complexity impacts oral fluency in L2 speech. Secondly, the relationship between task complexity and learner-internal variables, with examples including levels of PSTM, or language proficiency. Recent research (Malicka & Levkina, 2012; Révész et al, 2011) has pointed to the importance of further investigating these learner-internal factors and the way they mediate the role of task complexity in second language acquisition and performance (Awwad & Tavakoli, 2019). This is something the present study seeks to address by exploring the impacts of L2 speaker levels of PSTM on varying levels of task complexity and input format.

As mentioned above there are two major conceptualisations of task complexity in the literature on TBLT and they have informed much of the research in this area, they are the Cognition Hypothesis (Robinson, 2001, 2007) and the Limited Capacity Hypothesis (Skehan, 1998, 2009). The following section will explore them in greater detail.

3.9 Task complexity and Competing Theories; Skehan's Limited Capacity Hypothesis (1998, 2009) and Robinson's Cognition Hypothesis (2001, 2007)

Robinson (2001, 2007) and Skehan (1998, 2009) have both adopted a cognitive approach to language learning focusing on the psychological processes that are involved with learners performing tasks, and a particular focus on how the "cognitive demands of tasks affect learners' attentional resources and language performance" (Tavakoli 2009, p. 484). Robinson and Skehan offer differing perspectives on the role of cognitive demands in task creation and implementation. Skehan (1998) proposes that increases in fluency can be achieved, but not with simultaneous gains in accuracy or complexity, and this 'trade-off' is due to the limited information processing capacity of the learner. If the characteristics of a task are designed to facilitate a focus on, and increases in fluency, this may come at the cost of accuracy and complexity. As learners are focused on producing fluent speech, they have less attentional resources to attend to the accuracy and complexity of the language they are producing. Robinson (2007) however believes that attentional resources are not limited, but instead learners can access separate areas of attention that do not compete with each other. Increasing the cognitive demands of a task will encourage L2 learners to use more complex language, containing greater lexical variety, and at the same time displaying increased accuracy (Robinson, 2007).

3.9.1 Code Complexity, Cognitive Complexity, and Communicative Stress

Skehan's and Robinson's theories are different in regards to their focus on and understanding of task complexity and the effects it has on speech production. The Limited Capacity Hypothesis (LCH) (Skehan, 1998, 2009) explores task demands, noticing, and speech production and states that increases in fluency, accuracy and complexity can be achieved through reductions in task complexity, with increases in planning time in the pre-task phase, as well as through repetition, and increased task familiarity. The hypothesis of a limit to processing capacity is founded on theories of working memory, such as those by Gathercole and Baddely (1993). Analysis of task complexity and difficulty is partitioned into three areas: *code complexity*; the language required, *cognitive complexity*; the thinking required, and *communicative stress*; the performance conditions of a task (Skehan, 1998, 2007). The elements comprising task complexity are divided between the language needed to complete the task, the thinking required to organise both the language and the task itself, and the performance related skills needed to complete the task. In relation to the current research project, it is those elements which are covered by 'cognitive complexity' which are of most interest, as they are the features which relate to the content of the task and how the material is structured (Kuiken & Vedder, 2007).

The LCH states that the more difficult a task is the more conscious attention it requires from the learner and therefore the less attention they have available for a focus on form. Skehan (1998) argues that because of the limitations in attentional resources available to the learner tasks can either be geared toward a focus on increased accuracy, or a focus on increased complexity, but not both. Understanding the factors which impact task complexity and having the ability to arrange tasks, from less cognitively demanding to more demanding, enables the optimization of opportunities for attention to language forms and leads to a balanced development of language. As Skehan (1998, p. 119) summarizes: "second language development and second language use may enter into some degree of mutual tension since the priorities of real-time language use may siphon away attention from noticing and interlanguage change." In other words, if the learner is too focused on completing a task, they may not pay sufficient attention to the language they are using, and conversely if a learner is too focused on the language, they may not meet the communicative goals of the task. This has implications for the current study and the investigation of the impacts of task design on fluency; as we seek to understand how to promote and have L2 speakers produce more fluent language. A balance must then be sought between the use of language which relies on analysable and accessible systems, i.e., declarative memory, which depends on lexical and other rule-based systems, and at the same time the language performance. An L2 learner then may lack the attentional capacity to maintain equally high levels of fluency, complexity, and accuracy and may focus on, or at least prioritize, *using* a language at the expense of *learning* a language; they may focus on their current communicative goals at the cost of developing their interlanguage skills (Skehan & Foster, 2001).

The LCH maintains that as language learners have a finite capacity to process information, they have to prioritize what they focus on and give their attention to; whether that be completion of the task at hand through the use of communication strategies or on 'learning' the language being used through a focus on forms. In other words, if a task requires a lot of attention to its content, because of the complexity of the task, language learners will have less attentional resources available to focus on the language they must produce to perform the task (Tavakoli, 2009). In regards to the present study it means we will expect to see L2 speakers produce less fluent language on those more cognitively complex tasks, as resources are directed toward task completion and away from speech production.

Criticism of the LCH has pointed to the fact that the characteristics described are not detailed enough to enable direct links between pedagogic versions of the tasks, to the wide variety of tasks performed in authentic real-world language learning environments (Robinson, 2011). A further area of limitation for the LCH is that it is unclear to what extent the factors of code complexity, cognitive complexity, and communicative stress interact with each other (Kuiken & Vedder, 2007). It is possible that one may take precedent, or exerts a stronger influence than the others in certain task conditions. It is also not clear how these factors interact and to what extent they are interdependent. Resultantly, the limited attentional capacity model does not offer any clear guidelines in regards to task design, and how these three dimensions should best be operationalised in tasks (Kuiken & Vedder, 2007).

3.9.2 The Cognition Hypothesis and Triadic Componential Framework

The Cognition Hypothesis (Robinson, 2001) differs from the limited capacity hypothesis as it does not believe a 'trade-off' between attention to accuracy and attention to complexity is required. Rather, the cognition hypothesis states that on certain aspects of task demands increased cognitive complexity can result in production of more accurate *and* more complex learner output. Less cognitively demanding tasks result in "grammatically simpler utterances and more complex task demands lead to morphologically richer and structurally more complex syntactic learner output" (Robinson, 2011b, p. 14).

The Triadic Componential Framework (TCF) (Robinson, 2007b) was developed as an operational taxonomic framework of task characteristics, it divides aspects of task demands in task performance between three broad groups. The first of which is task *complexity*, it describes the cognitive factors, the intrinsic mental complexity, and the reasoning demands which are required by the task. The second component in the TCF is task *conditions* which concerns participation and the interactive demands of tasks, as well as the participant, for example the familiarity or unfamiliarity of a task. Task *difficulty* is the third component and it explores the learner's perceptions of task demands and the individual differences in learners, for example working memory capacity, which can affect how difficult learners believe a given task's demands are to meet. These individual learner differences can broadly define which tasks learners have a preference or 'aptitude' for within a task based framework. Similarities can be seen with Skehan's Limited Attentional Capacity Model 'cognitive complexity' and Robinson's 'task complexity', as well as Skehan's 'communicative stress' which is much like Robinson's 'task conditions'. The third element of the TCF is 'task difficulty', but this has no equivalent corresponding entry in the LACM, and is perhaps the least well defined construct in Robinson's framework (see section below on limitations for a more detailed explanation).

Further divisions are drawn by Robinson (2007a) between the three areas of reasoning demands made in tasks. *Spatial* reasoning firstly is concerned with awareness of space, common examples in tasks are giving directions, navigating around a city, or where to find a particular item in a library. *Causal* reasoning involves a learner being able to understand and explain a natural or mechanical process as well as how and why it occurred, examples include why a car crashed in bad weather, or why a computer has stopped working. *Intentional* reasoning is the understanding of the intentions, desires, and beliefs of people and the ability to explain their actions, some examples are why your boss is angry with you, or why you want to be a doctor (Robinson, 2007a). The taxonomic listing of pedagogic task features in the TCF are detailed in figure 3-1 below.

Task complexity	**Task condition**	**Task difficulty**
(Cognitive factors)	(Interactive factors)	(Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior descriptive analyses)	(Classification procedure: ability assessment analyses)
Sub-categories: (a) Resource-directing variables making cognitive / conceptual demands +/− here and now +/− few elements −/+ spatial reasoning −/+ causal reasoning −/+ intentional reasoning −/+ perspective-taking	Sub-categories: (a) Participation variables making interactional demands +/− open solution +/− one-way flow +/− convergent solution +/− few participants +/− few contributions needed +/− negotiation not needed	Sub-categories: (a) Ability variables and task relevant resource differentials h/l working memory h/l reasoning h/l task-switching h/l aptitude h/l field independence h/l mind reading
(b) Resource-dispersing variables making performative/procedural demands +/− planning time +/− prior knowledge +/− single task +/− task structure +/− few steps +/− independency of steps	(b) Participant variables making interactant demands +/− same proficiency +/− same gender +/− familiar +/− shared content knowledge +/− equal status and role +/− shared cultural knowledge	(b) Affective variables and Task relevant state-trait differentials h/l openness h/l control of emotion h/l task motivation l/h processing anxiety h/l willingness to communicate h/l self-efficacy

Figure 3-1. *The Triadic Componential Framework for task classification – categories, criteria, analytic procedures, and design characteristics (from Robinson, 2007a)*

As figure 3-1 above highlights there is a further important distinction in the TCF, it is one which is made between those task features affecting the ‘resource-directing’ and ‘resource dispersing’ dimensions of complexity (Robinson, 2003a). The cognitive complexity of tasks, the resource-*directing* dimensions are those factors which affect allocation of cognitive resources to particular areas of the target language. Tasks that increase *intentional*-reasoning demands mean learners must comment on how people might be feeling, and would focus learners’ attention to new language and terms in the L2, for example ‘presume’ or ‘suppose’ (Robinson, 2011). It is questionable to what extent this language can be considered ‘new’, especially for more advanced speakers of the target language. However, the idea is the learner is stretching their interlanguage as they may not know these terms, how to use the terms properly, or more likely have sufficient practice using them in authentic instances. Through attempts to incorporate these ‘new’ language elements into their speech to complete the task they will *notice* and become more aware of them (Robinson, 2011). As complexity is increased along resource-directing dimensions, language learners gain opportunities to expand their knowledge of the target language’s vocabulary and linguistic structures. This process is facilitated through practice with an increase in the linguistic resources, and a shift from implicit to explicit knowledge (Robinson, 2011).

Resource-*dispersing* dimensions on the other hand are those elements which do not focus on specific aspects, but instead spread the focus of the learner’s attention over a wider range of task elements and demands (Robinson, 2011). Increasing task demands along the lines of resource-dispersing dimensions slowly removes assistance for language processing, providing less scaffolding for speech production. Practicing tasks with more complex resource-dispersing elements then means that speakers will be encouraged to produce speech more quickly and more automatically (Robinson, 2011). This increased task complexity, is argued by Robinson, to lead to a learner being more efficient in accessing and developing their interlanguage, and would result in greater opportunities to develop both complexity and accuracy in the L2. Based on the TCF and the resource-directing/dispersing distinction Robinson (2011b) states that two operational principles should be taken into account when organizing tasks; firstly, in agreement with Skehan (1998) and others, that tasks should be sequenced for learners solely on the basis of their increase in cognitive complexity, rather than on the difficulty of the language required for completion. Secondly, that tasks should begin with increases in resource-dispersing dimensions of task complexity first, to encourage learners to use their current interlanguage. Only after increases in resource-dispersing dimensions have been increased should increases in the resource-directing dimensions of tasks be implemented (Robinson, 2011).

To summarise, Robinson suggests that first tasks should be designed to increase in complexity which challenges learners across a wide spectrum of elements. This allows the learners to communicate and acquire new linguistic elements and vocabulary, through the expansion of their interlanguage to meet the goal of the task at hand. At the same time, they will also become more comfortable using these features in communication. After this, tasks can then be designed to increase in complexity through elements which focus on specific language structures, and as a result lead to a focus on, and an opportunity to acquire, new language forms.

Robinson's Cognition Hypothesis and the TCF are more complex and elaborate than the LACM, and the most recently updated version of the TCF (Robinson, 2007) contains 36 different elements. Similar to criticism of the LACM, the Cognition Hypothesis lacks detailed explanations as to how these myriad factors interact which each other, and to what extent they can be separated and isolated in task design (Kuiken & Vedder, 2007). With so many measures detailed in the TCF it is almost impossible to avoid overlap amongst them, when trying to operationalise each one. This difficulty of operationalisation is problematic when attempting to implement the Cognition Hypothesis through the TCF, and it is often hard to distinguish between, for example the number of elements and the number of reasoning demands. In this example it would seem fair to surmise that as the number of elements increases so to do the number of reasoning demands, how then should the two be separated? The interrelated nature and inability to isolate and differentiate specific elements have been the source of much debate in the literature, and accounts for many of the seemingly apposing or varied results generated from research into task design.

To better understand the Cognition Hypothesis and TCF and the way in which they describe task complexity, and how this complexity impacts the cognitive factors underpinning speech production, it is necessary to look at the important role which schema play. Schema can be defined as the conceptual systems which are used for understanding and organising knowledge and information, about different concepts in the mind. The 'restructuring' of schema plays an important part in SLA, and tasks which increase in cognitive complexity necessitate learners activating more complicated schema, required to meet the necessary communicative function. This in turn means that the speakers use more linguistically complex language forms form the activated concepts (Kormos, 2011; Robinson, 2005). Tasks which increase along what Robinson calls the 'resource directing dimensions' required to perform a task then activate more complex concepts. Kormos (2011) explains that this impacts macroplanning at the conceptualisation stage of speech production, because this is where concepts are selected, and relationships between them are encoded. As was explained earlier, in section 2.5, macroplanning occurs in the conceptualisation stage of speech production, as this is where the learner creates the pre-verbal plan and addresses the communicative goal of an utterance. As a result, more complex concepts will result in learners producing more sophisticated grammar and vocabulary, resulting in higher levels of complexity in language and greater accuracy. Resultantly, to generate more lexical variety and more complex syntax, tasks need to be more complex in resource directing ways (Kormos, 2011).

This leads us to another area of limitation when implementing the Cognition Hypothesis (Robinson, 2001) and the TCF in research; the models fail to adequately provide for the role of fluency. Robinson's research is heavily focused on task complexity and how it impacts on accuracy and complexity, with fluency largely being overlooked. This is problematic when exploring task design and the effect it has on language performance in relation to fluency. This lack of detailed explanation in regards to fluency, coupled with the difficulties inherent in isolating one of the many components of the TCF. may account for the many studies which found conflicting or inconclusive results. The current research project is aiming to address this gap, as it explores the relationship between task design elements in Robo's TCF (+/- order of information), as well task input format (pictorial/written), and L2 speaker oral fluency. Research has attempted to establish the interaction between gains in fluency and how they are impacted by task design, but many research projects which have sought to explore fluency have tended to operationalise it through reflections of the LCH, rather than the CH and TCF. This has led to somewhat of a split in the literature, with many articles explaining things in terms of either the LCH or the CH and TCF, but rarely with attempts to reconcile the two hypotheses.

The limited attentional capacity hypothesis (Skehan, 1998, 2009), and the cognition hypothesis (Robinson, 2001, 2007) have provided the theoretical foundations for much research, however differing and often contradictory findings have proven problematic for attempts to completely support either. The diversity of findings is likely due to the fact that some research has failed to take into account that tasks often contain elements which employ resource dispersing demands on learners, where as other characteristics of tasks draw attention to resource directing demands, often simultaneously (Kormos & Trebits, 2012). In authentic real world communication, both of these demands would be present, and affect task performance, so for the purposes of research it is important to examine all of the demands made on a learner during task performance (Pallotti, 2009).

When first exploring the LCH and the CH they appear to hold contradictory positions in relation to how they explain the focus of attention (Kuiken & Vedder, 2007). However, in regards to the view of resource-dispersing variables they are in agreement that attentional resources are finite and limited. It is only in the explanations in respect of resource-directing variables that they differ. While there are clearly large areas of overlap between the two theories there are also other elements which require further examination; the distinction between task demands and task complexity and the relation between them is one which needs to be clearly defined. The *cognitive load* of a task and the *task difficulty* are differentiated by Robinson in his model (2007), with a task's difficulty being ascribed to learners' perceptions of how difficult it is to complete a task. A distinction then is drawn between the actual amount of mental resources required to transact the task, and the individual learner's assumption of just how difficult the task is to complete. In other words, a task which may be perceived by a learner to be very difficult does not necessarily result in increased cognitive load, and does not automatically result in more mental effort being expended for task completion (Sasayama, 2016). This distinction between 'difficulty' and 'complexity' in the CH can be problematic, as it can be hard to differentiate between the two. To further confuse matters when a task is more difficult for learners, it would be reasonably expected that it would also be more complex to transact. 'Difficulty' may be used in a particular way in the CH and is to some degree a more subjective term, but ultimately is it still problematic to disentangle from 'complex' and it is no easy tasks to operationalise and clearly sperate the two concepts. The differentiation between what is difficult and what is complex seems to be somewhat artificial, and fails to take into account the nature of many of the tasks operationalised in research, but perhaps more problematically those used in classrooms. The distinction between cognitive complexity, and the difficulty of task transaction is a valuable one to the current research project, as this project seeks to investigate the impacts of task design on the cognitive processes of speech production. The tasks employed in this research project seek to increase the cognitive complexity of the tasks, but maintain the linguistic difficulty of the task to better isolate the impacts on the attendant mental processes in L2 speech production.

In regards to this research project to better understand the participant's views on task complexity and tasks difficulty all of the L1 participates in the current research project completed a post task questionnaire. The results show that those tasks which were operationalised as more *complex* according the TCF were also found to be more *difficult* to transact by the L1 participants (see appendices E, H, I J, K, and L for further learner responses, and section 5.6.2 for details of implementation). The responses and the analysis highlight the problematic nature of separating complexity and difficulty, two of the primary components of the TCF.

The identification and isolation of the cognitive demands required by specific task design elements, and the task characteristics which are linked to them will clearly be of value to future research. A better understanding will hopefully be provided by this study through the investigation of increased cognitive complexity and its impacts on L2 speaker oral fluency. The current research project will also help to more clearly identify which design elements affect L2 speakers when they transact a task, through the careful analysis of the differing task design elements.

3.10 Task complexity, Intentional Reasoning, and Oral Fluency

Tasks require participants to employ different reasoning skills to transact them depending on the demands of the specific task (Robinson, 2007; 2013). Tasks which require a learner to describe motions employ spatial reasoning; tasks that require a learner to predict and presuppose peoples' thoughts and intent require intentional reasoning; and tasks which necessitate an explanation behind the reasons for actions employ causal reasoning (Robinson, 2007). These task elements have high intentional reasoning demands and result in a focus towards increasing the use of more complex and accurate language to meet the needs of the task (Robinson, 2013). So intentional reasoning demands in a task mean learners will adopt more complex linguistic forms and more complex lexis to be better able to describe the thoughts, reasons, or actions required by a task (Robinson, 2007; Awwad & Tavakoli, 2019). However, it is important to note that Robinson (2015) describes how this increase in linguistic complexity *or* accuracy comes at the cost of a corresponding decrease in either complexity *or* accuracy, due to the increased demands of intentional reasoning. As was noted before there is little mention of fluency in Robinson's explanation, but we may surmise that an increase in either accuracy or complexity would see a fall in the levels of fluency. This may be primarily due to the increased cognitive load of the task, through the increase in the task reasoning demands, and the complicated lexical structures necessitated are likely to impact the difficulty in the formulation of utterances. In summary, as intentional reasoning demands increase so too does the complexity or the accuracy of the language produced, but fluency may suffer due to the impact of the increased cognitive load on the L2 speaker at the formulation stage of speech production.

3.11 Impact of Task Complexity on L2 Oral Development

Several studies have explored the impacts of task cognitive complexity and its influence on L2 speaker performance. However, they have typically focused on language development and involved a pre, during, and post-task structure. Nuevo (2006) operationalised task complexity through (+/-) *causal reasoning* in narrative and decision making tasks, but found no increase in learners gains in English tenses. Similarly, Kim (2009) explored the effects of (+/-) *causal reasoning* in interactive tasks on learners' abilities in question formation and the past tense, but unlike Nuevo (2006) reported gains in L2 development. More recently Révész et al., (2011) investigated how operationalising task complexity through (+/-) *here and now* affected learners' development on descriptive tasks and reported gains in L2 development, specifically for the past progressive tense. The above mentioned studies looked at L2 development, while not directly related to this study and its focus on oral fluency they do provide considerable evidence of the impacts of task complexity on L2 performance. We can again conclude that in line with these studies on oral development, as task complexity is increased through task design then levels of l2 oral fluency will decrease.

3.12 Task Complexity and Language Proficiency

The challenge of understanding task complexity is often increased by the variability of language learner's proficiency in the target language. Having less proficient speakers perform tasks makes the investigation of task complexity even more complex and problematic (Awwad & Tavakoli, 2019). Language proficiency, defined here by Gaillard and Tremblay (2016) as the linguistic knowledge and skills that make up an L2 speakers' understanding and production of the target language, is important as it relates directly to the amount of automaticity available (Awwad & Tavakoli, 2019) and the amount of attentional resources required (Kormos, 2011) for speech production. As was detailed earlier in the section on speech production (2.5) learners who are more proficient are able to complete parallel processing, requiring less attentional mental resources, but those of lower proficiency are forced to switch to serial processing which greatly reduces the efficiency of speech production (Kormos, 2011). It is important then to control for the variable of language proficiency, to decrease the impact of any interactions between language proficiency and task complexity.

As can be seen from the studies detailed above, research has tended to focus on accuracy, and on language acquisition, rather than on fluency. However, recently more studies have chosen to explore the relationship between language proficiency, task design, task complexity, and oral fluency (Malicka & Levkina, 2012; Awaad & Tavakoli, 2019). Few studies have reported results which showed increased task complexity to have a positive impact on oral fluency, but one such study is the study by Malicka and Levkina (2012). They found that increased cognitive complexity actually resulted in higher levels of oral fluency for less proficient L2 speakers. This unexpected result was explained in relation to Skehan's Trade-off Hypothesis (1998) with the argument being that the less proficient speakers lacked the skills to meet the cognitive demands of the complex task so therefore focussed on a single aspect of performance, in this case fluency, to the detriment of both accuracy and complexity. They reasoned that as L2 proficiency increases so too does the ability to allocate attention, with those less proficient speakers lacking the attentional capacity to meet the increased demands of speech production and complex task completion. Malicka and Levkina's (2012) findings provide support for the Trade-off Hypothesis and suggest that an L2 speaker is able to 'focus on fluency' to a certain extent. By decreasing focus on language complexity and accuracy L2 speakers are able to speak with fewer pauses and breakdowns, and thus more fluently. In real terms this means L2 speakers would produce shorter, simpler sentences, with fewer words, a lower level of complexity in vocabulary, and make more grammatical or linguistic errors. Through a comparison of advanced and pre-intermediate learners Malicka and Levkina (2012) explored how language proficiency was impacted by controlling the number of elements and reasoning demands in an 'instruction giving' task. They conclude that the participants with higher proficiency levels constructed more accurate and complex language during the performance of the more complex version of the task, but this came with the consequence of decreased fluency. However, interestingly they also reported findings that showed learners from the lower level proficiency group produced more fluent language on the complex task. They attributed this to the lower level speakers not having the linguistic skills to focus their attentional resources on producing complex, accurate, and fluent speech. The more proficient learners however, were said to have more control over their attentional focus and were able to achieve more complex and more accurate language, but this came at the cost of a reduction in fluency. Perhaps, as they conclude somewhat counterintuitively, the more proficient L2 speakers produce lower levels of oral fluency because they are striving to meet the demands of the more cognitively demanding task, and not simply directing all of their mental resources toward producing fluent speech (Malicka & Levkina, 2012).

In the second study mentioned, Awwad and Tavakoli (2019) explore L2 proficiency through a broader perspective by looking at the distinction between 'explicit' and 'implicit' knowledge. This is an important distinction when exploring task performance and the allocation of mental resources during speech production, as using unconscious, implicit knowledge is procedural in nature and has been found to aid in fluent speech production through the ease of access (DeKeyser, 2001, Kormos, 2011). Those mental processes which are implicit and have become part of a process, are faster in access, and require less conscious attention to utilise. See description of Anderson's (1982) skill acquisition theory in section 2.9.1.1. On the other hand, explicit knowledge is linked to declarative knowledge, and is most often acquired through conscious awareness, and is assumed to result in controlled processing (DeKeyser, 2001; Kormos, 2011; Awwad & Tavakoli, 2019). Declarative knowledge as was detailed before in sections 2.5 and 2.14.2 is stored in the long term memory and requires greater attentional resources to access, resulting in an increased cognitive load (Kormos, 2006; 2011) and has been directly linked to errors and slowness of L2 speech (Anderson, 1982).

Awwad and Tavakoli (2019) reason that this distinction will help to show more clearly the importance of proficiency on the mental processes employed during task performance, and speech production. Their study found that measures of oral fluency were higher for the more complex version of the task, but were only significant for measures of speed fluency, which was operationalised through speech rate. While the less complex version of the task saw longer clauses and more frequent pauses, both at the end and mid clause level, however these differences were not seen to be statistically significant. The findings of Awaad and Tavakoli's (2019) study offers partial support to Skehan's (2009) limited attentional capacity theory as some aspects of fluency are observed to participate in a trade-off between measures of syntactic complexity and accuracy. However, other aspects of oral fluency are seen to increase in line with gains in complexity and accuracy, necessitating the further investigation into which factors impact, and how they impact, the cognitive processing of L2 speakers during task performance.

The conclusions to be drawn from these previous studies then are not that one approach is more accurate or valid than the next, but rather that the way in which task complexity is operationalised, the individual learner differences of the participants, the mode and manner of the task performance, and the ways in which performance is measured all have a considerable impact on the outcome. When all of these elements are different than it is difficult, if not impossible, to accurately describe which particular aspect is responsible for the evidenced performance, and what this tells us about the underlying cognitive processes. This further highlights the need to create comparable operationalisations of complexity, to deliver reproducible tasks, to consider individual learner differences, to employ consistent measures of performance, and ultimately to better analyse which factors involved in both the task design, and the make-up of the participants are responsible for the performance.

3.13 Task Complexity and Speech Production

Robinson's (2001b) list of task characteristics which impact task complexity includes: the number of different elements in a task, the amount of time available for planning pre-task, and previous knowledge which may be helpful, however as Kormos and Trebits (2012) point out linking speech production to the different operationalisations of tasks is a necessary endeavour. As explained in section 2.2, Levelt's (1989) model of speech production constitutes four parts; conceptualisation, formulation, articulation, and self-monitoring. Of great importance to this study and to considerations of comparisons of L1 and L2 oral fluency measures is the fact that the conceptualisation stage of speech production requires conscious attention, for first language speakers however the formulation and articulation stages are largely automatic processes, enabling the parallel processing of these stages and the production of generally smooth and fast speech amongst L1 speakers. For L2 and non-balanced bilingual speakers however the speech production process is far less automatic and the formulation and articulation stages in particular will require greater attention and conscious effort, which often hinder parallel processing and can lead to a loss of fluency.

Task complexity is usually seen as directly relating to the conceptualisation stage of speech production, with little link to formulation (Kormos & Trebits, 2012). The idea being that more complicated concepts cannot be described without using more complicated linguistic elements, so complex tasks are then more difficult to conceptualise and to formulate, but are not designed to be more difficult to formulate (Robinson, 2001b, 2007). However, as Kormos and Trebits (2012, p. 6) argue it may be "…possible that tasks make separate and independent complexity demands on the conceptualisation and formulation stages…" of speech production. It is clear however, that different types of task design impact different forms of cognitive complexity; a task with written input for example, provides linguistic information in the form of words, this may provide an easing of the lemma retrieval section in the formulation stage of speech production through ‘lexical priming’, but may also increase difficulty in the conceptualisation stage by providing linguistic concepts that participants are unfamiliar with or lack the vocabulary to express. As Kormos and Trebits (2012) explain participants can only use the linguistic resources they have available to them in their second language, which makes it that much more difficult to complete speech in the target language. This would obviously not be the case in a picture narration task, as the participants have more flexibility to match the language they employ to the task narration, put simply if they do not know a linguistic item they will not employ it. This can be clearly linked to Levelt’s (1989) model of speech production, as task demands that do not require a participant to think about ‘what they want to say’ or require them to ‘create’ a story ease demands on the conceptualisation phase, whereas tasks that provide learners with linguistic items may reduce the cognitive demands on ‘how they want to say it’ in the formulation stage of speech production. These examples illustrate that the cognitive demands of a task are often interrelated and multifaceted, and ordering tasks based on cognitive complexity may be problematic, with different demands focusing on separate aspects of speech production (Skehan, 1998; Pallotti, 2009). To this end Kormos and Trebits (2012) argue that the cognitive demands of a task should be considered separately in relation to the difficulty required to conceptualise and formulate speech in an L2.

This has direct relevance for this research project, and it will be interesting to see which task input type generates the more fluent language production. In relation to the current study, it may well be that tasks with the most ‘lexical priming’ in terms of formulation lead to the production of more oral fluency, but the impacts of variations in complexity in the conceptualisation stage will also have an impact on L2 oral fluency. The next section discusses this in terms of recent studies.

3.14 Conceptualizing and Fluency in L1 and L2 Speech

Several recent studies have sought to address this separation of the cognitive demands in the formulator and the conceptualiser, with recent research suggesting that disfluencies in speech may be related to the different stages of speech production (Fraundorf & Watson, 2013; Hartsuiker & Notebaert, 2010; Mirdamadi & De Jong, 2015; Felker et al., 2018). These studies examine losses in fluency caused by increased task complexity, but differentiate between difficulties with linguistic encoding in speech formulation, which involves lemma retrieval from the mental lexicon, and losses of fluency resulting from conceptual planning difficulties in conceptualisation, involving content selection. A comparison then of L1 and L2 speech production may help to highlight which aspects of speech production are impacted by increased task complexity, separately from linguistic complexity. Speaking in an L2 increases cognitive processing demands on both the formulation and articulation stages of speech production (processes which are largely automatic in an L1), but may not result in similar increases in demands on macroplanning at the conceptualising stage (Kormos, 2006). It has still not been widely researched how an increase in complexity at the conceptualisation stage will impact oral fluency in a second language relative to a first language. Studies exploring L1 speech production have debated if macroplanning and microplanning utilise the same attentional resources (Levelt, 1989; Roberts & Kirsner, 2000), arguing that an increase in the demands on one sees a corresponding increase in the other. However, more recent research (Skehan et al., 2016) has linked microplanning to the formulator stage and the linguistic complexity of at a task, and macroplanning with the conceptualiser and discourse level decisions, and has considered their impact to be largely separate. Greene and Capella (1986), cited in Awwad and Tavakoli (2019), describe how in spontaneous speech switching between the subgoals or 'moves' in the planning of a discourse would require extra attentional mental resources. The increased cognitive demand during conceptualisation in speech production would result in more breakdowns and pauses at the discourse level, when speakers are trying to link sentences together and move from one idea to the next. In relation to the present study these pauses can be viewed in a similar way as pauses at the end-clause position, occurring at the boundary of a clause and are indicative of problems with macroplanning and the conceptualisation stage of speech production. Greene and Capella (1986) found that the majority of pauses at 'idea boundaries' or the end-clause position were silent pauses, and that these pauses decreased greatly when increased planning time was allocated pre-task to facilitate the organising of discourse structure. In summary, when demands on planning and conceptualising were eased the number and frequency of pauses at clause boundaries decreased (Felker et al., 2018).

In a study by Roberts and Kirsner (2000) which explored unplanned speech samples, similar results were reported. The study provided evidence of more pauses and lower levels of fluency before a change in topic, and greater levels of fluency and fewer pauses immediately after a shift in topic. The study concluded that macroplanning does not support, but rather competes with other aspects of speech production for a limited amount of attentional resources (Skehan, 2009; Kormos, 2006; Segalowitz, 2010). Macroplanning then may operate as a 'speed bump' in the path of smooth and fluid speech processing, slowing down the other processes until it has been negotiated (Roberts & Kirsner, 2000). This "cognitive bottleneck" created at the macroplanning stage may limit the flow of fluent speech by demanding attentional mental resources, causing other elements of speech production to slow down or even stop, until the speed bump has been negotiated (Felker et al., 2018, p. 114). To summarise, as an L2 speaker transitions to a new thought, idea, or topic, it can require extra mental resources in conceptualisation which may result in a loss of fluency. This may well be connected with difficulties in macroplanning, but it is still problematic to ascribe the losses of fluency directly to macroplanning within the conceptualisation stage as there are many other factors which could play a part.

Losses of fluency then have been shown to align with changes in topic and focus in second language speakers' spontaneous speech and are indicative of difficulties in conceptualising (Felker et al., 2018). In a review of several early psycholinguistic studies which explored the impacts of changes in the difficulty at the conceptualisation stage of speech production on oral fluency Felker et al., (2018) found that it was possible to see direct effects of task complexity on oral fluency. In the first study by Siegman and Pope (1966 cited in Felker et al., 2018) longer pauses were found at subject boundaries, between clauses, in tasks with increased levels of difficulty in the conceptualisation stage. The increased difficulty was operationalised in this study through levels of ambiguity, with the task which contained more possible interpretations resulting in lower levels of fluency, seen in both the increased number of pauses, and the greater evidence of speech repairs. Another early study (Goldman-Eisler, 1968 cited in Felker et al., 2018) contrasted the task of describing a series of pictures, the less conceptually complex form of the task, with describing the same pictures, but this time requiring participants to ascribe meaning to the pictures, a more conceptually complex task. Complexity of conceptualisation was manipulated by contrasting having participants describe 'what they can see' with 'what it means', Goldman-Eisler (1968) found that the amount of speech dramatically decreased and the amount of pausing increased for the more complex condition. So, speakers spoke less and paused more when they had to describe what something meant, but paused less and spoke more when they merely had to describe what they could see. In the third early study reported, Lay and Paivio (1969, cited in Felker et al., 2018) compared different levels of cognitive difficulty and showed that oral fluency decreased with the increase in difficulty at the conceptualisation stage. While it is clear that all of these studies found decreases in fluency as the complexity of the task increased, it is not quite so clear if they were attributable solely to the increases in complexity at the conceptualisation stage of speech production or rather due to the impacts of other factors. As Felker et al., (2018) state the results of the above mentioned studies show increased task complexity impacts fluency, however differences in the complexity of language required, learner familiarity with the task, the input format of the task, as well as the difficulties in formulation and articulation stages, may also have influenced task performance.

To examine the effects of task cognitive complexity and the demands it makes on the different aspects of speech production Felker at al., (2018) in their own study explored how a variation in the difficulty conceptualising affected L1 and L2 speakers' oral fluency. While they tracked the variation in conceptualising difficulty, they maintained the same demands across the formulation and articulation stages of speech production. Their experimental design compared the effects of difficulty in conceptualisation across two task effects, the first looking at the generation of speech and the second looking at the abandonment and regeneration of speech. The participants were Dutch L1 speakers who also spoke English as an L2, they maintained the same subjects across languages. The results of the study showed that when participants abandoned, and especially when they regenerated their speech, greater numbers of breakdowns in oral fluency were recorded. This was true of both L1 and L2 speakers across the majority of the oral fluency measures employed by Felker et al., (2018). However, a significant difference was noticed between L1 and L2 speakers when comparing the time taken to solve conceptual difficulties involved with abandoning and regenerating a speech plan. This result is possibly explained by Kormos' (2006) description of the difficulties in regaining serial-processing and resuming the production of fluent speech, something which Kormos

(2006) notes is far less cognitively demanding, and requiring of less attentional mental resources, for L1 speakers than it is for L2 speakers. The participants were seen to take significantly more time to overcome these difficulties in resuming speech production after abandoning or regenerating speech when using an L2 compared to speaking in their L1. Felker et al., (2019) concluded that their findings provide support for theories that cognitive resources engaged while conceptualising are also being utilised by the other stages of speech production in formulation and articulation, and that any attempts to increase difficulty in conceptualisation should be attempted with care and caution. With this in mind the focus will now shift to another variable which may impact the complexity of tasks, that of the task input format.

3.15 Variation of Task Input Format and Design

The present study attempts to focus on the impacts of increased task complexity on the cognitive process involved in speech production, and another area that has seen little research is the impact of a variation in the type of task input format and the effects these differences have on oral fluency. Préfontaine and Kormos (2015, p., 96) point out that: "…task related variation and its interrelationship with speech production using automated measures has been a neglected component of empirical L2 fluency investigations…" and it is to this variation that we now turn. Only a few studies have attempted to address the role that differentiating specific task elements plays in language performance (e.g., Tavakoli & Foster, 2008). One area which has seen investigation is in the distinction between narrative tasks which are considered to have a 'tight' or 'loose' structure. Any narrative task with an obvious, or inherent structure, with a clear beginning, middle, and end is considered 'tight', and those which contain events which could be reordered without affecting the story are termed 'loose' (Tavakoli & Foster, 2008). In relation to Robinson's Cognition Hypothesis (2001, 2007) this may be thought of as similar in terms to +/- causal reasoning (see TCF 3.9.2). Narratives with a 'loose' structure require participants to create links between the pictures or sentences in task input, whilst those with a 'tight' or inherent structure have links which are more obvious, enabling easier conceptualising, macroplanning, and ultimately task transaction (De Jong & Vercelotti, 2016). Similarly, Kormos and Trebits (2012) examined L2 speaker narrative production, with complexity operationalised through information order; less complex tasks were created from pictures arranged in an obvious and straightforward sequence and the narratives compared to those which were arranged in a random non-sequential order. The results of their study were broadly in line with Robinson’s Cognition Hypothesis (2007) with the more cognitively complex task generating more complex language, however their results relating to fluency were inconclusive and point to the need for more research in this area.

Very few studies have investigated the role of the differences in speaker performance as a result of a variation in the input, or format, of the task, a few studies have however looked at the impacts of task type on L2 speaker performance (Gilabert et al., 2009; Skehan & Foster, 1997; Foster & Skehan, 1996). Many studies have employed video or computer based input (Baralt, 2013; De Jong et al., 2013) and much research has also explored picture based task input (Foster, 2009; Kim, 2009; Robinson, 2007b), however so far very few studies have looked explicitly at the effects of a variation of task input format on L2 speaker’s oral fluency measures, which is something the present study aims to address. This lack of research may be due to the reliance on the historical dominance of written input and traditional teaching ideas that classroom language learning is best achieved through written input (Krashen, 1989), but may also be due to some forms of task input not lending themselves to manipulation for research, or may be perceived as forms of entertainment (television, music, etc.) rather than educational (Feng & Webb, 2019). The present study seeks to address this gap in knowledge, and investigate whether the input format of a task impacts the oral speech production, while seeking to explain why this may be in terms of the different components of speech production and the attendant mental processes.

3.16 Effects of Differentiation in Task Design on L2 Speech Production

When investigating the impact of task design on L2 speaker oral fluency Kim (2009) found that the effects of task complexity on the occurrence of language related episodes varied depending on the type of task being performed. Similarly, in an investigation in to the impacts of task design on L2 speaker performance Tavakoli’s (2009) study observed increased performance on tasks with greater structure; with both increases in fluency and accuracy measures evidenced. Tavakoli and Foster (2011) in their investigation into task design also provided empirical evidence that L2 performance is affected by task design. They explored the impact of the sequence of narrative order in the task on L2 performance; through a manipulation of the sequence of pictures provided in the narrative input. The study showed how greater task complexity in the narrative of the task led to greater syntactic complexity, and an inherently fixed structure in the task design resulted in greater grammatical accuracy. Important for the scope of this study is the impact of the variation in order of the pictorial prompts for the narrative, the obvious and tight ordering of the prompts released attentional resources which would have otherwise been spent on establishing the correct sequence of events (Tavakoli & Foster, 2011). Although this study did not report the impacts of the increased complexity and narrative order on L2 oral fluency, it does show how structural elements and their sequence within tasks can impact other areas of performance. These studies all point to the need for further investigation in to the part that the input and the design of a task plays in the performance of L2 speakers, and specifically the impact of the input on oral fluency.

The studies detailed above focused on how increased complexity at the different stages of speech production affected oral fluency and impacted speech production. Skehan et al., (2016) found that broadly in line with the previous studies increases of complexity in task structure can be seen to impact fluency during the conceptualiser, during macroplanning, and result in increased breakdowns *at* clause boundaries, or 'between ideas'. However, the use of more formulaic language was seen to affect the formulator stage, during microplanning, and led to greater loss of fluency, recorded through a greater number of breakdowns *within* clause boundaries. This was explained in terms of the processing skills of language production, and described that those learners who used more formulaic language were not deploying original creative language to complete the tasks, but rather were relying on the use of preconstructed formulae to attempt to meet the needs of the tasks. So somewhat counterintuitively, but in line with the earlier discussion on formulaic and creative language use (2.14.2), those learners who had greater language proficiency, likely related to having a better understanding of, and ability to use the language structures, were able to produce more fluent speech.

The research to date provides evidence to support the notion that increases in the complexity of planning before a task, as well as increasing the number of options available for description during the task, affect macroplanning at the conceptualisation stage of speech production and result in reduced oral fluency at the end clause level, between sentences, clauses, or 'ideas'. Conversely the literature also provides evidence that disfluencies at the mid clause level, within clauses, and between words are as a result of difficulties in microplanning and the formulation stage of speech production.

For this current investigation, this means that we may expect to see higher levels of oral fluency on the task with a written input due to the easing of difficulty in the conceptualisation stage of speech production, due largely to the more structured nature of the task (Tavakoli & Foster, 2011). For the pictorial input, the freedom of structure will require the participants to generate their own scenario, and the extra effort in conceptualisation will divert resources away from fluent language production, causing losses in oral fluency, evidenced by more frequent pauses at the mid clause position.

3.17 Task Complexity and Working Memory

Several studies have explored working memory and how it interacts with L2 performance in general, as well as with task complexity more specifically. Working memory is viewed as being an important aspect of performance interacting with both L2 development and performance, and importantly also with task complexity (Cho, 2018; Gilabert & Munoz, 2010; Kormos & Trebits, 2011). The importance of working memory to task performance in particular is underlined by Conway et al. (2005) who describe it as a system made of several components responsible for the online maintenance of information flow despite interruptions, other ongoing mental process, and any other distractions. The links between working memory and task complexity are long established in the literature, with Kyllonen and Christal (1990) first putting forward the notion that performance of complex tasks in an L2 is impacted by working memory. More recently Wen et al. (2015) describe the potential influence working memory has on regulating L2 speakers' linguistic repertoire as well as their attentional resources during task performance. This highlights how integral working memory is to the successful production of fluent language in an L2, and how executive working memory, as well as phonological short term memory, may be able to mitigate the mental processing demands required during complex task performance.

In the study mentioned above Gilabert and Munoz (2010) found that working memory correlated with oral fluency. Their study sought to address the relationship between task complexity and working memory, and explored whether a variation in working memory and task complexity would explain any variation in L2 speaker task performance. The participants undertook a single video based input narrative task and no correlation was found between working memory and language proficiency, but working memory was found to correlate with lexical complexity, and as mentioned above, with measures of oral fluency.

In further research Kormos and Trebits (2011) explored L2 speakers' oral performance and how it was affected by the relationship between task complexity and working memory. The participants completed two tasks differentiated by task complexity, here operationalised through +/- information; one task requiring the learners tell a story (less complex) and the other task necessitated the invention of a story (more complex). The results showed that task complexity had little impact on the complexity of language used, and a high level of executive WM only benefitted syntactic complexity in the less complex version of the task. Interestingly no effect was observed on the complex version of the task for measures of oral fluency. In regards to my study, it will be interesting to see what impacts the increased task complexity, and the variation in task input format have on L2 fluency, and to what extent higher levels of WM, specifically PSTM, have on mitigating the completion of the more difficult version if the task. It is possible that the ability to hold more information for longer in PSTM enable the ability to produce more fluent speech, when completing the more complex tasks.

3.18 Skehan et al.'s (2016) Clause and Discourse Fluency Theory

So far, this chapter has explored tasks, and the impacts of variations is task design on L2 performance, but the focus will not switch to a study by Skehan et al., (2016), which proposes a new way to explore and measure the paradigm of task complexity and oral fluency.

Previously research on the dimensions of oral fluency for L2 speakers has focused on the triad of speed, breakdown, and repair fluency (e.g., Tavakoli & Skehan, 2005; De Jong et al., 2013). However, Skehan et al., (2016) have proposed a distinction between dysfluencies that occur ‘within’ a clause, and those which occur at the ‘boundaries’ of clauses. They argue that speed measures of oral fluency are largely separate factors as: “everyone punctuates the speech stream with pauses that do not arise from processing blips” and can vary significantly due to the individual differences of each speaker (Skehan et al., p. 99, 2016). Anecdotally we can each confirm this to be true; I am sure we all know someone who speaks quite slowly, perhaps they pause often or for a long time, as well as someone who speaks quite rapidly, but we would usually not consider them to be more or less fluent than each other when speaking in their L1. Rather this difference in pausing and speed of speech is just a reflection of the unique way in which each individual speaks, in much the same as their accent, vocabulary, or pitch. This focus on speed then can have the result of often obscuring the view of which factors impact oral fluency, as well as to what extent. The cognitive processing that occurs within a clause, i.e., the linking of words *within* a grammatical clause, as opposed to the joining of ideas into a discourse, is focused on lexical and morpho-syntactic choices and is directly linked to the formulator and articulator stages of Levelt’s (1989) model of speech production (Skehan et al., 2016). The dysfluencies that occur *above* the level of the clause i.e., developing an argument and joining units of speech together, are directly related to processing within the conceptualiser stage of Levelt’s model of speech production (Skehan, et al., 2016). This distinction is particularly relevant to a psycholinguistic perspective on language production, as it offers a fresh insight into the cognitive process being utilised at the time of the disfluency.

This new approach to the exploration of L2 fluency suggests that macroplanning is occurring during breakdowns in fluency at the discourse level, or the clause boundary, and breakdowns occur due to difficulties with processing which occurs in the conceptualiser stage of speech production (Skehan et al., 2016). Conversely those breakdowns in fluent speech which occur within a clause, between *words* and not *ideas*, occur during microplanning and are evidence of trouble processing during the formulator stage of speech production (Skehan et al., 2016). Results from the study showed that the degree of structural tightness did affect levels of oral fluency for the L2 speakers; with the most structured task being performed significantly more fluently than the other three versions, which showed little variation from each other (Skehan et al., 2016).

Skehan et al.'s., (2016) examination of the influences on clause level oral fluency highlights the importance of L2 speakers being able to draw on formulaic language. These idiomatic 'chunks' of language allow for the reduction in mental processing through the easing of microplanning and allow for faster and more fluent speech. The larger the number of formulaic chunks a L2 speaker has learnt and acquired then the greater capacity they have to produce fluent speech (Skehan et al., 2016). This is due to the decreased processing pressure on the formulator stage of speech production. The L2 speaker is no longer piecing together separate, single lexemes, but rather is selecting from pre constructed 'chunks' of data stored in the lexicon, which can be processed more efficiently, enabling the increase in oral fluency. Skehan et al.'s (2016) data seems to support this explanation, with more breakdowns in speech recorded for the L2 speakers *within* a clause/AS-unit rather than at the *boundaries* of clauses/AS-units in those tasks which provide less structural organisation. In other words, as cognitive processing resources are diverted away from speech production and instead towards task completion, in this case the structural organisation of the narrative task, then fluency for the L2 speakers decreases. This decrease in fluency is captured by the increase in the number of pauses and breakdowns in speech within a clause for the L2 speakers. The study found a significant increase in the L2 speaker's pauses and breakdowns at the mid clause level for the less structured tasks, when compared with the L1 speakers who did not record similar increases. These results would seem to suggest that the L1 speakers had sufficient cognitive resources to meet the processing demands of both the task organisation itself and for speech production. No significant increases in pauses and breakdowns at the mid clause level were found in L1 speakers across each of the differentially structured tasks. This ability to combine fluent speech with the task's processing demands comes as a result of the increased automaticity of L1 speech, as few attentional resources are required for microplanning and the formulator stage of speech production, mental resources can instead be reserved almost exclusively for task completion.

The research design employed by Skehan et al., (2016) provides an insight into how speech production differs in L1 and L2 participants. This comparison of L1 and L2 speaker performance, on the same tasks, allows for a clearer perspective of which oral performance features are a result of the difficulty transacting the task and which are the result of the more limited language resources of L2 speakers (Foster & Tavakoli, 2009).

Through this distinction areas of difficulty can be highlighted, and the mental processes being utilised when pauses and breakdowns in oral fluency occur uncovered, allowing for a better understanding of the effects of task complexity on speech production.

There are several limitations, which Skehan et al., (2016) themselves acknowledge, which need to be kept in mind when considering the effectiveness of the study, and evaluating their claims in regards to pause location and speech production. The first limitation they highlighted was the fact that the study involved individual learner variability through the utilisation of a between-subjects design. A comparison of L1 and L2 performance from the same speaker would eliminate this variable; perhaps with a native speaker completing the task in their L1 and then a similar task in their L2. However, practically speaking this may be unfeasible and would involve finding enough speakers capable of completing a task in the desired L1 and then in the selected L2. The task itself would also need to undergo modification as the participants could obviously not complete the same task twice. A further issue here would be that of comparing fluency across languages, a topic which has a large measure of inherent difficulties and related issues. I feel that the compromise of a between-subjects design is a necessary one and one which provides reasonably reliable and valid outcomes. The Second limitation is the fact that the task employed was a video narrative. This form of task input will create time pressure for the participants, i.e.; when the video is over, they can no longer see or hear it, likely impacting their responses and the performance of the task (Skehan et al., 2016). A written or picture based task prompt would not contain the same time pressure, with participants able to proceed at their own time, as well as having the possibility of going back to check if they are unclear on something. This can perhaps be explained in terms of the distinction between explicit and implicit information processing; with the picture or written input in hand a participant would likely be able to effortlessly incorporate the information into task transaction, whereas the video input would have to be recalled from declarative memory stores, and would be demanding of conscious effort and more attentional mental resources.

This also raises issues for the generalisation of the Skehan et al., (2016) study, it seems highly likely that different task formats and input types would return different results. This calls for far more research to ascertain how and why different task inputs produce different results. This is an issue which the present study will seek to address through the investigation of both a written and a pictorial input task. A further limitation of the study is again concerned with the generalizability of the study and its findings. The study sampled only mid-intermediate level English L2 speakers, so it would be necessary to examine higher and lower level L2 speakers to draw more robust conclusions from the data they reported (Skehan et al., 2016).

The current research project will seek to explore and examine the findings and suggestions made apparent by Skehan et al.'s., (2016) article, and take the research forward to explore the impacts of task input format and cognitive complexity in narrative structures, through a comparison of L1 and L2 speaker oral fluency, at both the mid clause and end clause level. The employment of different format types for the task input; pictorial and written, rather than video based as in Skehan et al., will seek to address the issue of the extent to which the results can be generalised. It will also be interesting to investigate Skehan et al's., (2016) claims on the new perspective on microplanning, and to ascertain to what extent it is possible to clearly link any breakdowns in fluency directly to the either microplanning or macroplanning. Specifically, can those pauses at the mid-clause position be reliably described as resulting from trouble with processing occurring at the formulation stage of speech reduction, and with microplanning.

3.19 Conclusions

This final section seeks to draw together all of the different theories and understandings of tasks design and its impacts on oral fluency in relation to the current study. The starting point for most theories in task design is with definitions of TBLT, which is a research supported teaching methodology and utilises a communicative approach to SLA. TBLT aims to encourage the use of real and authentic language to facilitate language learning. This approach is in contrast to more traditional approaches to LT which focus on forms, or specific language structures and explicit teaching methods. Tasks are the central strand of TBLT and are the defined as 'activities where target language is employed to meet a communicative need' (Willis, 1996). Tasks typically employ three distinct stages; a pre task, during ask, and post task, and work to encourage learners to find their own answers to communicative problems. A variation in task design, and specifically task complexity, have been shown to alter the performance and production of L2 speech.

Much research has investigated which features of task design impact which elements of performance, and in regards to a variation in task complexity and the production of L2 speech two theories dominate; those of Skehan and Robinson. The two theories differ in their description of how they view the allocation of attentional mental resources; Skehan proposes that there is a trade-off between attentional mental resources, and for one element to improve in performance another must decline, whereas Robinson argues that performance can be mutually beneficial and it is possible to increase both accuracy and complexity on the performance of more complex tasks. Despite the theories appearing to be opposed, there are several areas of mutual agreement, and both have been operationalised in many different research studies.

Studies purporting to support both approaches have found that task complexity plays an important role in understanding the production of speech in an L2. In relation to oral fluency, increases in task complexity has largely shown decreases in fluent speech production, but the reasons for this are still debated. Task complexity has largely been linked to difficulties in conceptualising speech and with difficulties macroplanning and microplanning. However, some more resent work has suggested that task difficulty may also affect the conceptualisation and formulation stages of speech production, and it is often hard to distinguish between the separate impacts on either. A distinction can be drawn between the different ways in which L2 speakers may produce fluent, or disfluent, speech; difficulties with linguistic encoding in speech formulation, and losses of fluency resulting from conceptual planning difficulties in conceptualisation. It is here where comparisons with L1 speech may highlight the differing aspects of production impacted for L2 speakers.

In relation to the production of fluent L2 speech it is important to explain that there are multiple ways for this to occur; with various techniques learners are able to employ to produce fluent speech. First among these are the strategies, often employed in communicative teaching approaches to SLA, where by learners may produce fluent speech through the implementation of certain skills. In contrast to this is a more traditional way for learners to produce fluent speech, by having a comprehensive knowledge of the language. This distinction between knowing a language well, and being able to use it well, is important to understand the different ways in which fluent speech can be produced by L2 speakers.

Recently Skehan et al., (2016) have proposed an idea of 'discourse' and 'clause' based fluency, where breakdowns in fluency can be directly linked to the stages of speech production, and to macroplanning and microplanning. The theory has some areas of uncertainty, particularly in regards to microplanning, and how to prove that pauses at the mid clause are directly related to problems of formulation. However, the theory does raise some interesting ideas and offers a fresh insight into how task design may impact oral fluency. The present study seeks to address this theory, while investigating whether the input format of a task impacts fluent speech production, and offering possible explanations for why this may be, in terms of the different components of speech production and the attendant mental processes.

4 Working Memory and Phonological Short Term Memory

4.1 Introduction

In this chapter I will focus on working memory (WM) and phonological short term memory (PSTM), and the key role they play in speech production, and language learning. However, I will first begin with a brief introduction to WM to show its importance to research in the fields of education and second language acquisition (SLA), but also to locate PSTM in the wider context of WM. The chapter will then provide a detailed description of Baddeley's (1986) multi component framework which employs a model of WM as well as explanations for each of the components of the model: the phonological loop, the visuospatial sketchpad, the central executive, and the most recent addition the episodic buffer. A brief explanation of other models of WM, and how they compare with Baddeley's model will then be discussed. The chapter will then move on to an introduction of PSTM and where it is situated within the overall framework of WM, as well as the importance PSTM plays in language learning and proficiency. The chapter continues with a brief exploration of how PSTM may impact oral fluency through a review of previous research. A brief summary follows of the measures employed to measure PSTM and how they differ, details of the measures employed for the present study and how they were selected. The chapter concludes with details of the implications for the present study.

4.2 Working Memory

WM can be defined as a complex multicomponent system responsible for the real-time maintenance of information during cognitive processing (Conway, et al., 2005). The importance of WM to task performance in particular is underlined by Conway et al. (2005) who describe WM as a system made up of several components responsible for the online maintenance of information flow, despite interruptions, other ongoing mental process, and any other distractions. The continual maintenance of information is handled by either domain-specific storage and rehearsal processes, or by domain-general executive attention. The extent to which each functions, is dependent on multiple factors including the context of the task, the individual learners' ability, and the number of other cognitive processes involved (Conway et al., 2005). This description underlines the role

WM plays in speech production and the way it may impact on other cognitive processes and attendant mental resources. Importantly to this study working memory capacity (WMC) is also believed to have an increasing effect on task performance as the cognitive complexity of a task increases (Robinson, 2012).

4.3 Working Memory and Phonological Short Term Memory in an L2

Working memory has also been defined as the ability to mentally store and manipulate information relevant to a task (Martin & Ellis, 2012). Studies into WM fall into two broadly separate approaches in research, one British and one North American. Unfortunately, the distinctions between these two approaches are not always apparent, or made clear in the literature (Williams, 2012). WM typically refers to the storage and the processing of information which is measured in listening or reading span tasks, whereas phonological short term memory (PSTM) is widely considered to refer to memory storage alone and is often assessed by; non-word repetition tasks, which require participants to repeat pseudo-words, or digit span tasks, which ask individuals to reproduce sequences of digits in increasing length (Van den Noort et al., 2006; Kempe & Brooks, 2011).

PSTM can be defined as the ability to recognise and remember phonological elements as well as the order in which they occur (O'Brien et al., 2007) allowing the listener to identify words as well as syntactic structures. Phonological memory facilitates language learning by temporarily retaining phonological data until permanent representation can be created in long-term memory storage. Is has been measured and discovered that phonological data traces can be held for up to 2 seconds, at which point they are lost, unless they are refreshed by sub vocal rehearsal (Baddeley, Thomson, & Buchanan, 1975). PSTM works as a holding mechanism that keeps phonological information available for a short amount of time, allowing subsequent more complex processing that follows (Kormos, 2006).

It is possible then, that the primary purpose of the PSTM system is to aid in the facilitation of the learning of new words, in both a first and second language (Baddeley, Gathercole, & Papagno, 1998). Many studies have demonstrated how non-word repetition tasks act as an accurate measure for predicting the capacity of participants to retain new vocabulary items in an L2, and those learners with larger PSTM capacity have been proven to successfully recall more newly learned vocabulary in the target language (Papagno, Valentine, & Baddeley, 1991; Ellis & Bywater, 2004; Williams & Lovatt, 2005). The ability of individuals, to create accurate phonological representations is an integral part of both short-term storage and long-term retention, required for language acquisition (Service, Mauri, & Luotoniemi, 2007). It is possible then that people with higher levels of PSTM capacity benefit through their ability to temporarily store more phonological information, for a limited amount of time, and are then able to better transfer this to their long-term memory storage. PSTM plays an important role in learning a language and acquiring new linguistic elements, but there is also evidence to suggest that the link between short term memory, and long term knowledge operates both ways (Hulme, et al., 1991; Gathercole, 1995). Research has been conducted that displays long term knowledge and understanding of a language has an influence on processing in PSTM (Gathercole et al., 1997; Kormos & Sáfár, 2008). An important point to note is that the majority of these studies conducted research in PSTM in an L1, or were focused on the acquisition of new vocabulary in an L2. Relatively few studies have investigated the role of PSTM in L2 speech production. A detailed review of two such studies is discussed in section 4.8.

4.4 History of Working Memory

The construct of memory has long been explored in literature through the dichotomous view of ‘primary’ and ‘secondary’ memory stores since James (1890) first used the terms. Despite the ongoing debate surrounding the convergence or distinction the terms short term memory (STM) and long term memory (LTM) are still widely used. STM is responsible for memories that are *active* and are presently being attended to and are those memories which we are currently aware of, LTM on the other hand deals with the long term storage and retention of information which is not currently being attended to. This distinction between STM, which Baddeley (2003, p. 830) argued was based on “temporary electrical activation” and LTM which was based on “neuronal growth”, is referred to as a dual-store theory. It had previously been assumed that the short term memory store also functioned as a working memory which was essential for not just retrieving memories, but for learning as well. Research by neuropsychologists into patients with mental impairments however showed some patients maintained the ability to recall long term memories, whilst being unable to form new short term memories. It was further discovered that patients who had STM impairments could still learn, and were capable of forming new long term memories (Baddeley, 2003). This evidence provided the starting point for Baddeley to design his model for working memory, which describes WM as both a utility for storage and an attentional component. Engle et al., (1999) argue that WM and STM can be differentiated by the links working memory has to fluid intelligence which STM does not share. The differences drawn by Engle et al., (1999) between WM and STM are both numerous and complex, and will not be covered here in any further detail as they are not directly relevant to the present study.

4.5 Baddeley's Multi-Component Model

Baddeley and Hitch's (1974) original framework for working memory theorised that STM provided an active workspace for the processing of any old or new memories. LTM on the other hand, was theorised to be a much larger and mostly passive storage system, comprised mainly of procedural and declarative memories. Procedural memories are those actions or processes which are repeatedly carried out (e.g., tying your shoelaces), and through frequent use have become stored adequately, and as a result are activated automatically without requiring conscious mental attention. Declarative memories are comprised of either episodic memories, which are autobiographical in nature, or semantic memories, which pertain to factual information. Unlike with procedural memory storage, which is a largely automatic process, declarative memory requires conscious effort to attend and access.

Baddeley and Hitch's original model (1974) was made up of three components; the first being the central executive, the second the phonological loop, and finally the visuospatial sketchpad. The central executive is described by Baddeley as a 'control system of limited attentional capacity', and is supported by the other two subordinate storage components which are each responsible for separate elements. The phonological loop deals with the processing of sound and language, and the visuospatial sketchpad is responsible for dealing images and dimensions. Baddeley's model is cited in all of the literature investigating WM, and it is the foundation for much of the research in to WM and PSTM. As such Baddeley's model is central to the understanding of working memory and the way in which it emphasises the combination of both mental processing and storage. The model also underlines the importance of detailing and defining a system which facilitates a wide variety of different cognitive processes, examples include understanding, reasoning, as well as learning (Baddeley, 1989). More recently Baddeley (2000) has described another component which has now been added to the previous model; the episodic buffer. The next section will continue with a description of each of these components separately in additional detail below.

4.5.1 Phonological Loop

The phonological loop is further subdivided into two constituents; the first being the short term phonological store or buffer where traces of memories are stored for a few seconds, and a subvocal rehearsal process which functions to revive these memory traces (Jacquemot & Scott, 2006). Evidence for the articulatory rehearsal process is provided by studies which show a decline in the ability to recall words as the length of words increases (Baddeley, et al., 1975). Neuropsychological research has also examined the differences between anarthric patients; who are unable to speak, and dyspraxic patients; who are unable to make speech motor connections. Research has shown that those patients with anarthric impairment have no difficulties processing words which are presented visually, while the patients suffering dyspraxic problems are unable to process these words (Baddeley et al., 1975). This has led to the conclusion that 'subvocal rehearsal is not dependant on the capacity for peripheral control of articulation' (Baddeley, 2003). The capacity of the phonological loop in an individual can then be used as a good indicator to measure their ability to acquire a second language, and Baddeley argues that this language development is assisted by the phonological loop in two ways. The first way is the "relatively unconstrained temporary representation" for any new phoneme sequences which the store provides, and the second being any new sounds which can be represented using the existing input process, allowing the articulatory system to facilitate learning through rehearsal (Baddeley, 2003, p. 833). In other words, the phonological loop can facilitate the learning of new linguistic material by storing sounds in the short term and allowing them to be recalled in the short term.

While Baddeley explains how the phonological loop can assist language learning, it should also be noted that the process is not without complications which may interfere with the process of the phonological loop. The first factor which may work against the phonological loop is the complexity of phonological information and the similarity of sounds in the target language; examples of similar sounds in English include f and v, l and r, as well as b and p. The similarity of the sounds, and the difficulty with pronunciation, can affect recall performance as it increases the phonological complexity of the articulatory plan as well as the stimuli with which it is associated (Jacquemot & Scott, 2006). The length of words in the L2 may also cause problems, with longer four or five syllable words obviously being more difficult to recall than those with one or two syllables, as they require longer rehearsal times (Baddeley et al., 1984). Articulatory suppression, the process whereby the act of speaking inhibits the phonological loop, can also inhibit the recognition and retention of new linguistic data through the obvious limitations of both recalling sounds, while at the same time creating utterances. The addition of extra phonological information can result in elements fading from the loop before they have been practiced and reactivated (Gathercole, 2006).

It is important to note that input provided for the phonological loop can be in the form of either aural or visual stimulus, and each will be analysed separately, at least initially, by the relevant process. Auditory input is directly processed in the phonological store and then moves to phonological short term storage and then to the output buffer where input can be recalled or rehearsed. Visual input however, must first undergo visual analysis and is then moved to short term storage, after which it gains access to the phonological store through a process of orthographic to phonological recoding, or rehearsal. The phonological loop, it has been argued, came into use through the need to acquire language, and allows this acquisition to take place (Baddeley, 2003). Evidence for this argument comes from the investigation of patients who have specific and isolated deficiencies within their phonological loops; being unable to learn linguistic items from a new vocabulary, but still maintaining functioning verbal long term memory. In summary, the capacity of an individual's phonological loop can act as a strong predictor of their ability to learn a language, and this is especially true of children and younger learners (Baddeley, 2003). This is because the phonological store allows for the "unrestrained temporary representation for new phoneme sequences" and the articulatory system allows for the learning of new words and sounds through repeated rehearsal (Baddeley, 2003, p. 833). For an understanding of the cognitive processes of speech production it is important to note that the phonological loop has been shown to involve processes which involve both the production and perception of speech. As Jacquemot and Scott (2006, p. 481) conclude: “The phonological buffer is involved in the storage of the phonological input and the subvocal rehearsal requires the inner production of a speech output.” In relation to the design of the current research project and the variation in task input format, it is conceivable that the linguistic elements present in the written input format task, might provide some level of interference to the phonological loop through the reactivation of words. The language input provided may cause participants to pause speech production as they refer back to the stimulus, and read the input of the task. This will of course no not be true for the pictorial input task, as it contains no language to process in the phonological loop. For the more complex version of the task this problem may be exacerbated, as the participant will have to constantly refer back to the written input for task transaction. As little research exploring the impacts of task input format on the specific elements of WM and speech production has so far been conducted, it remains to be seen what effects these variations in task design may have on L2 oral fluency.

4.5.2 Visuospatial Sketchpad

The second storage system central to working memory in Baddeley’s (1986) model is the visuospatial sketchpad, which also has a limited capacity. The visuospatial sketchpad is responsible for the retention of items labelled either by their location, shape, or colour. Much like the phonological loop it also requires the attention of the listener to remember the items stored within it. It has been argued that a further division between the visual and spatial components of the store is necessary, much as the phonological loop is subdivided between the articulatory rehearsal process and the phonological store. The two components have proved problematic to differentiate in research with Corsi block tasks (a short term visual WM test) often used to determine the extent of participant’s spatial memory, and visual pattern tests (designed to measure short term visual memory without the spatio-sequential component) often being employed to gauge visual memory abilities among participants (Della Sala, et al., 1999). However, there is often overlap within the two types of test and it is hard to be certain that the tests are focused on and measuring one aspect completely in isolation. It is possible that both components are utilised at the same time during performance on the same memory task. Baddeley points to findings in more recent studies where participants use coding in the phonological loop for item storage while at the same time manipulating them 'visuospatially' (Baddeley, 2003). The visuospatial sketchpad is also of importance to the acquisition of semantic information, on how objects are used as well as how they look, and as such is of importance to understanding how learning is processed (Baddeley, 2003). It is clear however that more research into the examination of the visuospatial sketchpad is required to help better understand this integral part of working memory.

In regards the current study and the variation in task input format, the task with a pictorial input will first be examined in the visuospatial sketchpad. Similar to the way in which the written input task will be re-examined, the pictures will also need to be 'rehearsed' and held in the visuospatial sketchpad for the completion of the tasks and the production of the narration. Whether or not there is any difference in access or storage between the two systems means has yet to be fully investigated, but the study will seek to explore a comparison of the differentiation in these two input format types.

4.5.3 Central Executive

The central executive acts as the controlling and regulatory element of working memory. It allows for attention to be directed toward, or away from a specific feature of input, it also connects working memory to long term memories, and acts to facilitate attention allocation (Baddeley, 2003). The central executive has a limited capacity and is served by a supervisory activating system (SAS). Baddeley (2003) provides support for the existence of the SAS through research conducted into patients who have frontal lobe impairment of the brain. Patients presenting with frontal lobe damage typically suffer from what Baddeley (2003, p. 835) describes as 'excessive distractibility' and have reduced SAS function. A primary function of the central executive then is to govern the attentional resources of a person and exercise self-control. It has also been argued that working memory capacity is important in the role of conscious awareness, with much research conducted into providing a better understanding of this key component of working memory (Baddeley, 2003).

4.5.4 Episodic Buffer

The episodic buffer is the latest component to be added by Baddeley (2003) to complement the central executive. It is a limited capacity system, which is passive in nature, and it functions to facilitate the linking of information across domains (Baddeley, 2000). It allows for the interaction of the two previously described components; the phonological loop and the visuospatial sketchpad, and also allows the integration of visual, spatial, and verbal information with time sequencing or chronological ordering (Baddeley, 2000). Possible examples could include the memory of a story, or of a particular scene from a television programme or film. Baddeley (2003) has detailed how the episodic buffer may function as the storage section for the executive control, with conscious awareness providing ease of retrieval.

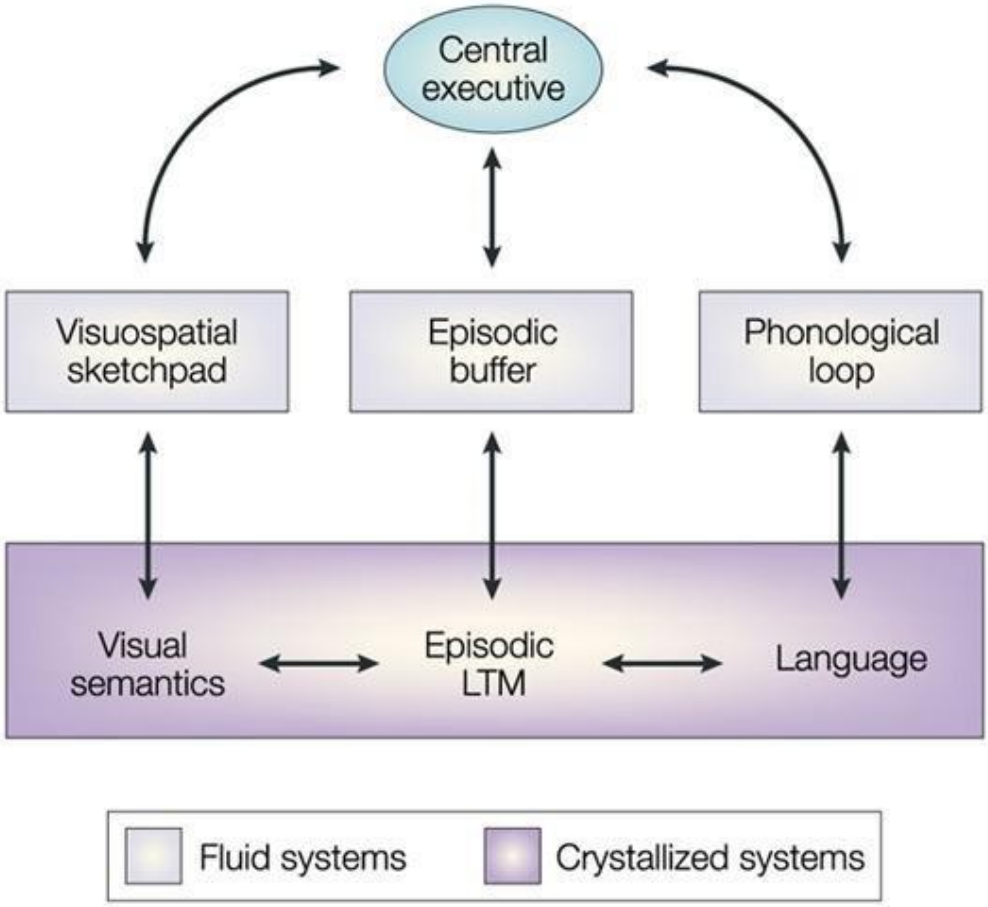


Figure 4-1. *The Multi-Component Working Memory Revision (Baddeley, 2003).*

Baddeley's (2003) updated model figure 4-1 above displays clearly how the central executive is responsible for all of the components which receive input. The visual semantics, episodic long-term memory (LTM), and language form long-term knowledge, or what Baddeley explain as more permanent 'crystallised' knowledge. The short-term memory subsystems; the visuospatial sketchpad and the phonological loop are connected with the LTM sub-systems through the interface of the episodic buffer. This allows intervention and the attentional resources to be controlled and to be designated by the SAS (Baddeley, 2003).

The present study seeks to explore the impacts of a variation of task design on the production of fluent speech in an L2, and to what extent these variations are mitigated by higher levels of PSTM. It is possible that having the ability to retain, and deploy larger amounts of data, held in the phonological loop or the phonological store provides an easing of the demands on speech production. Whether this is more apparent on the more complex tasks, or on the tasks with a variation in input design is difficult to predict due to the lack or research in the area. However, the current research will seek to fill this gap, and explore the relationship between levels of PSTM and the ability to produce fluent speech across a variation of task designs.

4.6 Comparison with Other Models of WMC

Other models of WMC have been proposed, and the most relevant to this paper as well as the most widely cited is Just and Carpenter's (1992) model which focuses more on language comprehension and the differences between L2 language performance as a function of their WMC. Understanding and comprehending language necessitates the memory storage of language components, both complete and incomplete, and is a task which requires complicated information to be mentally processed (Just & Carpenter, 1992). Baddeley's model (1986, 2003) and Just and Carpenter's (1992) model have similarities in the view of an 'executive' which controls attentional resources; the Central Executive (Baddeley), and the Executive Function (Just & Carpenter, 1992). The differences between the models are the way in which they view and detail WM, with Baddeley's model providing multiple components, conversely it is viewed as a single powerful system by Just and Carpenter. This difference has led to Baddeley's model often being employed to investigate functions and implications concerning the phonological loop, and the Just and Carpenter model has typically been used in research investigating the operationalization of the executive function. As the present study focuses on speech production and the impacts of levels of PSTM on pause location it will be explained in relation to Baddeley's (2003) updated model of the working memory system.

4.7 Evidence for PSTM in Second Language Learning Research

Much research has focused on the importance of PSTM and the vital role it plays in vocabulary acquisition in children, as they learn their first language (Gathercole & Baddeley, 1989a, 1990b; Gathercole et al., 1992, 1997; Ellis, 1996). These studies and several others have evidenced that a relationship exists between PSTM and learning in a second language. The importance of the role PSTM plays in successful language learning then is an area which has seen increasing amounts of research with many studies exploring the extent to which PSTM can predict the successful learning of a second language (Cheung, 1996; Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995).

Service (1992) assessed the PSTM skills of L2 English learners before they began their studies, operationalising PSTM through the use of English sounding non-word repetition. The performance of individual participants on this initial task was found to provide a strong prediction of their English proficiency, when it was corresponded with their eventual attainment three years later. The findings were viewed by Service to be evidence of the contribution of PSTM to the acquisition of vocabulary in a foreign language. However, a study by Gathercole et al., (1992) showed that PSTM scores provided little evidence of second language learning ability. It is important to note however that both of these studies examined younger learners; 9-10 year olds in Service (1992), and 5 year olds in Gathercole et al., (1992), so it could be that PSTM is more relevant to those learners in their early childhood, but especially when acquiring new vocabulary (Gathercole & Baddeley, 1993). PSTM has been proven to play a key role in facilitating second language acquisition, and it is important in the early stages, but as learners improve their L2 abilities the role of PSTM may diminish (Gathercole, et al., 1992). A similar pattern is reported by Baddeley (1991) who points out that as children start going to school, and begin formal education, the increase in their vocabulary sees the role of PSTM become less and less important. Several other studies (French, 2006; Williams & Lowatt, 2003; Williams, 1999) have found similar results and conclude that as learners develop their language skills and vocabulary, either in their first or second language, their reliance on PSTM decreases and the role it plays becomes less important.

Service and Kohonen (1995) later explored other aspects of foreign language learning and found that skills such as written production were less strongly associated with measures of PSTM. These studies (Service, 1992; Service & Kohonen, 1995) have been used as evidence into PSTM's impact on the learning of vocabulary in a second language, and provide support for the view that a lack of PSTM skills may act as a constraint to individual's language learning, particularly in the early stages of L2 acquisition (Gathercole & Thorn, 1998).

The importance of PSTM and the impact it has is clearly crucial to efficiently and effectively recognise and recall phonological elements and the order in which they occur (Martin & Ellis, 2012). When interacting in a foreign language the listener has to recall phonological elements, as well as their correct order, over quite short periods of time. This recall is a vital skill if they are to successfully take in and understand the vocabulary and grammar they receive as input in the foreign language. This recall is clearly more problematic in an L2, and will require more online processing than in an L1 as these processes and the linguistic recognition required would be far more practiced for L1 speakers, and likely require less conscious effort and fewer attendant mental resources.

In a second language the recognition of sounds may be problematic due to the lack of similar sounds in the L1, and possible variations in the phonology of the target language (O'Brien et al., 2007). This makes the ability to retain input for a short amount of time critical to enable language processing to occur. PTSM then plays an integral part in not just language comprehension, but also in the acquisition of new vocabulary and language forms (Martin & Ellis, 2012). The inability to correctly recognise input sounds will clearly represent a great barrier to not just short term comprehension and communication goals, but will also make it difficult to retain vocabulary which is vital to learn new words as well as their attendant grammatical structures. However, when knowledge of a language grows so too does PSTM capacity, developing in tandem and reducing the learners need to rely on PSTM. Learners with greater levels of PSTM then have a possible advantage, especially at lower levels of language proficiency, and at the earliest stages of language acquisition. PSTM is important then not just in the recognition and comprehension of L2 input, but also arguably vital in the initial stages of learning a language, and for those with lower levels of proficiency. The important roll which PSTM plays is well established in the literature on aptitude for SLA (Robinson, 2003; Skehan, 1998).

4.8 Phonological Short Term Memory and Speech Production

Few studies have so far explored PSTM and how it impacts oral fluency in speech production. This is perhaps due to the predisposition in the literature to focus on WM, and on language acquisition and proficiency more generally, this is a somewhat understandable approach considering WM's origins in exploring L1 abilities. However, it may also be due to the interrelated nature of WM and PSTM, and the difficulty inherent in reliably disentangling the two in terms of direct links to speech production (Kormos & Sáfár, 2008). There is also research showing evidence that tests examining PSTM capacity and instruments measuring complex verbal working memory capacity are tools that explore the same underlying constructs (Kormos & Sáfár, 2008). Research in this area then is not without difficulties and challenges.

Most Research into PSTM has tended to focus on vocabulary learning, and language acquisition in general, but few studies have focused on phonological memory and its role in immediate performance. A recent study conducted by O'Brien, et al., (2007) investigated the role of phonological memory in the development of oral fluency. The paper operationalised PSTM through serial non-word recognition (SNWR) with L1 English speaking participants, all of whom were attending an English speaking university. They were assessed at the beginning and the end of a university semester using the SNWR test, and were rated for measures of oral fluency in spoken Spanish. Results of the study showed that "…phonological memory was implicated in the oral fluency gains of adults learning Spanish as an L2 despite the fact that the SNWR scores did not change over the same time span." (O'Brien, et al., 2007). It was also noted that the levels of phonological memory acted as a good predictor for the participants' level of oral fluency at the end of the semester, evidenced by a number of measures. The study concluded that phonological short term memory was an imperative and necessary component, integral to the successful learning of a second language and the development of L2 speech production in adults, but also cautioned that further studies will be required to look at exactly how and why WM and particularly PSTM impact oral fluency performance in language learners producing a second language (O'Brien et al., 2007).

A more recent study by Suzuki (2021) explored PSTM and its ability to predict changes in the development of L2 oral fluency, through the analysis of differentiation in pause locations. Suzuki (2021) highlights the importance of differentiating between pause locations, as it allows for mid clause and end clause pauses to be linked to different cognitive process involved in speech production (De Jong, 2016; Kahng, 2018; Lambert et al., 2017; Skehan et al., 2016). which may be drawing on differing aptitude components.

Another study conducted by (O'Brien et al., 2006), found that there was a significant correlation between higher PSTM scores on a non-word test, and fluency scores derived from a spoken interview. The study provided evidence that WM and PSTM develop differently and may lead to different types of learning difficulties (O'Brien et al., 2006). The study also suggests that WM and PSTM play a different role in the successful acquisition of an L2, with WM linked to language proficiency in beginner level learners, and PSTM linked to the fluency of more proficient speakers. A possible theoretical explanation offered for this is that as language learners increase in proficiency, they develop more accurate and proceduralised, and perhaps even automatized, grammatical processing (O'Brien et al., 2006). This provides a potential theoretical link between PSTM and speech production in an L2, and suggests that as the speaking skills of proceduralisation and automatization and improved PSTM memory plays a larger part in the ability to produce fluent speech. These theories highlight several areas of specific interest to the current research project which can conceivably be linked to the speaking skills detailed in previous sections; Formulaic language and chunking (2.14.2), language processing skills, (2.14.3), and proceduralisation and automatization (2.9.1.1). Are increased levels of PSTM indicative of the ability to access larger amounts of formulaic language? It is likely that more proficient learners have access to more ready-made lexical formulae (O'Brien et al., 2006), and are able to produce more fluent speech through the deployment of these chunks. Do higher levels of PSTM indicate that more proficient learners can tap into the cognitive processing skills of proceduralisation and automatization more effectively? Increasing fluency through the smoother delivery of speech due to easing the cognitive load in the formulator. Or finally, do higher scores on PSTM tests indicate that learners have faster access to the increased phonological knowledge of the target vocabulary, that more proficient leaners possess? In relation to the present study, it will be interesting to see to what extent higher levels of PSTM mitigate the impacts of task complexity and task input format, and how this may be explained in reference to speech production, and learner use of the above mentioned skills.

4.9 Measures and Operationalization of PSTM

An immediate serial recall test was used by Williams and Lovatt (2002) in their study into the impacts of PSTM on grammar learning. The test involved assessing the participant's short-term storage capacity with the as yet unknown vocabulary to be used in the experiment. The test involved the immediate recall of a list of five vocabulary items in the same order as they were presented. Williams and Lovatt (2002) employed this slightly unusual approach to measuring PSTM because as they point out previous research has shown non-word repetition tests with immediate recall depend largely on how similar the words are to words the subjects already know. So, to more accurately capture the participants' short term storage capacity for the vocabulary in the task, and the ability to relate this to rule learning the test was conducted using those words. It was also noted that the words were too short to form a non-word repetition test, providing further support for the selection of the immediate serial recall test.

Interestingly O'Brien et al., (2007) chose a serial non-word recognition test as they believe it offers the most accurate view of PSTM compared to non-word repetition and serial non-word recall, the two most popularly employed measures of PSTM in research. O'Brien, et al., (2007) argue that SNWR benefits by not having a spoken production element which means it will not confound with L2 speaker measures of oral fluency. It is conceivable that a confound could occur due to "mutual output constraints in which associations between speech production measures and phonological memory measures requiring verbal responses reflect common articulatory output requirements." (O'Brien et al., 2007, p. 563). This criticism has also been levelled at non-word repetition, as well as serial non-word recall and non-word span. The argument is that as these tests all require extensive articulatory processes the interpretation of results can be problematic due to the uncertainty of what specifically is being measured; the results may be a measure of short-term memory or they could rather point to the mental processes involved in output demands, or even some combination of the two.

As was mentioned in the previous section the way in which speech is produced, and the underlying nature of the speech production process, is important to understanding the role of PSTM in fluency. The theoretical justification offered by O'Brien et al's., (2016) findings; that with increased proficiency comes greater proceduralisation of language production, and that this is related to the impact of levels of PSTM and they role they play in fluency for more proficient learners, is extremely interesting. Higher levels of PSTM may provide greater access to formulaic language, and may facilitate the proceduralisation of language, a more effective PSTM may also indicate that a learner has more linguistic knowledge and is able to produce it faster.

4.10 PSTM, Cognitive Complexity, and Oral Fluency

In regards to the present study it will be of utmost interest to see if a similar correlation to O'Brien, et al., (2007) can be discerned, and whether or not the participants with higher measures of PSTM correlate with participants having increased rates of oral fluency, both on the tasks in general and more specifically on the cognitively more complex tasks. The impact the increased cognitive load has on levels of oral fluency will also be of great interest to this study; will the participants with better PSTM scores be able to mitigate the effects of the more cognitively complex task and produce more fluent speech then those participants with lower levels of PSTM? It is hypothesised that the L2 speakers who have 'better' levels of PSTM will be able to produce more fluent speech during the cognitively more complex versions of the tasks, when compared to those second language speaking participants with 'lower' levels of PSTM. This could be due to the fact that phonological memory facilitates language learning through 'holding' phonological traces temporarily until more permanent representations can be created (O'Brien et al., 2007). This ability to hold onto more phonological information may provide an opportunity to mitigate the impacts of cognitive complexity within a task, and could help second language learners maintain levels of oral fluency throughout the various stages of speech production.

4.11 Conclusion

In this chapter I detailed a brief introduction to working memory and how integral it is to research in the fields of both education and psychology. The focus of the chapter then shifted to a review of previous research conducted on the concepts of working memory, and phonological short term memory. Baddeley's multi component framework and the way in which it provides a model of working memory, and each of its components was explained. A comparison of other models of working memory, and how they compare with Baddeley's was then detailed. The chapter then explained how phonological memory fits into the overall framework of working memory, and the importance phonological short term memory plays in language learning. The chapter then looked at how phonological short term memory has been shown to effect oral fluency in previous research. A brief summary followed of the methods employed to measure phonological memory and how they differ. Finally, the chapter concluded with details of the relevance to the present study and the possible correlation between L2 speaker levels of PSTM and task performance, evidenced by measures of oral fluency and the ability to produce more fluent speech.

5 Methodology

5.1 Introduction

In Chapter 2 I explained how the understanding and analysis of L2 oral fluency, as well as speech production and its attendant mental processes, has been advanced through the considerable amount of recent research conducted. More valid and accurate measures have been identified to reliably measure L2 fluency, and advances in the software used for data analysis has greatly increased the ease with which large amounts of data can be consistently analysed. More interestingly, and of more importance to this study, research is approaching a better understanding of the specific speech characteristics which directly relate to certain cognitive mental processes which provide the foundation for L2 speech production. The research questions and the predictions provided below aim to develop a more complete understanding of these speech characteristics, and how they are impacted by the mental processes at work during L2 speech production.

In this chapter I will first present the research questions that have been posed by the literature review in the previous chapters and then state the predictions. An explanation of the design of the study follows, with details of the dependent and independent variables measured. The chapter then proceeds with a detailed account of the participants and the research site involved in the present study. The following section comprises a description of the instruments; the tasks themselves, including details of how task complexity has been operationalised for this study. This is followed by a brief section on the quantitative and qualitative date obtained from the L1 participants to support the operationalisation of task complexity employed. I them move on to explain the organisation and details of the procedures employed for data collection. The way in which the data was coded, through the use of the PRAAT and R software programmes is then detailed, and data analysis is described. A brief description of implementation of linear mixed effects models follows. Focus then moves to PSTM and the way in which a digit span tests was employed post task, to asses levels of L2 speaker PSTM. Finally, this chapter provides a report of the ethics requirements that were met, and concludes with a summary of the pilot study explaining how it informed the present research, the data collection, and its methodological format.

5.2 Research Questions and Predictions

The research project sought to address the gap in knowledge in the effects of task cognitive complexity on L2 speaker speech production and oral fluency, viewed through a comparison with L1 speaker oral fluency, as measured at both the mid-clause and end-clause position. The impacts of a variation in task input mode (written or pictorial), and levels of phonological short-term memory (PSTM), on the cognitive processes of speech production, again measured through the frequency of pauses at the mid and end-clause were also explored. The questions that guided this research project are detailed in the following section and the focus on the effects of increased task complexity with a distinction between mid-clause and end-clause pauses. I will outline the pedagogic rationale for the research, and conclude the section with a description of the predictions and what the study aims to identify.

RQ 1. What is the impact of a variation in task cognitive complexity on the performance of L1 and L2 speakers' oral fluency measured at the *mid*-clause?

RQ 2. What is the impact of a variation in task cognitive complexity on the performance of L1 and L2 speakers' oral fluency measured at the *end*-clause?

RQ 3. What is the impact of a variation in task input format (written or pictorial) on the performance of L1 and L2 speakers' oral fluency at the mid-clause and end-clause level?

RQ 4. What is the relation between L2 speaker's levels of Phonological Short Term Memory (PSTM) and their measures of oral fluency at the mid-clause and end-clause level?

The predictions for each research question are detailed below, followed by an explanation and rationale drawn from the literature.

RQ 1. The increase in task cognitive complexity will result in a reduction of oral fluency for L2 speakers at the mid-clause level, evidenced by more frequent mid-clause pauses. L1 speakers' levels of oral fluency will be unaffected at the mid-clause level, and their pause frequency will be unaffected (Skehan et al., 2016; Tavakoli, 2011; Kahng, 2017).

RQ 2. The increase in task cognitive complexity will result in a reduction of oral fluency for both L1 and L2 speakers at the end-clause level, evidenced by more frequent end-clause pauses. However, end-clause oral fluency will be less affected than mid-clause fluency in L2 speakers (Skehan et al., 2016; Tavakoli, 2011; Kahng, 2017).

RQ 3. The task with written input will see higher levels of oral fluency at the mid-clause level amongst L2 speakers when compared with their performance on the pictorial input task. It was predicted that this would result in more frequent pauses at the mid-clause on the task with pictorial input. L1 speakers' levels of fluency will be unaffected by the input format (Feng & Webb, 2019; Rast, 2008).

RQ 4. The L2 speakers who have lower scores on the PSTM test will record lower levels of oral fluency and pause more frequently at the mid-clause level. Levels of PSTM will not have a correlation with the frequency of pauses at the end-clause (Baddeley et al., 1984, O'Brien et al., 2007, Van de Noort et al., 2006, Williams 2012).

The predictions stated here are based on research which has looked at task complexity (Skehan, 1998; Robinson, 2011b; Tavakoli & Foster, 2008; Kormos & Trebits, 2012), speech production (Levelt, 1989; Segalowitz, 2010; Kormos, 2006), oral fluency (De Jong, et al. 2012; Préfontaine & Kormos, 2015; Tavakoli, 2016; Kahng, 2014), task design in task based language learning (Hoey, 2012; Long & Norris, 2009; Long, 2015; Rast, 2008; Nel de Jong & Vercellotti, 2016), and phonological short term memory (Baddeley et al., 1998; O'Brien et al., 2007; Van de Noort et al., 2006; Williams, 2012). It is proposed that L2 learners will have lower measures of oral fluency at the mid-clause level, as evidenced by the greater number of pauses, compared to L1 speakers (Skehan et al., 2016). This is predicted due to the increase in task cognitive complexity which results in increases in microplanning during the formulation stage of speech production, and will be evidenced through the more frequent pauses of L2 speakers recorded at the mid-clause level. This increase in pausing, and the resultant dysfluency, is due to mental processing resources being diverted toward information organization for task completion and away from microplanning, reducing the amount of mental resources available for language formulation. Both L1 and L2 speakers may have problems organising the information in the more complex versions of the tasks; resulting in greater pauses at the end-clause phase for both the L1 and L2 speakers of English. These breakdowns, and dysfluencies are thought to be as a result of increased demands on the macroplanning phase of language production and are thought to be linked to conceptualisation in language production (Skehan et al., 2016). However, it is predicted that L1 speakers will still be able to produce language at the formulation stage, during microplanning, due to their implicit knowledge of grammar and vocabulary, as well as the advanced automaticity skills of language production available to L1 speakers. The L2 speakers however will have problems at the formulation stage of language production and may pause more mid-clause as a result of the increased cognitive load in the more complex version of the tasks (Skehan et al., 2016, De Jong, 2016).

Task characteristics and task design has seen much research interest with findings showing that a variation in task input can affect the fluent production of speech (Kim, 2009). As the written input format task provides linguistic input is it believed that this will ease speech production for the L2 speakers through the process of ‘lexical priming’ (Hoey, 2012). By providing some linguistic content, it is predicted that the written input will ease speech production at the formulation stage, specifically in lemma retrieval, and allow for higher levels of oral fluency. This greater level of fluency will be evidenced by fewer mid-clause pauses.

The importance of PSTM on language acquisition and performance in younger learners has long been established and is well developed (Baddeley et al., 1984, 1998; Martin & Ellis, 2016), however more recent research has provided evidence to show that PSTM has an important impact on the oral fluency development of adult language learners (O’Brien et al., 2006; 2007). It is predicted that those learners who have higher scores on the test of PTSM will pause less at the mid-clause level and have higher levels of oral fluency at the formulation stage. The written task input will see an easing of lemma selection for L2 speakers and reduces the cognitive load at the formulation stage of language production. Resulting in fewer breakdowns and pauses and higher levels of oral fluency compared to the pictorial task.

5.3 Design of the Study

There are three independent variables examined in this research. The first is the complexity of the task; whether or not it is cognitively more or less complex. The second independent variable is the type of speaker; if they speak English as a first (L1) or second language (L2). Thirdly the form of the input for the task; if the task media input is in a written form or is of a pictorial form. Table 5-1 below demonstrates the categorical variations in the independent variables.

Table 5-1. *Independent Variables*

Type of Variable	Variation 1	Variation 2
Speaker type	English as a first language (L1)	English as a second language (L2)
Task cognitive complexity	Cognitively more complex task	Cognitively less complex task
Input format of the task	Written input	Pictorial input

The present research primarily employs quantitative data to more accurately measure and compare data sets across groups (Mackay & Gass 2015). Utterance fluency is measured quantitatively throughout the study using the dependent variable measure of pause frequency (PF) which is measured at both the middle and at the end of clauses. The measure of phonological short term memory was collected from a standard digit span test administered before the tasks were performed.

Many studies of task complexity have looked at the impacts of variations in task complexity on L2 proficiency measures; complexity, accuracy, and fluency, and in some instances all three. However, little research has so far examined variations in task design through an exploration of a variation in the input format of a task (Long, 2015) and how these factors specifically impact the cognitive processes of speech production. Through a comparison of L1 and L2 speaker’s oral fluency measures, recorded as pauses at the mid-clause and end-clause position, the presumed differences in cognitive processes involved in speech production will be highlighted and examined.

5.3.1 Research Design

When looking at the data generated for the present study it is important to be aware of the research design employed in this project. Each of the participants completed two tasks; one more complex version of a task and one less complex version, but also one task took a written format and one took a pictorial format. This meant that there was a total of four different tasks versions; both a more and less complex version of the written input task, as well as a less and more complex version of the pictorial input task. This resulted in only half of the L1 participants completing the more complex version of the written task and only half undertaking the less complex version of the written task. This led to one of the principal challenges of analysing the data for this study; namely that the various divisions between the different research groups have meant a like for like comparison was not always possible. There are three sets of independent variables in the present study they are: the speaker type; L1 or L2, the task complexity; more or less cognitively complex, and the task format; written or pictorial. This research design meant that for each step of analysis a decision had to be made whether to analyse the data as a whole or to divide the data in to smaller groups. For the initial analysis on perceptions of task difficulty it was decided to explore the data in each separate group first, before looking at the data as a whole. Subdivisions were made between the format of the task input, be it written or pictorial, and obviously between levels of task complexity as well.

5.4 Participants and Research Site

This section will describe in detail both the L2 and L1 participants in this study as well as detail the demographics and English language experience of the L2 participants. The collection of data from L2 participants was conducted at a South Korean university on three separate occasions, due to small class sizes and the relatively large number of participants required. The 40 L2 participants were randomly selected from six intact classes of English language students. The L2 participants are all South Korean nationals, taking part in an intensive English language program during the university semester break. The classes were taught by both English L1 speaking instructors and Korean English L2 speaking teachers, but all lessons were conducted in English. Each class comprised of approximately 10 to 16 students and was made up of both male and female students. The classes were divided into one of three broad ability levels: low intermediate, intermediate, and high intermediate. Student class allocation was determined by a pre-sessional placement test which included grammar, listening, and speaking sections. First the students were given a grammar based multiple choice test to complete under test conditions, which they had an hour to complete. They then completed a listening exam, which lasted approximately 15 minutes and involved listening for specific information as well as listening for general comprehension. Finally, the students underwent an 8 to 10 minute, face to face, one on one interview style speaking assessment, conducted with an English L1 speaking instructor. The grades for the interview were based on a rubric which assigned a score based on the fluency, pronunciation, vocabulary, and communicative competently of each participant. All of the L2 participants in the study also completed the first part of a placement test, the grammar section, to provide a more detailed picture of their English language ability. The placement test section which participants undertook included 100 English grammar questions and results were marked out of a maximum score of 100. The lowest score was 61 and the highest score was 83, the mean score was 74 and the mode was 71, meaning these students would be largely classified as 'upper intermediate – competent users' on the OPT language level assessment. This is broadly in line with an IELTS level of 5.5 and a TOEFEL score of 550. The students were fairly evenly split between males and females, with 22 female and 18 male participants. The ages of participants ranged from eighteen to sixty-two, with a mean age of 23, and a mode of 20. All participants had remarkably similar learning paths, and had followed a similar route of English education in the Korean public school system. On average the L2 participants had been studying English since elementary school for a period of 12 years. However, eight of the L2 students initially selected had studied English abroad when they were in primary or secondary school, as a result their data was excluded from this study and replaced by other randomly selected participants.

The 40 L1 speakers were recruited from amongst the native speaking students at a university in England. All of the L1 participants volunteered to participate in the study and were randomly allocated to one of the two L1 speaker groups. The majority of L1 participants were female with only 9 males and 31 females participating. Ages ranged from eighteen to sixty-four, with a mean of 20 and mode of 20.

The research environment for studies involving oral fluency, and those for task based language learning, is one area of methodology that has received much debate. The vast majority of studies investigating oral fluency have been carried out in laboratory settings as opposed to the language classroom itself (Foster, 1998). The problems facing coding data automatically for oral fluency in a classroom setting are highlighted by

Lambert et al., (2017) who reported being unable to accurately record silences due to background environmental noise. A laboratory setting was chosen for all data collection and recorded in identical circumstances; an empty classroom with only the computer recording equipment and the interviewer present.

5.5 Research Instruments for Data Collection

The next section provides a brief description of the tasks and the way in which task complexity has been operationalised, in regards to Skehan's (2009) Limited capacity Model and Robinson's Triadic Componential Framework (2007a). This is followed by details on how complexity was implemented in the design of the two versions of each task. This section continues with an explanation of the instruments used for data collection and how participants completed the tasks. Information on the makeup of the tasks and their classification is then provided.

5.5.1 The Tasks

The two versions of each of the two tasks have been designed to be similar in terms of processes and outcomes (Ellis, 2003), but are differentiated through input and a variation in the cognitive complexity required to transact the task and create the narrative (Skehan, 2009). The tasks were divided by cognitive complexity, and task input format, with one L2 group of participants and one L1 group performing a simple, or less cognitively complex version of one of the tasks and then completing the more complex version of the other task type. The other L2 and L1 groups completed the opposite of this, first doing the more cognitively complex version of the task and then completing the less cognitively complex version of the other task.

The L1 speakers' responses to a post task questionnaire, which covered their opinions of the difficulty of the tasks, were also examined to provide the study with quantitative data and evidence the validity that the tasks are in fact more complex, and that the more complex versions are indeed 'harder' to transact. The questionnaire chose to focus on the L1 speakers as the assumption was made that their answers would refer only to the cognitive difficulty or complexity of the tasks. The L2 speakers' responses on the other hand would inevitably reflect the difficulty imposed by the use of language from an L2, resulting in a confounding variable. The aim of the questionnaire was to establish the levels of cognitive difficulty inherent in transacting the task, as the language required to complete the task was the same across both versions.

5.6 Instruments

The instruments used for data collection are the two versions of the two tasks which were performed by both the L1 and L2 participants. The tasks were created and selected to be similar to classroom language activities the L2 learners are familiar with and used to encountering in language classrooms as part of their English studies. This was evidenced both from my own teaching experience in Korean EFL classes, as well as from consulting current Korean university syllabuses and course books. The tasks were designed with these factors in mind and were also formulated to be in keeping with their regular EFL learning classes and language learning processes. The tasks and their constituents selected for this research project have been chosen using Ellis' (2003) general framework of tasks, as well as being informed by Long's (2015), Robinson's (2011), and Skehan's (2003) research into task selection, task complexity, and task classification. The input selected for the first task, referred to here as the 'accident' task, is pictorial and comprises a series of eight comic strip style pictures, displaying a car accident (Appendix A), and requires participants to create a narrative based on the pictures. The images are original, and were created by the researcher, and were designed specifically to give the participants enough to say about each picture, whilst still being able to combine the images into a single narrative. This task has been chosen as it is similar to a classroom activity, that students will be familiar with, and will have encountered in their formal language learning on many previous occasions. The task also closely resembles a question which is commonly included on the TOEIC exam, which is a well-known test of English proficiency commonly administered to Korean English language students at the university level. The familiarity of the task selected will hopefully help produce consistent responses and reduce confusion over what is expected from participants, allowing them to respond comfortably and confidently and providing the maximum amount of data.

The conditions of the task saw the participant being provided with a brief explanation of what was required from them; they had to produce a narrative and deliver it to the teacher/researcher. They were required to talk about each individual picture, and had to use a minimum of two complete sentences to describe each picture, as well as to speak for a minimum of two minutes in total (Appendix F). The less cognitively complex version of the 'accident' task provided learners with eight pictures in a fixed chronological order, displaying a clear sequence of events. The participants were given the comic strip pictures and were allowed up to two minutes to prepare and become comfortable with what they would like to say, and to get ready to deliver their narrative to the researcher. The participants were then finally tasked to produce a narrative based on the pictures, explaining what they believed was happening in each of the comic strip pictures and combining that into a single narrative. Again, the minimum length was set at two minutes, to encourage participants to provide enough data for the study, and to keep it in line with the other task. While the choice of tasks does not contain an inherent or specific 'information gap' for participants to bridge, it was still felt to meet the requirements of a communicative task through the definitions in the literature. An 'information gap' exercise is only one particular type of task, and tasks can still be classified as tasks as long as they include 'communicative' and/or 'interactive' elements, and these terms cover far more than just a back and forth dialogue between participants (Long, 2015). Communication and interaction in task based learning can be carried out through the implementation of tasks with varied design and structure (Samuda & Bygate, 2008), with examples such as 'interactive reading', or 'public speaking' exercises. The task employed a monologic discourse mode primarily in an attempt to avoid the many difficulties inherent with the analysis of dialogic task performance (Tavakoli, 2015; 2016), but also to remove any variables which may arise from learner-learner interactions.

For the cognitively more complex variation of the 'accident' task the individual pictures were not ordered chronologically, but rather in a random non sequential pattern, requiring each participant to decide on the order they think the events occurred in when creating and delivering their own narrative. The pictures were fixed, as a set of eight pictures, (Appendix B) in the random non-chronological order to maintain input consistency between participants and also to stop the learners from rearranging the individual pictures in their preparation time. It is also important to note that participants were not allowed to write anything during task preparation, and were specifically warned against writing on, or attempting to order the pictures through notation. Table 5-2 below details the specific elements of the task.

Table 5-2. *Task 1 specifications 'Accident'*

Task 1 'Accident'	Description
Input	1. Medium: pictorial (comic strip) 2. Organization: tight structure
Processes	1. Discourse mode: monologic
Outcomes	1. Medium: oral 2. Discourse mode: descriptive narrative 3. Scope: closed

As can be seen in table 5-2 above the task has a tight structure and requires the production of an oral narrative, this is consistent across all four of the task variants.

The second task, the 'timeline' task, takes the form of a biographical narrative. The task provides a variation in input, this time instead of pictorial input, the participant was given several written bullet points (Appendix C), which the learner had to utilize to create a narrative about the major events of the actor Johnny Depp's life. Johnny Depp was chosen because he was old enough to provide information for several bullet points, and being well-known and familiar to both English and Korean participants. The participants had to use all of this written information and each of the bullet points to create a chronological description of Depp's life. The conditions of the task were the same as the 'accident' task with the speaker having two minutes to prepare for the task before their narration commenced. Again, note taking, physical ordering of the input, or marking of the task hand-out was forbidden to maintain both consistency across tasks, and to ensure the cognitive complexity of the task was maintained. Table 5-3 below shows that the processes and outcomes which occur in both tasks have been carefully selected to be as similar as possible, with both involving a monologic discourse mode and requiring a narrative description as the outcome of the task. This is to allow for a more accurate comparison of the fluency measures obtained (Long, 2015).

Table 5-3. *Task 2 specifications 'Timeline'*

Task 2 'Timeline'	Description
Input	1 Medium: written (bullet points) 2 Organization: tight structure
Processes	1 Discourse mode: monologic

	1 Medium: oral
Outcomes	2 Discourse mode: descriptive narrative
	3 Scope: closed

The ‘time-line’ task also differentiated the information organization through the manipulation of the temporal order of the input to increase the cognitive complexity of the task, in exactly the same way as the ‘accident’ task (Skehan, 1998). The task provided for the learners in the first, less cognitively demanding version of the task, has the bullet points numbered and listed in chronological order, for the increase in cognitive complexity the input is again in a fixed, but random order. The information is summarised below in table 5-4. This requires participants to organize and sort the information in the bullet points themselves. Again, as in the ‘accident’ task the order of the random bullet points will be ‘set’ or fixed to stop participants manually arranging them during the two minutes pre-task time (Appendix D). During the completion of the more complex task the participants will be operating under an increased cognitive load. Speech production will now require more attendant mental resources to create the discourse and produce a narrative. Processing at the conceptualisation stage of speech production and with macroplanning will become more difficult as a direct result of the task conditions (Skehan, 2009; Skehan et al., 2016).

Table 5-4. *Variation of cognitive complexity through information organisation*

Task	Information organization	
‘Accident’ (pictorial)	Random + cognitive complexity	Chronological - cognitive complexity
‘Time-line’ (written)	Random + cognitive complexity	Chronological - cognitive complexity

Table 5-4 above shows how both task variants operationalise an increase in the task cognitive complexity through the +/- order of information, with a random order of the information provided increasing complexity.

5.6.1 Establishing Task Complexity: Quantitative Data

All of the 40 L1 participants undertook a feedback questionnaire on their perceptions of how difficult each task was to transact immediately after completion of both of the tasks. The respondents were asked to grade the difficulty of each task they performed on a 10-point rating scale (Appendix E) after completing both of the narratives. Space was also given for participants to add comments to justify their perceptions of difficulty for each task. The rating scale employed was arranged with a grade of 1 representing the participant's belief that the task was extremely easy, while at the other extreme, a score of 10 showed that the participant believed the task to be extremely difficult. The data set had been subdivided into sample sizes of n = 20 as each participant transacted only two of the four possible tasks types; written hard/more complex (WH), written easy/less complex (WE), picture easy/less complex (PE), picture hard/more complex (PH). Figure 5-1 below shows a jitter plot of the results of the L1 perceptions of task complexity arranged by group.

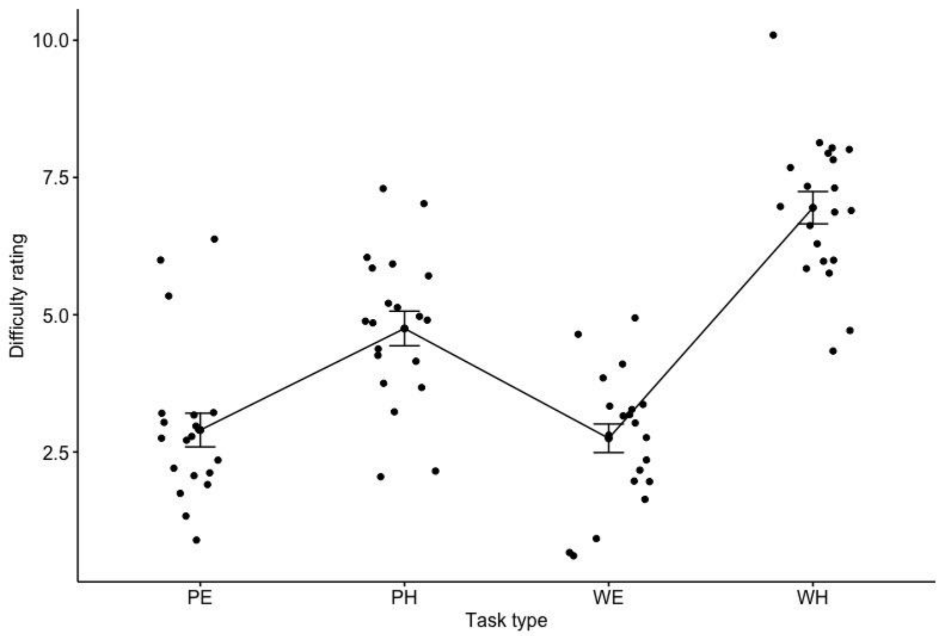


Figure 5-1. *Jitter plot of L1 perceptions of task difficulty*

Each participant would either complete the written more complex task *and* the pictorial less complex task, *or* the written less complex and pictorial more complex tasks. Table 5-5 below shows descriptive statistics for the L1 speaker's perceptions of difficulty.

Table 5-5. *Descriptive statistics of L1 perceptions of task difficulty*

Dependent Variable	Mean	Std. Deviation
Written less complex	2.75	1.16

Written more complex	6.95	1.32
Pictorial less complex	2.90	1.37
Pictorial more complex	4.75	1.41

Table 5-5 and figure 5-1 above show that the more complex version of the written task has the highest mean difficulty rating of 6.95, while the less complex version of the written task has the lowest mean difficulty at just 2.75. This represents a substantial difference between the two tasks which were identical except for the chronological order of the input. The pictorial task saw less variance between the means with an average difficulty rating of 4.75 for the more complex picture based task, compared with a mean of 2.9 on the less complex version.

The L1 perceptions of task difficulty data was checked for normality of distribution and was found to be slightly differently distributed from normal data. However, as all the data had been collected by myself, was recorded accurately, and was checked and found to be error free, it was felt that no responses needed to be removed. The accuracy of the data combined with the limited number of responses for each task type (n = 20) meant that removing any responses would impact negatively on data analysis, with the reduction of the sample size making it more problematic to draw conclusions. As a result of this and the recommendations of the literature on the removal of outliers in non-normally distributed data (Winter, 2019, Field et al., 2012) it was concluded that no data would be removed from this data set.

The L1 speaker perception results were then examined for variance between groups and a Kruskal-Wallis rank sum test was performed to ascertain the significance of the differences between responses for each task. The results showed a significant difference: chi-squared = 49.004, df = 3, p-value = < 0.001. Pairwise comparisons were then performed using a Wilcoxon rank sum test to compare the scores of different complexity levels for both task types, with the results displayed in table 6 below.

Table 5-6. *Wilcoxon rank sum test for significance of differences*

Task complexity and format	Pictorial - C	Pictorial + C	Written - C
Pictorial + C	p<.001	-	-

Written + C	-	-	p<.001

The results in table 5-6 clearly display that L1 speakers' perceptions of both the less and more complex versions of the written and pictorial task types are statistically significantly different from each other.

Further evidence of the L1 speaker perceptions of task difficulty is provided by the analysis of the questionnaires completed post task in the form of the qualitative data analysis provided in the next section.

5.6.2 Establishing Task Complexity: Qualitative Data

The post task feedback included a section where participants were asked to write comments on why they found the task to be difficult or easy to complete. The answers were coded and divided into three broad sections for ease of analysis and to better understand which elements had a bearing on the L1 speakers' perceptions of task complexity. The three sections impacting difficulty were: the order of the input, the format of the input, and the nature of the task itself. These sections were selected to best reflect the majority of responses, as every feedback form contained a comment that fitted into at least one of the coding categories.

The first group of comments were those which specifically expressed difficulty with the ordering of the input i.e.; the information was ordered chronologically and this made task completion easier, or conversely the order of the information provided was not chronological, but rather randomly presented, and this increased the difficulty in completing the task. This is of course the variable which has been explicitly changed to impact task difficulty, and is the way in which the dimension of task complexity has been operationalised for this study. Of the 40 responses, 30 of the L1 speakers wrote that the order of information affected task difficulty, either by saying that the lack of a chronological order made it harder; "*Not being in order made it tricky ...*", "*…difficult to put into chronology in head.*" (Appendix H), or by writing that the chronologically ordered information facilitated the ease of task completion; "*...it definitely helped being in order.*", "*Chronologically made sense which made the description easier*." (Appendix I). These responses provide evidence that task difficulty is impacted by the +/- order of information in the task input, and further to this suggest that the information supplied in order eased the cognitive demands of the tasks, while the non-chronological supply of information meant the task provided a more taxing mental challenge.

The second grouping of comments all reported something specifically about the difference in the input format of the task (be it written or pictorial) and how this impacted the completion of the task. These comments were found in more than half of the responses; 21 out of 40. Examples include factors which made the task easier: "*... easier to look at pictures than words*", "*The information was in bullet points which made it easier ....*" (Appendix J), and also format specific elements which they felt made the task more difficult: "*Pictures were easier to interpret*", "*I found it more difficult as there were a lot of numbers to keep in mind at once!*" (Appendix K). The evidence collected provides a consensus that input format has a discernible impact on task difficulty, and it does affect the difficulty with which the task is perceived by the majority of L1 participants.

The third section of coding focused on comments that described some difficulty with the general completion of the task, unrelated to either the order of the information or the input format of the task. There were 12 comments in the task feedback from the 40 responses that were assigned to this category. Examples include: "*Having to construct a narrative on the spot was very difficult ...*", "*It was quite difficult as it required thinking on my feet.*" (Appendix L). These results show that for some L1 speakers the task itself was problematic and difficult to transact.

In summary the results of the L1 speaker perceptions of task difficulty provide empirical evidence that the tasks, which are in a random, non-chronological order, are more complex and harder for the participants to transact. The data provided by the qualitative and quantitative analysis establishes that '+/- order of information' (Robinson, 2007) does impact L1 speaker perceptions of task complexity significantly, and it also provides evidence to define the tasks as clearly being distinguishable by the effort required to transact them, firmly establishing the tasks as either 'more' or 'less' complex.

5.7 Research Procedures for Data Collection

All of the participants were divided by their first language, therefore English (L1) or English (L2) speakers. Both the groups of L1 and L2 speakers were randomly allocated, to maintain internal validity (Torgerson, 2008), to one of two separate groups, through the process of stratified random sampling, to ensure a balanced sample of all classes was represented in the data (Cohen et al. 2013). The number of L2 students drawn from each of their classes was proportionate to the total number of students in that class. For example, in a class that had 10 students then 5, or 50%, were selected and a class that had 16 students 8, again 50%, were chosen randomly. With relatively small samples sizes the possibility of individual differences (IDs) impacting the data were possible. However, to guard against this, participants were selected from as similar groups as possible; all participates were chosen from the same university intensive summer and winter classes, and from the same broad proficiency level. The homogeneity of the Korean education system also meant that all of the participates had followed the same route through their English language studies, and the removal of any participants who had studied abroad reinforced this. It is of course possible that IDs impacted the results, but this was guarded against and minimised as much as possible throughout the study. It should also be noted that a total of 80 participants participated in the study, a relatively robust sample size for this type of research. The participants were also randomized by drawing numbers from a pot and selecting the corresponding student from the class roster. Both L1 and L2 speakers were randomly assigned to one of two groups, here labelled A and B. Those students assigned to group A undertook the cognitively more complex version of the pictorial task and the less complex version of the written task. The learners assigned to group B completed the opposite; the less complex version of the pictorial task and the more complex version of the written task. The four groups created; L1 A, and L1 B for the L1 speakers, and L2 A, and L2 B for the second language speakers completed two tasks each, differentiated by the complexity of the task as displayed in table 5-7 below.

Table 5-7. *Research study procedure*

Task	+ complexity	- complexity	+ complexity	- complexity
1	L1 A	L1 B	L2 A	L2 B
2	L1 B	L1 A	L2 B	L2 A

The data was collected from the L2 English speakers on three separate occasions as the classes were only held twice a year, due to the fact that they were post-semester summer and winter intensive English study programs. However, all of the participants were allocated to groups randomly. The L2 speakers were all attending mid-sessional intensive English language programs at a Korean university out of choice. A one-on-one meeting was arranged between each participant and myself in a laboratory setting. L2 Learners performed the research tasks during their regular English class time, but were taken out of class to an empty classroom to perform the tasks. The classroom was located on a different floor to minimize distraction for the participant, and to ensure the quality of audio recording was not adversely impacted (Cohen et al., 2013; Lambert et al., 2017). After a brief explanation to the participant of what was required the learner then performed both of the tasks, one after another. The instructions given to each participant were carefully prepared and read out from a script in English at the start of each task. The participant was not shown or given access to the instructions at any point. The instructions were kept as simple and as similar to each other as possible across the four versions, while still explaining what was required for each different task (Appendix F). Each task allowed the participant two minutes once they have been given the input to prepare their narrative, but participants were also allowed to start speaking as soon as they wished if they decided they were ready to begin. The tasks were performed in a counter balanced order to control for any sequence effects (Torgerson, 2008) with the first participant completing the pictorial task initially and following that with the written task, the second participant then completed the written task first and concluded with the pictorial task. The tasks were also counter balanced on task complexity as explained previously, with participants alternating between starting with a more complex task and a less complex task, so half of the group completed the complex task first and the other half had the easier task first as can be seen below in table 5-8.

Table 5-8. *Data collection*

	Group1 (L1 speakers)		Group 2 (L2 speakers)	
Order	First Task	Second Task	First Task	Second Task
Input Type	Picture Input - Complexity	Written Input + Complexity	Picture Input - Complexity	Written Input + Complexity
	Written Input + Complexity	Picture Input - Complexity	Written Input + Complexity	Picture Input - Complexity
	Picture Input + Complexity	Written Input - Complexity	Picture Input + Complexity	Written Input - Complexity

	Written Input	Picture Input	Written Input	Picture Input
	- Complexity	+ Complexity	- Complexity	+ Complexity

The sequence was alternated until all participants had completed both tasks. When each participant had completed both of the tasks they were returned to their English classroom. The entire process took approximately 10-15 minutes for each student, depending largely on how long the participant spoke for in the completion of each task.

5.8 Coding

The tasks recordings were made using Apple GarageBand software version 10.3.2 on my personal laptop. Each participant's speech recording was converted into a .wav audio file for automatic data analysis. Oral fluency was measured in the study by applying automated utterance fluency measures using the software program PRAAT version 6.0.28. PRAAT is a speech analysis software program designed and created by Boersma and Weenink, (2018) at the University of Amsterdam and enables the automatic evaluation of phonetic data. The PRAAT software program was used to automatically detect the length of any silences (pauses) and to count the number of syllables in each participant's utterance. Traditionally the pause threshold for a disruption in speech has been considered to be a duration of 0.4 seconds (Derwing et al., 2004), however, in light of more recent research on pausing and oral fluency by De Jong and Bosker (2013) the threshold was set at 0.25 seconds instead. PRAAT was used on the recorded sound file through the implementation of a script written and created by De Jong and Wempe (2009) (Appendix O). The script was originally written to perform analysis on speakers of Dutch by automatically interpreting changes in the frequency and intensity of waveforms to establish the amount of silence and the number of syllables produced. However, numerous previous studies (Segalowitz, 2010; De Jong et al., 2012; Préfontaine, 2013; Préfontaine et al., 2016; De Jong, 2016; Michel, 2017) have employed the script in research to measure pauses in speech and to count syllables produced in other languages, both L1 and L2, and have reported acceptable levels of accuracy and reliability.

To verify the validity and reliability of the script's automatic coding of silences and syllable count for the present study a random sample was selected of 20 (12.5%) of the 160 one-minute long audio files. Ten files were selected from each of the L1 and L2 speakers and they were then hand coded for silent pauses and for the number of syllables. A comparison between the automatically coded silences and syllables using the De Jong and Wempe script (2009) (Appendix O) and the hand coded silences and syllables revealed a consistency of greater than 0.93 on the location and length of pauses as well as the total number of syllables. This was considered to be within an acceptable range (Portney & Watkins, 2000) and the automatic coding was employed throughout the study. A second PRAAT script (Handley & Haiping, 2018) (Appendix P) was also used to calculate the length, number, and frequency of pauses occurring at the mid-clause and end-clause position which were examined in this study. The script was written by Handley and Haiping for research which explored Chinese L1 speaker pause characteristics when speaking English as an L2. Both of the PRAAT scripts were used unmodified, and were employed to provide increased reliability to the extraction and calculation of the relevant fluency measures for each audio file.

5.9 PSTM Test Procedure

After completing both of the narrative tasks the L2 participants were given a digit span test (Appendix G) to assess their levels of PSTM. The digit span test involved testing the learners on how well they could remember a series of numerical sequences. The participants were asked to repeat digit sequences of increasing length, the digits must be repeated by the participant in the same order as they were heard. The digits were read at a constant rate of 1 per second, and participants were then asked to repeat the entire sequence. The digit span test consisted of seven sections to be recalled correctly in order. Each section comprised a list of two sets of numbers, with the first section containing just two digits, the second three, and the fourth four, and so on until the last stage which comprised eight digits. Participants finished the test by correctly completing all of the eight sections, or by failing both parts of a single section. For example, if the participants answered part 1a correctly, but failed to respond appropriately for section 1b, the test would continue to be administered and move on to section 2. However, if a participant failed both 2a and 2b then the test would be concluded.

Scoring for the PSTM test involved awarding one point for each correctly repeated digit sequence. There were eight sections in the ‘number forward’ element which the participants completed, each comprising two digit sequences, therefore the maximum possible score on the test was 16.

5.10 Data Analysis

The data collected from each audio file was analysed using PRAAT with the resultant qualitative data entered into an excel spreadsheet which was then in turn converted into a .csv file and imported into RStudio for statistical analysis. RStudio was chosen as it offers the flexibility and reliability to generate accurate statistical analysis for the large number of independent and dependant variables employed in this study (Winter, 2019; Field et al., 2012).

This section will explain how the data was analysed and begin with the definition employed for the distinction between mid-clause and end-clause pauses and how they have been operationalized. What constitutes a pause in this research will also be defined and discussed, with details of the measures which have been chosen to define oral fluency described in detail.

Pauses at the end-clause level are defined as those pauses which occur at the *boundary* of a complete clause or AS-unit. Mid-clause pauses are defined as those pauses which occur *within* a clause or AS-unit. A clause or AS-unit in the present research project is defined as "… a single speaker's utterance consisting of *an independent clause*, or *sub-clausal unit*, together with any *subordinate clause(s)* associated with either" (Foster et al., 2000, p. 365). The term sentence is not used in this definition as it fails to capture the fragmentary nature of spoken language, particularly that produced by L2 speakers, and the AS-unit has been purposefully designed to deal with the features which specifically characterise spoken language data (Foster et al., 2000). An *independent clause* is defined by Foster as containing at a minimum a finite verb e.g. "That's right" or "You go to the main street of Twickenham". An *independent sub-clausal unit* is defined as consisting of: *either* one or more phrases which can be elaborated to a full clause by means of recovery or ellipted elements from the context of the discourse or situation e.g., "how long you stay here" and "three months" *or* minor utterance e.g., "oh poor women", "yes", "Thank you very much" (Foster et al., 2000, p. 366). The definition for a *subordinate clause* is that it contains at least a finite or non-finite verb element plus one additional element, either a subject, object, complement or adverbial e.g. “it is my hope :: to study crop protection” (2 clauses and 1 AS-unit) (Foster et al., 2000).

As was briefly noted above, in the current study a pause in speech was considered to be any silence lasting for 0.25 seconds or longer. All of the pauses and silences were calculated automatically in PRAAT as 0.25≥ seconds (Boersma & Weenink, 2018). The selection of 0.25 seconds as the threshold for the length of pause has been informed by several recent studies (Préfontaine & Kormos, 2015; De Jong et al., 2015; and De Jong & Bosker, 2013). The studies found that a longer pause threshold meant many L2 speakers were attributed higher levels of oral fluency in comparison to expert judgements of perceived oral fluency (De Jong et al., 2012b). In other words, pauses which were perceived by experts to impact oral fluency were ignored when a threshold of 0.4 seconds was employed. At the other end of the spectrum those pauses shorter than 0.25 seconds, which are sometimes referred to as ‘micro-pauses’ (Riggenbach, 1991), have been shown to be largely irrelevant for measures of both L1 and L2 fluency. It has been shown that these micro-pauses are often representative of individual learner differences in speech patterns, rather than as evidence of dysfluency and problems with speech production (De Jong & Bosker, 2013). A pause threshold of 0.25 seconds or longer was additionaly found to correlate well with perceptions of fluency as graded by L1 speakers, which found raters of oral fluency did not consider a pause of less than 0.25 to affect fluency (De Jong et al., 2012b). It is important to note that speech patterns clearly vary greatly across individuals and a given speaker may simply speak more slowly, and also pause for longer than is common among other speakers. This does not mean they are a less fluent speaker (De Jong et al., 2013), this is just an individual characteristic of their speech. Unless a pause exceeds 0.25 seconds it is not considered to be a reflection of difficulties with mental processing in speech production, and is not due to efforts to complete the task at hand, not is it considered to be evidence of a loss of fluency. Instead, following the implications of the research (De Jong et al., 2012b; Riggenbach, 1991) these shorter pauses are deemed to be symptomatic of individual differences in speech patterns. The present study has therefore utilised a measure of 0.25 seconds as the minimum threshold to be considered a pause which represents a loss of fluency. Shorter pauses are disregarded and are considered to provide evidence of individual speaker patterns, rather than evidencing a loss of fluency occurring as a result of the impact of the increased complexity of the task being transacted.

The oral fluency measure employed to explore oral fluency in this study is pause frequency, measured both at the mid-clause and end-clause position. Pause frequency has been shown consistently to be a reliable and valid measure for gauging oral fluency employed in many previous studies (Freed, Segalowitz, & Dewey, 2004; De Jong et al., 2012; Bosker et al., 2013, 2014; Préfontaine & Kormos, 2015). Pause frequency has also been demonstrated to correlate closely with measures of perceived fluency by trained raters of L2 oral fluency (Kormos & Denes, 2004; Iwashita et al., 2008; De Jong, et al., 2012b; Kahng, 2018).

The measure of *pause frequency* (PF) is a calculation of the total number of pauses within a given speech sample, divided by the total length of speech in seconds. Again, only those pauses deemed long enough, 0.25 seconds or greater were calculated and measured in PRAAT. This value was further divided by tracking pauses through a distinction between pauses at both the mid-clause and the end-clause, leading to the generation of two dependent variables: mid-clause pause frequency and end-clause pause frequency. The distinction in pause location seeks to provide evidence to link pauses and breakdowns in speech with the attendant mental processes and the specific stage of speech production which is involved. RStudio was then used to analyse the data collected. A detailed explanation of how pause location frequency was analysed in R, and how mixed effects linear regression models were created can be found in Appendix M (Analysis: statistical model creation and justification) and Appendix N. The appendices also contain a description of the ways in which the present study employed these models, to generate predictions for both mid and end-clause pause frequencies.

5.11 Ethics

Ethical approval was granted before data collection commenced from the University of York ethics committee and consent was given to undertake the study by the Head of Department at the Korean university. Both L1 and L2 participants were given consent forms and details of the study prior to data collection and given the opportunity to opt out of participation, which only one of the L2 students selected for the study exercised. Participants were also given the option to have their data removed from the study up to one week after data collection, after which time all of the data was anonymised. All data collected was stored on a password-protected computer only accessible to me and the audio data collected was stored anonymously so participants could not be identified.

5.12 Pilot Study

In this section I will describe the pilot study and detail the findings which had implications for the organization and formation of the final study. I will discuss the problems in obtaining sufficient quality audio recordings, as well as the often uneven response of several of the L1 speakers. The section will conclude with how these problems were addressed prior to data collection in the main study, and which strategies were implemented to combat them.

The methodological approach of this project was guided by the studies I have previously mentioned, but the methodological elements were also based upon the findings of a pilot study which was completed before the present research project was organised. The first thing the pilot study needed to accomplish was to find out how easy or difficult it would be to collect the data required. It was also important to test the automatic measures of coding and if it was possible to accurately record and analyse audio data from participants, and whether or not they could then be explored with the software programme PRAAT. The pilot also tested the implementation of the two scripts (De Jong & Wempe, 2009; Handley & Haiping, 2018) to automatically measure and analyse silence and pauses in speech. Secondly, the pilot aimed to test the logistics of having students come out of their regular language classes and perform the tasks, in the correct manner and within the allotted time, and then be returned to their original classroom to continue studying. The final aim of the pilot was to asses if the instructions given to the L1 and L2 participants were clear and resulted in replicable and consistent task performance from each participant.

The pilot was conducted in the same Korean university, and at the same university in the U.K., as the main study to make sure it provided as similar outcomes as possible. The L2 portion of the pilot was conducted over a single day and randomly selected 16

English learners from four different English language classes. The participants were interviewed separately and given the two tasks in a counter balanced order, as explained above, one student would first complete a more complex task followed by a less complex version of a task. The next learner to participate would begin with a less complex task and finish with a more complex version of the other task. The participants were all recorded using the software program GarageBand on Apple Mac, and audio files were saved onto a separate hard disk for safety and security. The participants were all taken out of their regular English classes and recordings were made in a separate unoccupied classroom, in a laboratory style environment to remove any extraneous sounds.

During the pilot several issues were encountered which were resolved prior to the collection of the data for the main study. One major problem encountered during the pilot was that for greater than 20% of the audio recordings the sound quality was too low to be automatically coded using the software programme PRAAT. This was due to several reasons, firstly the background noise from both inside and outside the classroom interfered with the accurate detection of the number of syllables. The noise inside the classroom came from participants moving their chairs, drumming on the table, clicking their pens, as well as the air conditioning unit. The external noises were from traffic on the road outside and people speaking in the adjacent classroom. Secondly, some participants spoke too softly and quietly for the software to accurately detect the differences in silence and speech, this problem also caused PRAAT to detect too few syllables and code as silence some parts of the participant's speech. I attempted to rectify these problems with the automatic coding by altering the sound detection settings within PRAAT. In De Jong and Wempe's (2009) script used to auto code the files for silence and syllables the silence threshold was altered from -20db to -30dB. It was hoped that this would enable the quieter, more softly spoken sections of the audio recordings to be detected by the auto coding. This did result in higher and more accurate syllable detection, but unfortunately resulted in even more disruption and distortion with the recordings due to increased impacts from background noise. As stated previously this resulted in approximately 20% of the L2 participant's audio files being of insufficient quality for use in the study due to the impossibility of accurate, reliable coding of silence in comparison to the other files.

It was therefore concluded that obtaining the highest quality audio recordings was necessary for the main study, with significant reductions in background noise to enable greater reliability and more accurate coding. To first reduce the background noise, it was decided to have the participants perform the tasks in a separate classroom as far away from the road and other active classrooms as possible. This was to better reduce instances of other voices and sounds being detected by the speech analysis software. It was also decided to turn off any computers, heaters, and air conditioners in the classroom, as the low-level background noise created by these devices was picked up by the microphone and interfered with the accuracy of the automatic coding. To further improve the quality of recordings it was decided that an external directional microphone would be employed to record the participants during the main data collection portion of the study. This allowed for a significant reduction in background noise and enabled a much higher percentage of the files to be accurately and effectively coded automatically in PRAAT. All participants were also encouraged to speak slightly louder than usual and to face the microphone at all times to aid the audio recordings.

Another problem identified during the pilot study was the fact that some participants were providing insufficient data to meet even the minimum of one minute required on the picture description task. The ‘accident’ task required students to create a narrative for a car accident based on eight pictures arranged in a comic strip style. However, some students would use only a simple sentence for each picture and would give an overall summative description of the pictures and the story as a whole, rather than as requested an individual detailed description of each separate picture and how it related to the story in general. This problem was found especially among the L1 speaking English participants, but also on rare occasion among the L2 speakers. This outcome may have occurred due to several factors. Firstly, some students were quite nervous and tended to speak a little more quickly than usual, motivated perhaps by a desire to finish as soon as possible, or simply due to the fact that they were uncomfortable being recorded and having their speech so closely monitored. Due to the nature of the participation the tasks were often viewed as a form of assessment or test by many of the L2 speakers and this increased their anxiety. For the L1 speakers the novelty of language class performance and the unfamiliarity of speaking for practice caused some confusion and discomfort. The problem of L1 speakers not speaking for long enough and thus not producing enough data could also be attributed to their lack of experience of classroom based language instruction and task based learning. Many of the L1 participants were not currently studying a second language and almost all of them had little or no experience of a task based or communicative approach to language learning. As a result, the L1 participants were often unsure of what was expected or required of them when completing the tasks, and unaware of how much language data was needed from them as they had no experience of language learning tasks. This may have been as a result of unclear instructions prior to the task; lacking a clearly defined expectation of just what exactly was expected of them and how much language it was necessary for them to produce. In an effort to remedy this L1 participants were given clearer more precise instructions in regards to the outcome of the task, they were also encouraged to relax and speak as normally as possible. More specifically they were instructed to give as many details as possible for each picture frame and asked to provide at least two complete sentences, describing each individual picture in the series.

5.13 Conclusion

This chapter began with an explanation of the research questions that were created from the literature review and stated the predictions that were drawn from them specifically; the focus on levels of oral fluency at the mid and end-clause level and how these levels are affected by the variation in task cognitive complexity, and task input format. The design of the study was then detailed and the independent and dependant variables explained. The next section detailed the participants; the Korean learners (L2) and the English speaking students (L1). The following section contained details of the research site followed by how and when the data was collected. The chapter moved on to discuss the operationalisation of cognitive complexity in the study through the information organisation of the two tasks, and their two variations. This was followed by a detailed description of each of the instruments and how they were created to differentiate input (written or pictorial), but otherwise maintain consistency across other aspects of task design, specifically the narrative output. The section then detailed how learners were ascribed to one of four groups and how data was collected from each of the groups. The methodology moved on to detail how the data were analysed through the use of automatic coding with PRAAT software (Boersma & Weenink, 2017) and scripts written for PRAAT by researchers who had previously utilised the automatic coding of speech measures (De Jong & Wempe, 2009; and Handley & Haiping, 2018). Details of pause frequency, which was recorded at both the mid and end-clause level, the measure used by the present study to operationalise oral fluency as well as details of the previous research which had employed the measure, were then described. The distinction for what would constitute a pause, and whether or not it would be considered a mid-clause pause i.e.; 'within a clause' or an end-clause pause i.e.; at the 'boundary' of a clause, was also defined with regards to research conducted by Foster et al., (2000) on the AS unit. The next section presented information on the digit span test employed to measure L2 speaker's levels of PSTM, as well as how it was administered. The methodology then detailed a brief description of the ethical requirements which were satisfied, and finally concluded with information on how the pilot study, which had been employed prior to main data collection, informed the make-up and design of the present study in regards to data collection and task design. The poor quality audio recordings were addressed through a reduction in the amount of background noise, and by employing a directional microphone. The problems with L1 speakers were corrected through giving each participant clearer instructions, and providing them with better explanations of what was required for the satisfactory completion of the task.

6 Data Coding and Analysis

6.1 Introduction

In this chapter I will explain how the oral data collected during task performance were coded and analysed. I will first detail how the data were recorded, measured, and then automatically coded using the software program PRAAT (Boersma & Weenink, 2018). I will provide examples from the data to illustrate the different procedures and techniques used to perform the coding. This chapter will then move on to look at the justification for employing automatic coding and the limitations inherent with its application. The chapter concludes with a brief summary of the inter-rater reliability measures obtained.

6.2 PRAAT

Firstly, every audio recording file was converted into a .wav audio file to facilitate the automatic data analysis in PRAAT. Oral fluency was operationalised here by applying automated utterance fluency measures through application of the functions in the software program PRAAT version 6.0.28. PRAAT is a speech analysis software program which was designed and created by Boersma and Weenink, (2018) at the University of Amsterdam to allow for the automatic evaluation of phonetic data. PRAAT employs *scripts*, which are user generated instructions, to allow for the tailoring of the measures recorded and calculated to meet the specific needs of the data analysis required in each study. The PRAAT software program was used in the present study to automatically detect the length of any pauses in speech, through the detection of silence, and to calculate the number of syllables spoken by each participant. As was described in the methodological section (5.8), the threshold for detecting a pause was calculated at 0.25 seconds or longer in PRAAT. This threshold is in line with the findings of several recent research studies (De Jong & Bosker, 2013; De Jong et al., 2015; Préfontaine et al., 2016) which reported that 0.25 seconds was a more appropriate measure for pauses to accurately reflect a break down in speech production, rather than simply reflecting an individual difference in a given speaker’s speech patterns.

6.3 Recording

All of the data collected during task performance were audio recorded on my personal laptop, an Apple Macintosh computer, using the media recording software GarageBand version 10.3.2 and an external directional USB microphone. After being informed by the pilot study phase where data collection experienced some considerable problems, directional microphones were employed and background noise was reduced as much as possible. The microphone was also positioned in front of participants and they were asked to speak in the direction of the microphone to facilitate the more accurate audio recording of data, and ultimately enable the software to more accurately analyse the data automatically. During the recording and data retrieval process it was also important to keep track of which version of the task participants were performing; whether it was the written or pictorial input format, or the cognitively more or less complex version.

6.4 Automatic coding

As I explained earlier in Chapter 5 the number of studies using computer software to provide greater insight into the analysis of L2 speech has been steadily increasing in recent years (Segalowitz, 2010; De Jong et al., 2012; Préfontaine, 2013; Préfontaine et al., 2016; De Jong, 2016; Michel, 2017). There are a number of benefits which the software provides to data analysis and as a result, all of the audio data collected for this study was analysed using the software program PRAAT. The primary benefit of the software's automatic coding is the increased reliability it provides in accurately and consistently recording the number and location of syllables, as well as the automatic coding of silences and pauses in speech. The implementation of automatic measures means that all silences are reported objectively and are not open to interpretation through human error. The software provides increased consistency in calculating measures when dealing with large amounts of audio data, such as were utilised in the present study, and removes the impact of any human errors which will almost certainly occur when dealing with this much data, and these many calculations.

However, there are limitations to the software which mean completely automatic analysis of fluency was not possible with the current data set. Fully automatic fluency calculations are only capable of calculating particular and specific measures. It is possible to calculate the value of such measures as *speech rate* for example, which is the amount of time a participant spends speaking calculated as a ratio of the total time of the task, using De Dong and Wempe's (2018) script (Appendix O). However, it is not possible to calculate the *pruned speech rate* for example, a measure which takes into account the amount of speech which is free from errors, repetitions, and restatements. The identification and calculation of 'filled pauses', which Clark and Tree (2002) have shown are also evidence of problems with mental processing in speech production, as pauses is also problematic as the script only codes silences as pauses, and is as yet unable to identify 'ums' and 'ers' as breakdowns in communication. There is currently no script available for PRAAT that can automatically identify and measure these filled pauses accurately (Hilton, 2014) resulting in the necessity of calculating fluency measures manually in conjunction with the measures automatically generated in PRAAT.

PRAAT is a versatile piece of software and it allows the user to investigate fluency analysis from a variety of different perspectives, as well as with varying degrees of automaticity. This versatility and flexibility allow the user to undertake as much, or as little, detailed manual investigation as their study requires. It is possible to combine the automatic measurement PRAAT provides of the duration of chosen elements with manual annotation of the speech samples. The manual annotation of speech files is a laborious task and consumes a substantial amount of time, it also requires the user to become proficient in using and understanding the scripts which PRAAT employs. However, the combination of automatic and manual fluency analysis provides the present study with the highest quality data for quantitative analysis.

To collect the data required to answer the research questions all of the tasks undertaken by participants were analysed first automatically and then manually in PRAAT. There were 40 L1 speakers and 40 L2 speakers, and each speaker performed two different tasks, meaning there were a total of 160 tasks, the recordings were of various lengths, but were only annotated and analysed for the initial one minute. This was done to ensure an equal amount of time to provide for a balanced and easily comparable reflection of the fluency measures contained in each. There was little variability in the way each participant started off their narrative, with the first minute providing a reasonably accurate representation of their overall performance on the task. The audio recordings were then investigated one by one, through looking closely at the spectrogram

(Appendix Q) generated by PRAAT, whilst also listening to the audio file. Using a script written by De Jong and Wempe (2009) for PRAAT, all of the files were first automatically coded to identify silences. The script detects when speech occurs, and conversely when there is silence, and records this data to a file called a ‘textgrid’. An example textgrid (Appendix R) can be seen with annotation at the bottom to denote the different sections of speech and silence. The next stage of analysis was to check for voiced pauses, these are instances where the speaker interrupts the flow of speech by saying ‘er’ and ‘um’ due to difficulties in producing speech. This was manually achieved by listening to the audio file whilst looking at the spectrogram and matching the audio with the representations of intensity and pitch shown in the spectrogram in PRAAT. It is possible in this way to get a very precise and accurate account of when each instance of speech ends and the silence begins, and to record the boundaries. The analysis of each speech sample begins at the first syllable spoken by the participant regardless of if it is a voiced pause, or if it is a word or part of a word. Filled pauses as mentioned above are those pauses which are voiced, whereas complex pauses are those pauses which contain both voiced elements, as well as some silence. The distinctions between these types of pauses were carefully identified through repeated listening to small sections of the audio file, whilst at the same time observing the spectrogram. Only those voiced and complex pauses which were longer than 0.25 seconds were recorded and marked as pauses, and they were manually recorded on the textgrid for each audio file as either silent, filled, or complex. This creates intervals on the textgrid, with the start and end of each pause clearly defined for the entire speech sample, and each segment labelled as a either speech, or as a silent, filled, or complex pause. The next step of analysis was to determine whether pauses should be classified as a mid-clause pause or an end-clause pause. This was done again by carefully listening to the audio file with a specific focus on the language employed, to determine if the pause occurred within a clause or an AS-unit, or at the boundaries of a clause/AS-Unit. This was done in line with the distinction described by Foster et al., (2000), detailed in section (5.10) of the methodological chapter, and was chosen to better represent the characteristics of spoken, rather than written language. Each of the pauses in speech when the categorised as occurring at either the end or mid-clause position, and this distinction was also added to each of the textgrids.

After running the first PRAAT script (De Jong and Wempe, 2009) (Appendix O) to determine where the silences and pauses should be measured, and manually adding the tiers to the 'textgrid', the final stage of data analysis was to apply another script (Appendix P) written by Handley and Haiping (2018) to calculate the total number and frequency of pauses occurring at both the mid and end clause for each audio file. The output of the second script was exported into a Microsoft Excel file to better compare the data on the location and frequency of pauses, as well as to view the other speech phenomena. These data were then employed in the statistical analysis carried out to calculate the findings of this study.

As mentioned earlier the manual coding of measures in PRAAT is a time consuming undertaking, and as the primary focus of this research is on the location, and classification of pauses it was decided that this would be the focus of second-rater re-examination. A 20% sample of the data was randomly chosen and recoded for the location of the pause, here being either a mid-clause pause or an end-clause pause. This resulted in a sample of thirty-two 'textgrids' and their accompanying audio files being examined by the second rater to establish inter-rater reliability. The second rater had previous experience of hand coding using PRAAT and was familiar with the classification for the location of pauses employed in this study. Each 'textgrid' contained between eleven and twenty-three pauses each of which required classification, resulting in a total of 293 pauses to be classified. The second rater agreed on the classification of all but 21 of the pause locations categorisations across the 293 pause locations, resulting in a high rate of inter-rater reliability of .928.

6.5 Conclusion

In this chapter I explained how the data for my study has been coded and analysed. I first detailed how the data was recorded, and then described the procedures used to calculate the required measures using the software programme PRAAT. I next looked at the way in which the textgrids were used to detail the differences between speech and silence, and how the scripts were used to differentiate the types and locations of pauses. I then proceeded to describe the benefits of PRAAT's automatic coding as well as the benefits of manual coding, and the decision to utilise both for the analysis of the current data set, explaining how manual and automatic analysis was combined to calculate the fluency measures required for my study. The next section described some of the difficulties and issues faced when using PRAAT. Finally, details of inter-rater reliability were presented and the high level of inter-rater reliability provided.

7 Results

7.1 Introduction

In this chapter I will describe the findings obtained generated through statistical analysis to answer each of the research questions. The chapter begins with a brief reminder of how task complexity was operationalised for the study and how this operationalisation was verified with empirical data. The next section provides a brief description of the layout of the chapter, describing how each research question is explored separately and in the same order as laid out in the methodological chapter. The next section details a comparison of the results obtained for each research question, and continues by drawing together the results. A brief description of the limitations of the models employed is then presented. The final section of the results chapter examines details of the correlations displayed between the L2 speaker's levels of PSTM, as measured by a post task test, and the location and frequency of pauses. The chapter concludes with a brief summary of the results obtained.

7.2 Establishment of Task Complexity

The first analysis conducted was to establish the difference between the levels of task complexity, providing empirical evidence that the tasks operationalised were distinctly more or less complex to complete. This analysis was detailed in section 5.7 and 5.8. As described in the methodological chapter the level of task complexity had been operationalised in line with Robinson's taxonomy (2007) of task complexity and involved the manipulation of information provided in the input with +/- chronological order. The data provided by the qualitative and quantitative research provides empirical evidence that the operationalisation of '+/- order of information' (Robinson, 2007) employed does impact task complexity significantly. It also provides robust evidence to define the tasks as clearly being distinguishable by the effort required to transact them. This firmly establishes the tasks as being either 'more' or 'less' complex.

7.3 Chapter Layout

The following sections will each examine one of the research questions detailed in the methodological chapter separately and in order. Each segment will begin by first providing a brief reminder of the appropriate research question and hypothesis, and continue with an initial view of the descriptive statistics generated, followed by details of the distribution of the data, and finally concluding by answering the research question using the results obtained through analysis of the data and use of the inferential statistics generated.

7.4 RQ 1: The Impact of Task Complexity on L1 and L2 Oral Fluency at the Mid-Clause.

The first research question explores the relationship between mid-clause pausing frequency and the differing levels of task complexity. It was hypothesised that the increased task complexity would result in a higher frequency of mid-clause pauses for the L2 speakers, but that the L1 speaker's mid-clause pausing frequency would be unaffected during completion of the more complex tasks.

7.4.1 Descriptive Statistics

To enable a clearer outline of the data and to help better identify any patterns, descriptive statistics were generated for the frequency of mid-clause pauses (number of pauses divided by seconds speaking) for levels of task complexity (+/- order of information), input format types (written/pictorial), and divided by speaker type (L1/L2). Table 7-1 below shows the mean (M) mid-clause pause frequency and the standard deviation (SD) for each category within the independent variables. The table is organised to display the results separately for speaker type, and then further divided by the type of task; more or less complex (C+/C-), as well as whether or not the input format was written or pictorial (Wri/Pic).

Table 7-1. *Descriptive statistics for mid-clause pause frequency*

	L1 speakers				L2 speakers			
	C+	C-	Wri	Pic	C+	C-	Wri	Pic
M	0.09	0.08	0.08	0.09	0.22	0.18	0.20	0.20
(SD)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)

Table 7-1 demonstrates that the highest frequency of mid-clause pausing was recorded by the L2 speakers on the more complex tasks, with a mean frequency more than double that of the L1 speakers on the more complex versions of the tasks. For the L2 speakers there is also a higher frequency of mid-clause pauses for the more complex tasks when compared with the less complex version. Table 7-1 shows that the difference in task complexity has a smaller impact on the frequency of mid-clause pauses than the impact of speaker type. However, this initial view of the data supports the hypothesis that both the speaker type and the level of task complexity have an impact on the frequency of mid-clause pauses, and the L2 speakers are indeed pausing more at the mid-clause on the more complex tasks.

7.4.2 Distribution of Data

The mid-clause pause data was then analysed for normality of distribution and to identify any outliers. To help provide a clearer view of the data, and to more easily identify outliers, box plots were also generated showing data for all three independent variables. Below in Figure 7-1 is the box plot generated in RStudio, showing mid-clause pause frequency across speaker type, task complexity, as well as between the different task formats.

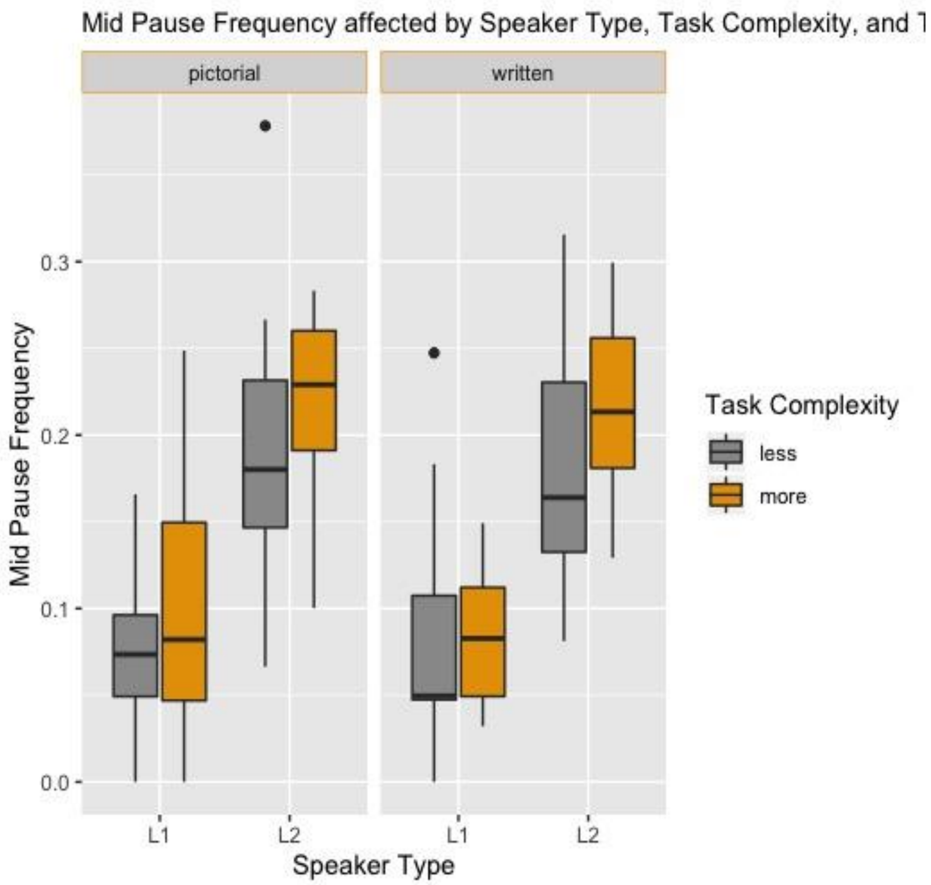


Figure 7-1. *Mid-clause pause frequency by speaker type, task format and task complexity*

Figure 7-1 clearly highlights the large difference between the frequency of pauses occurring within clause boundaries (at the mid-clause position) in a comparison of L1 and L2 speakers. The box plot figure 1 shows two extreme outliers, one for the L2 speakers on the less complex pictorial task, and one for the L1 speakers on the less complex written task. Each individual extreme outlying data point was identified and checked for precision, but the data was accurate, with no errors found in recording or calculation. Due to the accuracy of the data, it was decided that they should still be included in the final analysis. This decision is supported by both Field et al., (2012) and Winter, (2019) who point out that removing data solely based on it being an outlier is not always appropriate, and can lead to a reduction of the accuracy and validity of the results obtained. Keeping all of the data collected also conferred the benefit of maintaining the balance of the study with, equal numbers of responses across each speaker type, task input format, and level of complexity.

Q-Q plots were generated in RStudio to more exactly visually asses the normality of distribution for each of the four groups to be compared: L1 less complex (L1C-), L1 more complex (L1C+), L2 less complex (L2C-), and L2 more complex (L2C+). A Shapiro-Wilk test was also performed for each of the four groups to statistically check the normality of distribution. Table 7-2 below summarises the findings.

Table 7-2. *Shapiro-Wilk test for normality of distribution for mid-clause pause frequency*

	L1 speakers		L2 speakers	
	C+	C-	C+	C-
W	0.94	0.89	0.97	0.95
p	0.04	<0.001	0.38	0.10

Table 7-2 shows that both of the L2 speaker groups contained data which was normally distributed, however both of the L1 speaker groups had data which was found to differ significantly from normal distribution. The two groups were more closely explored with density plots generated to show how exactly they differed from normal distribution. The density plot for both the L1C-, and the L1+C, showed that the data was slightly positively skewed. As each of the groups shared a similar pattern of non-normal distribution, and contained a substantial number of responses (n=40), with larger groups being more robust when dealing with variations in normality of distribution (Field et al., 2012), it was decided to proceed with the analysis and not to transform the L1 mid-clause pause data.

7.4.3 Inferential Statistics

As detailed in Appendix M, MidFqMod 7 returned the highest Akaike Information Criterion (AIC) value of -489.38, so this model was chosen to assess the differences between groups on the frequency of mid pauses. Statistical significance of the simple main effect analyses as well as those conducted in the pairwise comparisons were accepted at a Bonferroni-adjusted alpha level of 0.025. This Bonferroni correction for the significance value was implemented to control for the use of multiple tests (Field et al., 2012; Winter, 2019). This corresponds to the current level of statistical significance at $p < 0.05$ divided by 2 for the number of simple main effects computed. The analysis shows there was a statistically significant main effect of speaker type, t (1, 78) = -0.56, $p < 0.0001$, but not of task complexity, t (1, 78) = -0.38, $p < 0.70$ on the frequency of mid-clause pauses. The interaction of speaker type x task complexity was also shown to not be statistically significant, t (1, 78) = 1.25, $p = 0.09$.

Pairwise comparisons of the different levels of task complexity and the different speaker types were performed next. There was a significant difference in mid-clause pause frequency for the L2 speakers between less complex and more complex versions of the task, F (1, 156) = -2.62, $p < 0.01$, but no significant difference was shown for the L1 speakers, F (1, 156) = -1.03 $p < 0.30$. There was also a highly significant difference evidenced for mid pause frequency between the L1 and L2 speakers on the less complex task version, F (1, 156) = -8.25, $p < 0.0001$, and on the more complex version of the task, F (1, 156) = -9.83, $p < 0.0001$. These results provide evidence that task complexity did have a statistically significant impact on the mid-clause pause frequency of the L2 speakers, an impact which was not found to be true for the L1 speakers.

7.4.4 Generalised Linear Model

A generalised linear model (GLM) was also constructed using negative binomial regression to create predictions for the frequency of mid-clause pauses by each independent variable (IV). The initial model examined the impact of task complexity, speaker type, and task format on the frequency of mid-clause pauses, with each separate model controlled for the amount of time spoken, through the fitting of the exposure variable *duration* (Winter, 2019). The results of which are shown below in table 7-3.

Table 7-3. *General linear model predictions by speaker type and task complexity for mid-clause pause frequency*

	L1 speakers		L2 speakers	
	C+	C-	C+	C-
M	0.09	0.08	0.22	0.18

Table 7-3 confirms the results noted in table 7-1, and the predictions generated from the general linear model provide evidence to confirm the hypothesis that L2 speakers do pause more frequently than the L1 speakers at the mid clause level, due to the effects of the increased task complexity.

7.5 RQ 2: The Impact of Task Complexity on L1 and L2 Oral Fluency at the End-Clause.

The second research question addressed the relationship between increased task complexity and frequency of pauses occurring at the end of clauses. The hypothesis stated that increased task complexity would result in an increase in the frequency of end-clause pauses for both L1 and L2 speakers, however the impact for L2 speakers would not be as great as that evidenced at the mid-clause.

7.5.1 Descriptive Statistics

Again, descriptive statistics were generated for both speaker types and divided by task complexity and task input format, with the results displayed below in table 7-4.

Table 7-4. *Descriptive statistics for end-clause pause frequency*

	L1 speakers				L2 speakers			
	C+	C-	Wri	Pic	C+	C-	Wri	Pic
M	0.23	0.26	0.22	0.28	0.17	0.19	0.17	0.19
(SD)	(0.07)	(0.07)	(0.05)	(0.08)	(0.06)	(0.05)	(0.05)	(0.06)

Table 7-4 displays the highest frequency of pausing at the end-clause was recorded by the L1 speakers, with a mean frequency higher than that of the L2 speakers across both levels of task complexity, and both levels of task input format. A finding which shows the L1 speakers pause at the end of clauses more frequently than L2 speakers. A higher frequency of end-clause pauses was also recorded for the less complex task version for the L2 speakers, though the difference noted between the levels of complexity is not as large as that displayed for the type of speaker. We can conclude that the greater impact on pause frequency is as a result of the difference in task complexity, and that the difference in speaker type has a smaller impact on the frequency of pauses occurring at the end of clauses. The descriptive statistics confirm the hypothesis; with the more fluent L1 speakers and the easier, less complex tasks generating more end-clause pauses, but with less frequent end-clause pauses recorded for the L2 speakers and for the more complex tasks.

7.5.2 Distribution of Data

Figure 7-2 generated below also shows the difference between speaker types in the frequency of end-clause pauses. The box plot was generated to show a comparison of end-clause pause frequency, again across task type, speaker type, and task complexity.

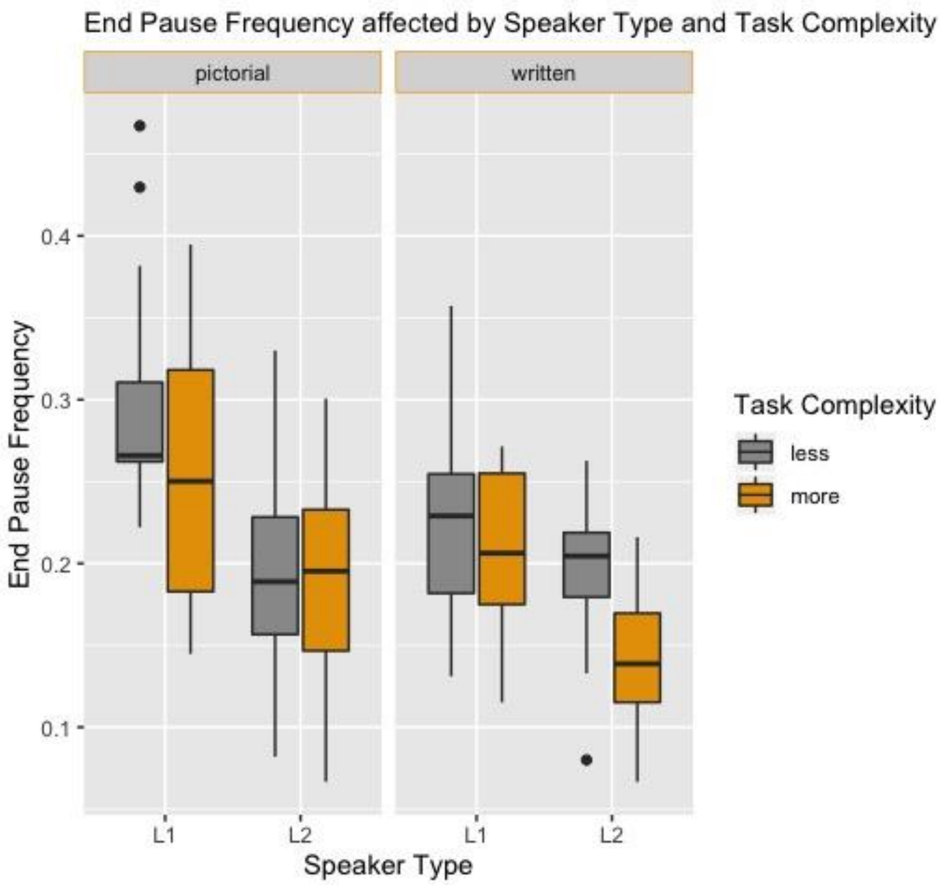


Figure 7-2. *End-clause pause frequency by speaker type, task format and task complexity*

Figure 7-2 presents the differences in patterns of end-clause pause frequencies for the different levels of task complexity, task input format, and speaker type. To check for outliers Q-Q plots were again created to assess the normality of distribution for each of the groups. A Shapiro-Wilk test was again conducted for each group to check the end-clause pause data for normality of distribution. Table 7-5 below presents a summary of these findings.

Table 7-5. *Shapiro-Wilk test for normality of distribution for end-clause pause frequency*

	L1 speakers		L2 speakers	
	C+	C-	C+	C-
W	0.95	0.91	0.97	0.97
p	0.08	<0.006	0.29	0.129

In table 7-5 we can see both of the L2 speaker groups contain data which is not significantly different from normal distribution, however the L1C- group again contains data which differs significantly from normal distribution. A density plot was generated to more clearly show the distribution. The density plot for the L1C- and L1+C groups displayed data which was only very slightly positively skewed. Again, as the groups of data comprised a similar pattern of distribution, and included a large number of responses (n=40), it was concluded that the data did not need to be transformed and would be used as it was (Winter, 2019; Field et al., 2012).

7.5.3 Inferential Statistics

EndFqMod8 (Appendix N) was run as the model of best fit, and reported the effect that speaker type, t (1, 78) = -3.03, $p = 0.0032$, was significant at the Bonferroni adjusted alpha level of $p < 0.025$, for the frequency of end-clause pauses. Task complexity however was shown not to be significant, t (1, 74) = -1.81, $p < 0.08$ for end-clause pause frequency.

Pairwise comparisons on speaker type highlighted the statistically significant difference between the frequency of end-clause pauses between speaker types on both the less complex, F (1, 156) = 4.98, $p < 0.0001$ and the more complex tasks, F (1, 156) = 4.57, $p < 0.0001$.

Pairwise comparisons for task complexity revealed that there was no statistically significant difference in end-clause pausing between the less and more complex tasks for L2 speakers, $F(1, 156) = 1.73$, $p < 0.08$, or for L1 speakers, $F(1, 156) = 2.14$, $p < 0.03$. Results were again calculated at the Bonferroni adjusted alpha level of $p < 0.025$.

7.5.4 Generalised Linear Model

A GLM explored the effects of task complexity and speaker type on the frequency of pauses located at the end of clauses. The results are shown in table 7-6.

Table 7-6. *General linear model predictions by speaker type and task complexity for end-clause pause frequency*

	L1C-	L1C+	L2C-	L2C+
End pause frequency	0.26	0.23	0.19	0.17

The results in table 7-6 display predictions that confirm the previous results seen in Table 7-4, and from the statistical analysis. If we compare data across pause locations, we can see the L2 speakers are impacted by increased task complexity to a larger extent at the mid-clause than the end-clause. The predictions generated also confirm the hypothesis from research question 2 that L2 oral fluency is affected more at the mid-clause than the end clause by task complexity.

7.6 RQ 3: The Impact of Task Input Format on L1 and L2 Oral Fluency at the Mid-Clause.

Research question 3 sought to address the relationship between a variation in task input format and the frequency of mid-clause pauses. The hypothesis was for less frequent pauses at the mid-clause for L2 speakers on the written input task, and for more frequent pauses at the mid-clause on the task with a pictorial input format. So, the written input task would result in higher levels of oral fluency. L1 speaker pause frequency was predicted to be similar across the two task input formats.

7.6.1 Descriptive Statistics

Table 7-7 below shows the difference between tasks input types, combining levels of task complexity, with little difference evident in pausing frequencies within speaker groups.

Table 7-7. *Mid-clause pause frequency separated by task input format*

	L1 speakers		L2 speakers	
	Pic	Wri	Pic	Wri
M	0.08	0.07	0.20	0.20
(SD)	(0.06)	(0.05)	(0.06)	(0.06)

Table 7-7 does display a marked difference in the frequency of mid-clause pauses for task input format between speaker types, showing that the L2 speakers pause more than twice as frequently as their L1 speaking counterparts on both tasks input formats. The findings display that the variable of speaker type has a greater impact on the frequency of mid clause pauses, than the task input format variable. However, it should be noted that the L1 speakers did evidence slightly more frequent mid-clause pauses on the pictorial format tasks, than on the written input tasks.

7.6.2 Distribution of Data

The four groups were checked for normality of distribution and outliers. A Shapiro-Wilk test was also performed for each group to ascertain if the mid-clause pause frequencies were in line with normal distribution. Table 7-8 below shows the results of the Shapiro-Wilk test.

Table 7-8. *Shapiro-Wilk test for normality of distribution for speaker type and input format groups*

	L1 speakers		L2 speakers	
	Wri	Pic	Wri	Pic
W	0.91	0.91	0.97	0.97
p	<0.002	0.01	0.49	0.34

In table 7-8 the L2 speaker groups both contain data which is not significantly different from normal distribution, however the L1 speaker groups do differ and both contain non-normally distributed data. Density plots were created, and the groups again showed very similar patterns of slightly positively skewed data distribution.

7.6.3 Inferential Statistics

A model (MidFqMod 6 – Appendix M) was created, and shows that including task format in addition to task complexity and speaker type did not improve the fit of the model. The AIC value increases with the addition of task format as a fixed effect, indicating that task format did not have a significant overall effect on the frequency of mid-clause pauses $\chi2(1) = 1.752$, $p = .185$. MidFqMod 8 (Appendix M) incorporated the speaker type x task format interaction into the model, but again the model was not significantly improved, $\chi2(1) = 0.001$, $p = .97$. Further analysis shows that there was no statistically significant main effect of task format, t (1, 76) = -0.38, p = 0.70 on the frequency of mid-clause pauses, and the interaction of speaker type x task format was also shown to not be significant, t (1, 76) = -0.04, p = 0.97.

Pairwise comparisons for speaker type show a statistically significant difference on the frequency of mid-clause pauses between speaker types on both the written, F (1, 156) = -8.82, $p < 0.0001$ and the pictorial tasks, F (1, 156) = -8.85, $p < 0.0001$. This result displays that the L2 speakers produced significantly more mid-clause pauses than the L1 speakers on both the written and pictorial versions of the tasks. However, pairwise comparisons for task format revealed that there was no difference in the frequency of mid-clause pauses between the written and pictorial tasks for L2 speakers, F (1, 156) = 0.61, p = 0.55, or for L1 speakers, F (1, 156) = 0.57, p = 0.57. This means that for RQ 3 the hypothesis in not proven, and the difference in task input format did not have an impact on the number of mid-clause pauses made by either the L2 or the L1 speakers.

7.6.4 Generalised Linear Model

The generalised linear model results are shown below in table 7-9.

Table 7-9. *Predictions from general linear model of mid-clause pause frequency by format and speaker*

	L1 speakers		L2 speakers	
	Wri	Pic	Wri	Pic
M	0.08	0.09	0.20	0.20

Table 7-9 displays the difference between speaker types is also notably consistent, with the L2 speakers pausing twice as frequently as the L1 speakers on both the pictorial and written input tasks. In contradiction of the prediction made in RQ 3, the pictorial tasks do not elicit a higher frequency of mid-clause pauses than those recorded for the written tasks.

7.6.5 Model fitting and Limitations

The step-by-step forward-selection procedure (Appendix N) employed by this model for this analysis can possibly lead to underspecified models. For a more detailed description of the model fitting process and the limitations inherent in this process see Appendix S.

7.7 RQ 4: The Relationship Between L2 Speaker's PSTM and Oral Fluency at the Mid and End-clause

Research question 4 seeks to investigate the relationship between the L2 speaker's levels of phonological short term memory (PSTM), and the measures of oral fluency recorded at the mid and end-clause positions. The hypothesis stated that those speakers with lower levels of PSTM would evidence decreased oral fluency at the mid-clause, and as a result would pause more frequently than those speakers with higher PSTM levels. This section will examine the correlation between the PSTM test scores collected from the post task test undertaken by all of the L2 speakers, and the frequency of pauses observed at the mid and end-clause.

7.7.1 Descriptive Statistics

The PSTM scores were first examined and explored through the generation of descriptive statistics. The scores recorded on the PSTM test ranged from 14 at the maximum to 5 at the minimum, the mean score was 10.1, with a standard deviation of 2.8.

7.7.2 Distribution of Data

The PSTM scores were next checked for normality of distribution. The results of a Shapiro-Wilk test reported that the data for the L2 speaker PSTM scores was non normally distributed, with a statistically significant deviation, $W = 0.94$, $p < 0.04$. As a result of the data set containing non normally distributed data a Spearman rank correlation coefficient was employed in the analysis (Field et al., 2012).

7.7.3 Inferential Statistics

A spearman rank correlation reported that the correlation coefficient between PSTM and mid pause frequency was, $R = -0.99$, $p < 0.0001$. A highly statistically significant result, as a score very close to -1 indicates there is a strong negative correlation, in this case between the PSTM score and the frequency of mid-clause pauses. As levels of PSTM increased the frequency of mid-clause pauses decreased. These findings are represented below in figure 7-3.

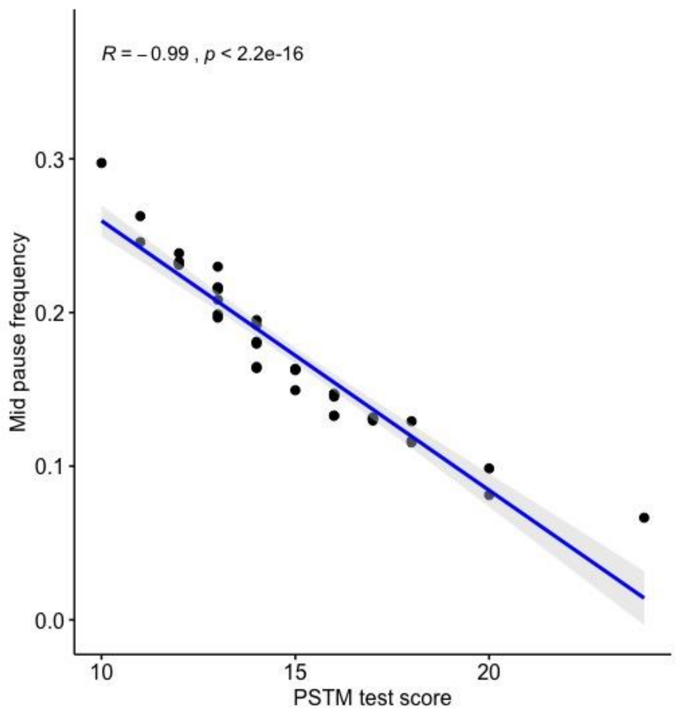


Figure 7-3. *Spearman's Rank Correlation between PSTM and mid-clause pause frequency*

Figure 7-3 above shows a Spearman's Rank Correlation of the L2 speaker's PSTM scores and the frequency of mid-clause pauses. A pattern is clearly evident from the data; those L2 speakers with lower PSTM scores have a corresponding increase in the frequency of mid-clause pauses. This supports the hypothesis that lower levels of PSTM result in lower levels of oral fluency, at the mid-clause position.

A similar analysis was also run for the correlation coefficient between the frequency of end-clause pauses and PSTM. Again, as the PSTM data was non normally distributed Spearman's rank correlation results are reported. Results are shown below in figure 7-4.

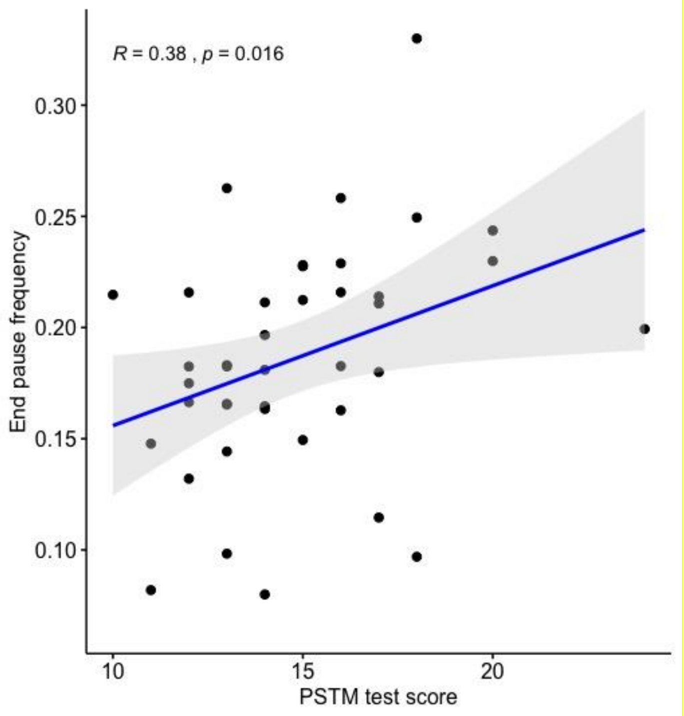


Figure 7-4. *Spearman's Rank Correlation between PSTM and end-clause pause frequency*

Immediately evident from the distribution in figure 7-4 is that the data is far more widely dispersed than the data seen for the previous analysis of mid pause frequencies. However, the correlation coefficient between PSTM and end pause frequency was still significant and is reported as: R = 0.38, p = 0.016. Again, this is a statistically significant result, with a positive score indicating that there is a correlation between the PSTM score and the L2 speaker's frequency of end-clause pauses. This result indicates that as PSTM increased the frequency of end-clause pauses also increased.

The two Spearman's rank correlations performed show that there is relationship between the PSTM of L2 speakers and their oral fluency, as evidenced through the frequency of pauses, both in the middle and at the end of clauses. This answers research question 4, and provides evidence to support the hypothesis that higher levels of PSTM among L2 speakers does provide benefits for more fluent speech production.

7.8 Conclusions

This final section draws together information from the results chapter and provides a brief summary based on the findings. The chapter first reported briefly on how task complexity was established and then moved on to detail the layout of the chapter and how it would be presented. The next sections of the chapter each focused on addressing one research question, and provided a reminder of the question as well as the hypothesis. Each section detailed the descriptive statistics, then the distribution of data, followed by the inferential statistics, and where appropriate concluded with reference to the general linear model utilised to answer the research questions.

The first section on research question 1 explored the impacts of task complexity and speaker type on the frequency of mid-clause pauses. Results indicated that mid-clause pausing frequency was significantly increased by the greater level of task complexity for the L2 speakers, but not for L1 speakers. These results were supported by the predictions generated from the general linear model which predicted a significant increase in mid-clause pausing frequency for L2 speakers between levels of task complexity.

The second research question examined the frequency of end-clause pauses and how they were affected by increases in task complexity for both L1 and L2 speakers. The L1 and the L2 speaker pause frequency at the end-clause was not impacted by task complexity, with no statistically significant differences recorded. Again, these results were supported by the predictions generated in a general linear model.

Research question 3 examined the impacts of task input format on the frequency of pauses in the middle of clauses. Analysis of the data and predictions from a general linear model both reported no statistically significant impact of task input format on the frequency of mid-clause pauses, for either the L1 or the L2 speakers.

The final section of results looked at research question 4 and the relationship between the L2 speakers' scores on the PSTM test and the rate of mid and end-clause pauses. The results of Spearman's rank correlations indicated that PSTM had a strong negative correlation with the frequency of clauses occurring in the middle of clauses, whereas PSTM correlated positively with those pauses occurring at the end of clauses.

The next chapter will move on to explore the results presented here in greater detail and offer interpretations and suggestions for the implications of the results obtained in relation to the literature.

8 Discussion

8.1 Introduction

In this chapter I will discuss the main findings presented in Chapter 7. The results will be explored in relation to each of the research questions and also to the studies presented in the literature review. I will first provide a brief overview of the key findings of the study and how they relate to the research questions. In the following section I will discuss the findings related to task complexity and how it impacts oral fluency at the mid and end-clause. The second section will examine the effects of task input type on oral fluency measures at the middle and end of clauses. In the third segment the correlations between PSTM and L2 learner oral fluency will be discussed. The chapter will then move on to a discussion of the theoretical, methodological, and pedagogical implications of the study, followed by a section on the limitations, and a description of suggestions for areas of possible future research. The final section will include a brief chapter summary and conclusion.

8.2 Key Findings and Significance of the Study

The aim of this thesis was to explore the impacts of a variation in task design on the production of fluent speech by L2 speakers. The thesis sought to address the gap in knowledge through an examination of variations in the impacts of task design, namely: task complexity, and task input format, L2 speech production and fluency. The research also sought to address the role of PSTM in L2 oral fluency. This study is one of the first to explore how the differing cognitive demands associated with task input type affect second language learner oral fluency, specified by measuring breakdowns in speech as either clause internal or clause external. It aims to provide a nuanced perspective on how spoken fluency can be understood and measured. This research contributes to the debate on task design, and has implications for language teaching pedagogy, syllabus design, and assessment. As well as offering some much needed insight into the impacts of task input type on language learner oral performance, it also provides empirical evidence on the individual learner difference of PSTM and how this benefits the spontaneous production of L2 speech.

The first research question addressed the impact of increased cognitive complexity in tasks on L1 and L2 cognitive fluency. The study demonstrated that the frequency of L2 speaker's pauses did increase at the mid-clause position when completing the more complex version of the task. The study also found that increased task complexity did not significantly increase the frequency of mid-clause pauses for the L1 speakers. At the end-clause position the frequency of pauses were largely unaffected by the increase in task complexity and saw no significant differences for either the L1 or L2 speakers. The second research question explored the impact of task input format on measures of oral fluency, with the findings indicating that they had little impact on the oral fluency of L1 or L2 speakers when recorded at either the middle or end of clauses. The third and final research question examined how PSTM levels impacted L2 oral fluency and demonstrated a strong negative correlation between participants PSTM scores and the frequency of mid-clause pauses. The results also highlight a positive correlation between PSTM and the frequency of end-clause pauses.

8.3 The Impacts of Increased Cognitive Task Complexity (RQ 1 and 2)

This section will discuss the findings of the thesis in relation to research question 1 and 2, and conclude with a summary of the findings recorded, and explanations with references to the relevant literature.

8.4 Mid-Clause Pause Frequency - L2 Speakers (RQ 1)

For the present study an increase in mid-clause pause frequency was evidenced for the L2 speakers on the more cognitively complex version of the task. This means L2 speaker oral fluency decreased on the more complex task. This is evidenced by a significant increase in the frequency of pauses within clauses on the more complex version of the task, when compared with the L2 speaker's performance on the less complex tasks. The finding is supported by Skehan et al, (2016), and by several other studies which distinguished between the locations of pauses when reporting the effects of increased cognitive load on speech production (Kahng, 2014; De Jong 2016; Tavakoli, 2011).

Several studies have reported that task complexity impacts the oral fluency of L2 speakers (Robinson, 2003; Tavakoli & Foster, 2008; Skehan, 2009; Awaad and Tavakoli, 2019), but relatively few studies to date have focused on the impacts of task cognitive complexity on oral fluency measures at the mid-clause location. Skehan et al., (2016) demonstrated that there is a significant difference between the pausing rates of L2 speakers at the mid and end-clause location, more specifically those pauses occurring in the middle of clauses are related to problems with microplanning and the formulation of speech, with the pause at the end of clauses reflecting difficulties in macroplanning and problems with conceptualisation.

More recently Awwad and Tavakoli (2019) also reported increased oral fluency performance from L2 speakers when performing a more complex task. However, their findings reported increases in speed fluency, and did not reach statistically significant levels. Their findings suggest that while speed fluency measures may be closely linked to language proficiency, breakdown and pause fluency are linked to several other influences including individual learner differences (Tavakoli, Nakatsuhara, & Hunter, 2017).

It therefore seems appropriate to draw the conclusion that increased task complexity does have a negative impact on the oral fluency of intermediate L2 speakers. The important contribution this study makes, however, is the finding that a decrease in fluency at the mid-clause position occurs as a result of increased cognitive task complexity.

8.4.1 Reasons for L2 Loss of Fluency (RQ 1)

There has been a focus in research on the alteration of the cognitive demands of a task and the predictions of the impact this will have on learners as they produce speech. Conflicting theories have emerged as to what the impact of this increase in cognitive load will have on speech production in regards to the CAF framework. Skehan's (1998) Trade-off hypothesis claims that limited attentional resources come into competition with each other once a threshold has been reached, and Robinson's (2001a) Cognition Hypothesis states that attentional resources may be assigned to improve both accuracy and complexity simultaneously. In regards to the present study we can attribute the decreases in oral fluency to the increased task complexity, the cognitive demands required for successful task completion are too great, and speech production breaks down. L2 speech demands too much of the cognitive resources available to complete both the more complex task, and produce fluent speech. This result is in partial support of both the trade-off hypothesis and the cognition hypothesis, as mental resources are directed towards task completion and away from performance fluency decreases. One possible explanation, in light of the results of this study, is that there may be some type of 'threshold' for task complexity. Increases in task complexity may first impact speech production in the formulation stage, and be responsible for the breakdowns in speech at the mid clause location, with further increases in complexity affecting the conceptualisation stage of production. The results of this study highlight this as an area of possible further investigation, but without more detailed examination it remains problematic to draw any concrete explanations in regards to Skehan's and Robinsons' hypotheses.

Kormos' model (2011) of L2 speech production is important to explain the results as it provides a structure to help define and identify the stages of L2 speech production, and the underlying mental processes being utilised as L2 speakers attempt to produce more fluent speech. Furthermore, Kormos' model also helps to provide information on why L2 speech production breaks down and where. A key difference in L2 speech production is that the L2 speaker's mental lexicon is still developing; it may not yet be well organised, the amount of formulaic language contained may still be growing, and consequently it will take more time and effort to access (Bolibaugh & Foster, 2013). This means that the demands of the pre-verbal plan are much harder to satisfy, and may not be met at all. Resultantly the formulation stage of speech production requires more attentional resources, meaning there are fewer mental resources available for allocation at the conceptualisation and articulation stages of speech production. This increased cognitive demand can 'overload' L2 speech production and result in a speaker being forced to switch to serial processing, where problems must be addressed at each stage separately before parallel processing can resume (Kormos, 2006; Skehan, 2014b). This means that the formulation stage of speech production can often act as a 'bottle-neck' for speech production and increased cognitive demands can result in loss of serial processing, resulting in slower speech, more pauses and less oral fluency (Segalowitz, 2010).

The current study recorded the impacts of this increased mental load through more frequent pauses occurring within clauses for L2 speakers. To understand why the L2 speakers are impacted more at the mid-clause we can refer to Skehan et al., (2016), who draw a distinction between pauses which occur in the middle of clauses and those situated at the end of clauses. The suggestion is that those pauses which occur within clauses are evidence of breakdowns in microplanning and the processing of speech production at the formulation stage, whereas the pauses occurring at the end of clauses, at the *discourse* level, provide evidence for difficulties in macroplanning at the conceptualisation stage of speech production (Skehan et al., 2016). In other words, pauses in the middle of a clause stem from difficulties in making lexical, as well as possibly morphosyntactic choices, and suggest evidence of problems with language decisions. Pauses at the end of clauses on the other hand are when a speaker has difficulty conjoining units, or developing an argument, and provide evidence of problems in making discourse decisions (See section pg. 175 for a more detailed discussion). Other studies have also linked pauses within clauses to mental processing difficulties with speech production (Kahng, 2014; Tavakoli, 2011), and as evidence of reductions in cognitive fluency (Kahng, 2017; Kahng, 2020).

In relation to the results of this study we can explain the loss of fluency at the mid-clause through the increased complexity necessitating extra attentional resources to focus on task completion, meaning less resources are available for speech production (Tavakoli, 2011; Skehan, 2009; Tavakoli & Foster, 2008). In South Korea, where traditional language teaching methods are employed, particularly in a test focused environment such as university (Jeon & Hahn, 2006), it could be argued that learners have not yet developed enough communicative skill through language practice to proceduralise their language production. The learners may be reliant on knowledge based routes to fluency, through a well-developed mental lexicon, and firm grasp of the grammatical structures, however they may need to acquire higher levels of proficiency to access and produce fluent speech. Instead, leaners are left to employ a variety of communicative strategies to try and meet the demands of the task and speech production, which often results in pauses and breakdowns in fluency.

This provides evidence to support the theory that the pauses at the mid-clause occur as a result of difficulties in online task performance increasing the load on the cognitive processes necessary for task completion (Kahng, 2014; Skehan & Foster, 2008; Tavakoli, 2011). The theory that mid-clause pauses are also evidence of difficulties in macroplanning and the formulation stage of speech production (Skehan et al., 2016) is also partially supported by these results, however this is much more difficult to accurately prove.

8.4.2 Mid-Clause Pause Frequency - L1 (RQ 1)

The results for the L1 speakers show no statistically significant increase in the frequency of mid-clause pauses for the cognitively more complex task, when compared with their performance on the less complex task. This again provides support for the prediction that the L2 speakers will be impacted by a loss of fluency at the mid-clause, but the L1 speakers will not. These findings are supported by the results of previous studies including Skehan et al., (2016.

The L1 speakers are able to balance the mental requirements for fluent speech production with the extra attentional resources required of the more complex task due to the automaticity of their speech production. We don't require conscious effort to make linguistic choices when we speak in our L1, there is no need to consciously attend to each sentence or to consider the correct grammatical form, nor are we consciously concerned with pronunciation. This is as a result of the way in which an L1 is learnt, which is largely acquired implicitly and forms procedural memories which are not effortless and not open to introspection (Han & Finneran, 2014). This is often made evident when people can speak fluently in an L1, but are not always able to explain why a particular sentence is correct, or a given grammatical structure is incorrect. Implicit knowledge of the language and its grammatical structures when combined with automatic speech processing abilities makes production in an L1 less cognitively demanding (Kormos, 2011). Most people have never given any conscious thought to formulating a sentence, or to articulating the correct sounds, and to produce fluent speech they do not need to. It should be noted that disfluencies are of course evident in L1 speech; arising from difficulties with certain topics, and in particular contexts, but generally speech production in an L1 is a largely automatic process. Importantly this frees up attentional mental resources for other cognitive processes; the increased task complexity, and the manipulation of the task input information in to an appropriate chronological order.

8.4.3 Mid-Clause Pause Frequency – L1 and L2 (RQ 1)

To provide a possible explanation for the increase in frequency of mid-clause pauses for the L2 speakers, but not the L1 speakers we can look to Skehan et al's., (2016) distinction between ‘language decisions’ and ‘discourse decisions’. The tasks in this study operationalised increased task complexity cognitively, not linguistically more difficult, but rather required extra mental processing to “get the whole picture” (Appendix H) as one of the L1 participants put it their post task questionnaire. The discourse decisions have been made more difficult, but the language decisions remain the same. One may expect that the L1 and L2 speaker’s performance would be impacted to the same degree, as the L1 speaker’s superior linguistic skills and knowledge are of little benefit to them in ordering the information chronologically. However, the difference in task performance, and speech production comes when the L2 speaker’s mental processing capacity has been reached (Kahng, 2014). They no longer have the necessary cognitive resources to produce fluent speech, while at the same time ordering the information in the task. This means parallel processing is no longer possible for the L2 speakers, the stages of speech production can no longer run simultaneously (Kormos, 2006). Instead, they are forced to switch to serial processing, which as was described earlier, requires difficulties and problems with each stage of speech production to be resolved individually before parallel processing can be resumed (Kormos, 2006; Skehan, 2014b). The increased demand for attentional resources which results in a switch to serial processing for L2 speakers is further exacerbated by the need to re-launch fluent speech, with requires yet more mental attention. The speaker needs to pay close attention to regain the parallel processing mode of speech production, and resume fluent speech production. It is often much more difficult and requires greater effort, often necessitating a longer pause, to restart speaking once speech production has been interrupted.

L1 speakers on the other hand have sufficient attendant mental resources to meet the demands of the increased task complexity. They are able to make the extra macroplanning, discourse level decisions, at the conceptualisation stage of speech production while still having enough mental resources left to produce fluent speech in the formulation stage as this occurs automatically. The difference then in processing demands is largely due to the allocation of mental resources required for L1 speech production at the formulation stage. As was explained in chapter 2 speech production benefits tremendously from the automatization of formulation in a first language.

Another key area of difference between L1 and L2 speech production which impacts cognitive fluency is the way in which some grammatical and phonological knowledge is stored (Samuda & Bygate, 2005). L2 speakers must access some linguistic knowledge from their long term memory, this is because knowledge which is learned explicitly is classified as declarative knowledge and requires conscious effort to recall (Kormos, 2011). However, in an L1 this process is automatic, and requires no conscious effort on the part of the speaker. L1 speakers do learn some vocabulary in a declarative fashion, an often cited example is the way in which a child's L1 vocabulary expands exponentially upon entering school for the first time. However, in relation to skills based theories of language acquisition, the way in which this initial declarative knowledge develops is an important difference in L1 and L2 speech production. For both L1 and L2 speakers declarative knowledge can become procedural through continual and repeated implementation (DeKeyser, 2017). This is what Anderson (1982) referred to as the 'compilation of knowledge', where skills move from declarative to procedural, through the collapsing of longer sequences of production into single sequences (Anderson, 1982). This process frees up mental resources during formulation allowing for more resources to be directed towards task completion, as well as the other stages of speech production. For the intermediate L2 speakers in this study, accessing declarative knowledge is still a necessary requirement of speech production, as they may not yet have been able to fully proceduralise their second language speaking skills. This means they will have less attentional resources available during formulation, and when this is combined with the greater task complexity, they may be likely to pause more in the middle of clauses.

A further significant difference between L1 and L2 speech production which may help to explain the increased L2 speaker pause frequency at the mid-clause is lemma retrieval. Segalowitz (2010) points out that despite several studies exploring lemma retrieval (De Bot, 1992; Kormos, 2006), it is still largely unknown whether or not an L2 speaker retrieves a lemma directly in the L2, or if they first need to access in in their L1, and then translate it. Early work by Talmy (1985) on lemma activation suggests they are language specific when used for particular lexicalisation patterns of a given language (Wei, 2017), but this view is still somewhat contentious and further investigation into lemma activation and retrieval is required. However, if direct lemma retrieval is not possible in the target language then a further step is imposed on L2 speech production, with increased microplanning required in formulation.

Difficulties in the articulation stage of speech production also require more mental processing for L2 speakers when compared with their L1 speaking counterparts (Levelt, 1999; Segalowitz, 2010; Kormos, 2006). The phonological score needs to be converted into an articulatory score, but can negatively impact L2 speech due to unfamiliarity (Segalowitz, 2010). In reference to the present study, which employed intermediate level Korean speakers of English as participants, it is important to note the difficulties inherent in articulation. These difficulties often arise due to the differences in pronunciation patterns in English and Korean, but also because of the difficulties in acquiring L2 phonology and phonetic articulatory skills. These difficulties are further compounded by the presence and ubiquity of English loan words in common usage in modern commonly spoken Korean. Korean L1 speakers will regularly use words such as *computer*, *pizza*, or *coffee* when speaking in Korean, and will access their gestural score for them from their L1. When speaking English as an L2 however, they will perhaps need to access their L2 gestural scores for these same words, which have markedly different stress, intonation, and syllabic stress when pronounced in English. Articulation then will draw further attentional mental resources away from the other stages of production and may contribute to the results seen here.

8.5 End-Clause Pause Frequency – L2 (RQ 2)

The second research question addressed the impact of a variation in task design and complexity on the production of oral fluency measured at the *end*-clause level. The prediction for question 2 stated that an increase in task complexity would lead to a reduction of oral fluency for L2 speakers at the end-clause level, but the impact would be less than that recorded at the mid-clause. In contrast to RQ 1, where an increase in pauses at the mid-clause signals decreased oral fluency, an increased frequency of pauses at the end-clause does not necessarily indicate a decreased level of oral fluency. This is because pauses at the end of clauses do not disrupt the flow of speech, and several studies have shown that pausing at the end of a clause does not impact perceptions of oral fluency, whereas pauses at the mid-clause do (Kahng, 2017; Suzuki & Kormos, 2019).

The L2 speaking participants recorded no significant difference in pause frequency in comparisons between the more and less cognitively complex tasks. Slightly more pauses were recorded at the end-clause for the less complex task, but this was shown to be a statistically insignificant difference. In relation to the hypothesis for RQ 2 the findings offer partial support, there was no significant change in pause frequency recorded at the end-clause level, but in relation to the second part of RQ 2 the L2 speakers did see less impact than was recorded at the mid-clause.

Pauses occurring at the end of clauses have been described in terms of difficulties with speech production in the conceptualisation stage and with difficulties macroplanning, and with adding clauses and ideas together to construct a coherent narrative (Kahng, 2017; Skehan et al., 2016), or make discourse level decisions (Skehan et al., 2016).

As no significant increases in pause frequency were recorded for the L2 speakers we can draw the conclusion that increasing cognitive task complexity does not impact end-clause pauses in any significant way. So why do the L2 speakers pause more at the mid-clause, but not at the end-clause position? To answer the question, we can refer to the results of this study, as well as the literature on the impacts of task complexity (Skehan et al., 2016; Kahng, 2014; Tavakoli, 2011; Kormos, 2006; Segalowitz, 2010). We might expect to see more frequent pauses at the end-clause position as a result of increased attention demanded at the conceptualisation stage, as conceptualising involves macroplanning and is related to discourse decisions. However, the frequency of pauses at the mid-clause actually increases as a result of these greater demands on the conceptualisation stage. This may be because a speaker has limited attentional mental resources, and attention is primarily focussed on conceptualising and on 'discourse level' decisions, rather than formulation and 'clause level' decisions (Skehan et al., 2016). This imbalance is a result of extra attentional resources being required to conceptualise speech being drawn away from formulation, and resulting in difficulties because of insufficient attentional mental resources for microplanning. The formulation stage of speech production then remains what Segalowitz (2010) refers to as a 'vulnerable point' for L2 speakers, and is an area of speech with demands greater attentional effort with increased processing demands which can lead to the slowing down of speech, and ultimately a reduction in fluency. It could be argued then that L2 speech production, and the way in which attentional mental resources are governed, prioritises to a certain extent *what* is said in conceptualisation over *how* it is said in formulation.

8.5.1 End-Clause Pause Frequency – L1 (RQ 2)

The increased task complexity did not result in a significant change in the frequency of pauses occurring at the end of clauses for the L1 speakers. However, the L1 speakers did record slightly more end-clause pauses on the less cognitively complex version of the task, though again this difference was shown to not be statistically significant. This result was counter to the hypothesis and the predictions made, and may simply be because of individual speaker patterns within the relatively small sample size. Another possible explanation however is that pausing more at the end clause is due to the task being easier for the L1 speakers. This may seem counterintuitive, but literature has shown that pauses at the end clause position do not impact perceptions of oral fluency as much as those mid clause (Khang, 2008, and Another, 2022), so this may be a conscious, or indeed unconscious, effort to deliver speech in a 'normal' fluent fashion and following an expected and conventional pattern of speech. Speakers may be choosing to avoid making pauses in the mid clause position by taking more time to perform 'discourse level decisions' (Skehan et al, 2016), and instead pausing at the end of clauses. It is possible that the reduced task complexity allows the speaker this freedom due to the decreased cognitive load on speech production, and they are using the extra time to plan the narrative they are performing. The slight increase in end clause pauses may also be due to nature of the task input; which each being broken down in to pieces or segments, the 8 separate pictures in the pictorial task, or the 14 bullet points in the written task. This fragmented nature of task input may lend itself to more end clause pausing overall, but particularly when the cognitive demands of the task are less because this allows more planning time and results in easing of conceptualisation. This result does partially support RQ 2 in that end-clause frequency did increase slightly, but there was no significant loss of fluency at the end-clause for the L1 speakers despite the increased task complexity. It may be that the L1 participants actually produce longer runs of speech, or say more when they spend more time in conceptualising speech. This is however beyond the scope of the current investigation and remains a possible area for future study.

8.5.2 End-Clause Pause Frequency – L1 and L2 (RQ 2)

The results for the frequency of end-clause pauses were similar for both the L1 and L2 speakers, with none significant decreases for the L2 participants, and none significant increases for the L1 speakers. A possible conclusion drawn from the results, is that the L1 speakers prioritised creating the narrative and telling the story itself, whereas the L2 speakers were more focused on the language necessary to complete the narrative.

8.6 Impacts of Task Complexity Summary (RQ 1 and 2)

To briefly summarise, the comparatively higher frequency of pauses at the mid-clause, and lower frequency at the end-clause, for L2 speakers is expected and provides evidence of not only the impact of increased demands due to increases in task complexity, but also the difficulty of speech production in a second language (Skehan et al, 2016; Kahng, 2014; Tavakoli, 2011). The entire process of speaking in a second language is less automatic, requiring more conscious effort, and results in more pauses and disfluency than occur in a first language. Breakdowns in fluency at the mid-clause are symptomatic of difficulties in the formulation stage of speech production, where the linguistic content of a message is selected, an area of speech production which is particularly difficult in an L2, often complicated by the accessing of declarative knowledge stores. L2 speach also lack the automaticity of speech production in an L1, and do not have access to the same detailed, elaborate, rich, and well organised mental lexicon or the same repertoire of formulaic language (Skehan et al 2016).

8.7The Impacts of Input Format (RQ 3)

The research question posed for the investigation into task input type was: What is the impact of a variation in task input format, here operationalised as either written or pictorial, on the performance of L1 and L2 speakers' oral fluency at the mid-clause level? The prediction for this research question is that the task with the written input will see higher levels of oral fluency at the mid-clause, a lower frequency of pauses, amongst L2 speakers when compared with the pictorial input task. The prediction is for the written input provided by the task to ease lemma selection and access to the mental lexicon in the formulation stage of speech production, and the mental load will be decreased. However, this will not be the case for the pictorial input tasks where no vocabulary or syntactical data is presented. L1 speech production and fluency will be unaffected by input mode at the mid-clause. This is because grammatical and phonological encoding are effortless in an L1.

Firstly, looking at just the L2 speakers the results of a differentiation in task input format showed that they did not pause more frequently at the mid-clause position on either the written input task or the pictorial input task. In fact, the findings were remarkably similar across both input formats, with the L2 speakers recording an almost identical frequency of pauses at the mid-clause for both tasks. These findings show that the differentiation in task input format had little or no impact on the L2 speaker's frequency of pauses at the mid-clause. Secondly, in regards to the L1 speakers this study also showed that there was no significant difference in pausing frequencies at the mid-clauses for the L1 speakers between the two task input formats. A very slight increase in mid pause frequency was recorded for the L1 speakers on the pictorial task, but this was found not to be significant. So, neither the L2 or L1 speakers paused significantly differently at the mid-clause

For L2 learners the importance of the impacts of task characteristics have long been explored in the literature (Kim, 2009), with many reaching the conclusion that specific task elements can help encourage the production of complex, accurate, or fluent speech. For the written task it was hypothesised that as the task input provided linguistic input in the form of written language this would 'lexically prime' (Hoey, 2012) the L2 speakers allowing them to ease the mental load during both macroplanning and microplanning. The idea behind lexical priming is that whenever words, syllables, or groups of words are heard the listener subconsciously takes note of the linguistic context in which they occur, and further repetition enables the listener to increasingly notice the features of the context which are also being repeated (Hoey, 2012). The vocabulary provided for the written task includes: *born, moved, married, divorced, quit*, etc., all words which will have been encountered before and be well known to the Korean learners of English, and as such should help provide easing of the mental processing, by providing some linguistic context. The effect lexical priming should facilitate, is the L2 speakers utilisation of the vocabulary within its appropriate linguistic context, and allow for easier speech production at the formulator, resulting in a lower frequency of pauses at the mid-clause position.

One difference between the designs of the tasks employed by this study, which may have inadvertently increased task complexity, is the requirement for participants to utilise background information that is not inherent in the task input. As an example, the picture task contains gaps in information that the participant must try to fill in, these elements such as: why the character is late (*to use examples from the participant's responses* - he is sick, he overslept, he drank too much), as well as where he is going (*again examples include* - to work, to a meeting, to a job interview) are all open to interpretation in the task pictorial input. Generating these details may tax the L2 speakers at the conceptualisation stage, with macroplanning requiring of more attention than when completing the written task, which does not require the participant to comment and speculate on why the participant dropped out of high school, or got divorced, etc. Several studies (Tavakoli & Skehan, 2005; Tavakoli & Foster 2008, 2011) have shown that requiring learners to reflect on information not inherent in the story when completing a narrative increases task complexity.

The L1 speakers did produce more fluent speech on the written input task, but only very slightly, with less frequent mid-clause pauses recorded, so it is possible that the linguistic data may have benefited them more than the L2 speakers. The processing advantage of L1 speakers when using familiar phrases is well documented (Libben & Titone, 2008; Tabossi, Fanari & Wolf, 2009) and it has been argued that in an L1 any linguistic material which has been encountered previously can be accessed via a direct route, hastening lemma retrieval (Wray & Perkins, 2000). It may be possible that the linguistic data provided in the written task input benefited the L1 speakers more than the L2 speakers, as they will have encountered these words and phrases before and be aware of the collocations, as well as the idiomatic and formulaic language associated with the task. As was noted in the section detailing the piloting (5.12) of this study the L1 speaking participants suffered from unfamiliarity with task based learning, and the performance of constructing spontaneous narratives in regards to language practice was novel to the vast majority of them. It is possible that this difficulty in task transaction may have been mitigated somewhat by the linguistic data supplied in the written version of the task, allowing the L1 speakers a ‘way in’ to begin the task, and providing them with some much needed structure and direction.

In the post task questionnaire completed by L1 participants on the difficulty of transacting tasks, results showed that the more complex written task was considered to be the most difficult, but that the less complex version of the written task was regarded as the easiest. These findings appear to suggest that the task format did not play a significant part in the L1 speaker’s perceptions of task complexity, and that task input format may interact with operationalisations of task complexity to a certain extent. This is by no means clear however, with several possibly confounding variables and a relatively small sample size, and is requiring of further investigation. As was explained earlier L2 speakers did not complete a post task questionnaire on perceptions of task difficulty. The questionnaire was primarily employed to establish levels of cognitive difficulty, and it was felt that this would be best achieved by removing the ‘language’ element form the task with the L1 speakers. Any responses from the L2 speakers may well have been confounded by perceptions of the difficulty of the language used in the task.

The L2 speakers on the other hand, will also almost certainly have encountered and employed the vocabulary provided in the task input before, but rather than help the mental processing at speech production it may have in fact impaired them. In a direct comparison with the pictorial task, where there is no linguistic information provided, the participants are free to construct a narrative using whatever linguistic resources they are most comfortable with. The L2 speakers do not need to refer back to the task input to use the correct language, they can use any vocabulary and linguistic resources with which they are comfortable. L2 speakers may employ communicative strategies to avoid language which they are unsure of, helping to ease speech production in the formulation stage and avoid possible breakdowns. This flexibility of performance afforded by the less tightly constructed tasks may encourage more fluent speech and allows for the generation of more dependable results (Skehan et al., 2016). The written input task in providing vocabulary helps to scaffold the narrative created, but it also imposes a structure which may act to constrict the L2 speakers and requires them to follow a certain narrative path. It is possible then that the written input provides both negative and positive impacts on the cognitive processing demands in L2 speech production.

When drawing conclusion on the influence of the role of task input format in L2 oral fluency it is difficult to determine precisely, from the results obtained in the present study, to what extent the input format taxes or eases the mental processing of speech production in spontaneous narrative production, for either the L1 or L2 speakers. However, it would seem to have no significant impact on the frequency of pauses at the mid-clause position in this instance. Whether this is because of a balancing of the task demands, as a result of lexical priming, or for some other as yet unknown reason, it is clear that further research is required to better understand which elements of task input format facilitate the production of fluent L2 speech (Skehan et al., 2016; Tavakoli & Foster, 2008; Tavakoli & Skehan, 2005).

8.8 The Relationship Between PSTM and Mid and End-Clause Pause Frequency (RQ 4)

The fourth and final research question explored the impact of L2 speaker levels of phonological short term memory (PSTM), operationalised here through a digit span test, on levels of oral fluency again measured at the mid and end-clause positions.

The prediction is that L2 speakers with lower levels of PSTM will pause more at the mid-clause this is based on the belief that having higher levels of PSTM will facilitate the production of fluent speech, reflecting a speaker's ability to better deal with phonological information required for task completion. Those learners who have the ability to hold and retain more phonological information (the knowledge of the phonemes that used to create words and sounds in a language) in the short term will be better able to produce fluent speech. Conversely those speakers who have lower scores on the PSTM test will see reductions in oral fluency during the more complex task at the mid-clause (Baddeley & Vallar, 1984; O'Brien et al., 2007; Van de Noort et al., 2006; Williams, 2012).

The study recorded results which provide evidence of a strong and clear correlation between the L2 speaker's PSTM score and pause frequency at both the mid and end-clause. Firstly, at the mid-clause position a highly statistically significant, strong negative correlation was found between PSTM and the frequency of mid-clause pauses. In other words, those speakers who had lower levels of PSTM paused more at the mid-clause than those with higher levels of PSTM. Secondly, at the end-clause a strong, and statistically significant positive correlation was found between PSTM scores and pause frequency at the end of clauses. So higher levels of PSTM resulted in more frequent pauses at the end of clauses. In summary the findings provided evidence that as levels of PSTM increased in L2 speakers the frequency of mid-clause pauses decreased, but the frequency of end-clause pauses increased.

Having a better PSTM allows learners to hold and process more phonological information as they are constructing speech. In reference to Levelt's model (1989) we know that during the formulation stage of speech production phonological encoding takes place after the surface structure has been compiled through grammatical encoding. When generating the phonological score for articulation the speaker needs to link each syllable to the appropriate articulatory gesture. As the formulation stage of speech production is where the phonological detail is added to the preverbal plan, it is conceivable that this is where higher levels of PSTM may help to facilitate the higher levels of oral fluency measured at the mid-clause level, which have been detailed in this study. Having the ability to hold on to and maintain a greater amount of phonological data in one's mind during speech production, allows the L2 speaker to ease the processing load at formulation, and free up mental resources for task completion and the conceptualisation stage (O'Brien et al., 2007).

Possible explanations for these results are that learners are utilising higher levels of PSTM to plan ahead, and to try 'manage' the cognitive load inherent in speech production, and make the formulation of speech easier, and more fluent. Selecting to pause at the end of clauses and giving themselves time to think ahead and to conceptualise their next utterance. To test this theory, further research would need to focus on the language employed directly after these long end clause pauses, and ascertain to what extent speakers employ more complex or more lexically primed output. A qualitative exploration of the language employed, including correlations with any formulaic chunks or idiomatic language use may also provide greater insight into reason for the pausing behaviour. An investigation into the pausing patterns of individual L2 speakers may reveal a relationship between the decrease in mid clause pauses and the increase in end clause pauses. This could allow for a closer look at if the two pauses are related, and mutually beneficial. Finally, an attempt to corelate PSTM measures with the pausing patterns mentioned above may well provide valuable insight to the role it plays in pausing, and the ability to plan ahead during speech production.

The articulation stage of speech production is another area where PSTM may benefit the production of more fluent speech at the mid-clause. As Levelt (1989) points out speech which is generated internally is unvoiced, and as a result is often much faster in production than the physical process of articulating speech. This necessitates the holding, or storage, of the phonetic plan in what Levelt (1989) refers to as the ‘articulatory buffer’. It is here in the articulatory buffer where chunks of internal speech are retrieved and used in turn to create overt speech. It is conceivable then that higher levels of PSTM are evidence of a greater capacity of storage, allowing for larger amounts of internal unvoiced speech to be held in the articulatory buffer (O’Brien et al., 2007). It may even also be the case that higher levels of PSTM provide for quicker access to storage (O’Brien et al., 2007), smoothing the process of accessing and retrieving information from the phonetic plan in the articulatory buffer. Whichever may be true it would be a clear benefit for L2 oral fluency, and speech production in general, to have a larger storage of information and faster access to it. Higher levels of PSTM may then provide an easing of the cognitive load and allow for the production of more fluent speech. It is also possible that individual differences in levels of PSTM may account for variations in dimensions of linguistic production, which may be partially responsible for the inconclusive observations reported in studies exploring the role of task complexity in oral fluency. This idea is also supportive of the cognition hypothesis (Robinson, 2007) that differences in working memory and attentional capacity are most important in relation to the completion of more cognitively demanding tasks.

Through several studies into young learners and children Baddeley (2000, 2003) has demonstrated repeatedly the value of PSTM in the language acquisition in a first language. However, research has also shown that as the age of a learner increases, and with it their language proficiency, the role which PSTM plays diminishes, suggesting that PSTM plays a role in the initial acquisition of vocabulary in the first language. More recent studies (O’Brien et al., 2007) have shown that PSTM also plays an important role in the development of oral fluency in adults, highlighting the role PSTM plays in not just acquiring new vocabulary in a second language, but also the performance and repetition of these phonological elements. It may be possible then that gaps in the L2 are concerned with retrieving vocabulary and other phonological data. PSTM may act to function in the more efficient retrieval of recently spoken or heard language, ultimately facilitating the production of more fluent speech. The present study reinforces the findings of O'Brien at al., (2007) and provides evidence for PSTM being an important factor in the production of language, and not just its acquisition. Further research may explore to what extent PSTM can facilitate fluent speech production and how this can be related to the implementation of novel vocabulary, to better understand the relationship between the two. As previous research has shown levels of proficiency also appear to play a large part in the role that PSTM plays, so this is another area of potential research, with obvious implications for SLA and fluency.

In line with the previous findings in this study on how mental resources and attention are directed during speech production higher levels of PSTM may facilitate the management of the cognitive load on L2 speakers, easing speech production in the formulation and articulation stages of speech production. The results of this study show a strong correlation between higher levels of PSTM and higher levels of oral fluency in L2 speakers, suggesting that PSTM plays an important part in the smooth production of speech in a second language. However, further detailed research is required into how the fluent production of L2 speech is impacted by levels of PSTM.

8.9 Theoretical, Methodological, and Pedagogical Implications

In the following sections I will discuss the impact of the current research project in relation to the theoretical, methodological, and pedagogical implications it provides. The concluding section provides a summary on the extent to which the findings obtained in this study are generalizable to other contexts, as well as a discussion of the limitations of the study, and concludes with suggestions for possible areas of future research projects.

8.9.1 Theoretical Implications

The report detailed here is one of the first to investigate the impacts of a variation in cognitive task complexity and task input format on a comparison of L1 and L2 oral fluency measures at the mid and end-clause position for spontaneous narrative production. It is hoped that the study provides fresh insight in how influences on oral fluency are understood, measured, and compared. The comparison with L1 speaker speech production is integral to the design, as it provides a *baseline* of sorts with which to enable a sharper focus on the differences between how L1 and L2 speech is produced.

The most apparent conclusion drawn from the study is the link between levels of task complexity and the frequency of pauses at the mid and end-clause positions, and what this implies for the cognitive processes employed during L2 speech production. Pause location has been linked to perceptions of oral fluency (Kahng, 2017), utterance fluency (Wittan & Davies, 2014; Vallas & Ferrer, 2012), and to cognitive processes of speech production (Tavakoli, 2011). The findings provide support for the theory put forward by Skehan et al., (2016) that increased task complexity impacts the production of fluent speech at the clause level for L2 speakers, but not for L1 speakers. These findings suggest that increased cognitive complexity results in greater difficulty at the formulation stage of speech production, and possibly with microplanning, as evidenced by the increased frequency of pauses at the mid-clause position. If we trace the process, we start with the operationalisation of task cognitive complexity, here the +/- order of information (Robinson, 2007), which requires extra attentional mental resources at the conceptualisation stage of speech production (Robinson, 2003; Skehan, 2009; Tavakoli & Foster, 2008), and possibly with macroplanning (Skehan et al 2016). This increased demand for attentional resources draws focus away from speech production in the formulation and articulation stages of L2 speech production (Kormos, 2011; Segalowitz, 2010), and specifically with microplanning (Skehan et al., 2016), and it often results in L2 speakers resorting to serial encoding (Kormos, 2006). The lack of attentional resources available to formulate speech results in the increased frequency of pauses at the mid-clause location (Kahng, 2014; Tavakoli, 2011), as evidenced in the findings of this research.

As was discussed in chapter 2 there are myriad different ways to approach the measurement of oral fluency, with difficulties in comparing results due to differences in the specifics of many of the measures employed. It is hoped that this new perspective on ways of thinking about and measuring fluency may allow for some of the disparate measure to be condensed, and more consensus on measures of oral fluency employed in future research.

Any conclusions which can be drawn from the examination of the impacts of a variation in task input format on pauses occurring at the mid-clause are less obvious. As there is no clear, direct impact on the frequency of mid or end-clause pauses, it is hard to outline any conclusive theoretical implications with the levels of task complexity or increases in cognitive load. However, it is apparent that different elements within this particular research project had both positive and negative impacts on task performance; with some elements likely easing the cognitive load, and yet other aspects of task design increasing the cognitive demand during task performance. Requiring learners to generate information not inherent in the task input for task completion increases complexity (Tavakoli & Skehan, 2005; Tavakoli & Foster 2008, 2011), and requires extra mental processing. This would then increase task complexity for the pictorial task employed by this study. On the other hand, the vocabulary provided in the written task input may also have increased task complexity. It is possible that providing detailed linguistic information (names, dates, times etc.) at input denies learners the freedom to create a narrative appropriate to their level. Incorporating the vocabulary from task input may deny participants the chance to employ communicative strategies and avoid potential road blocks in speech production. However, it is also possible that lexical priming occurs when a task is comprised of written input. The effect of lexical priming may facilitate the L2 participant's implementation of the vocabulary in the correct linguistic context, easing mental processing at the formulator and allowing more rapid speech production with fewer pauses (Hoey, 2012).

The purpose of this discussion in relation to task input format is not to point out every possible conceivable difference and nuance in task input, but rather to underline the importance of carefully analysing each aspect before any conclusions are reached. The conclusion drawn here is that more research is needed into the impacts of task input format, as seemingly benign aspects of task input may have both positive and negative effects on oral fluency. Each of the individual elements of task design and input should be clearly defined, categorised, and analysed for the impact it has on the cognitive complexity of tasks, speech production, and on the location of the resultant pauses.

Regarding PSTM, the main theoretical implication is the support provided from the findings of this research which underline the importance of the role played by PSTM in older and more proficient L2 language learner's ability to produce speech fluently (O'Brien et al., 2007). The findings of the present study provide evidence that PSTM plays an important part in the formulation of sentences and with linguistic performance in older and more proficient learners. This importance is evidenced by the strong negative correlation found between levels of PSTM and the frequency of mid-clause pauses. The findings also point to the fact that higher levels of PSTM correlate positively with end clause pauses. This was an unexpected result, but suggests there is more to the relationship between PSTM and oral fluency than precious research has shown.

I suggest two possible explanations for the findings in relation to PSTM. Firstly, that the higher levels of PSTM facilitate forward planning in speech production, and the participants are 'buying time' to prepare more complex speech turns, hence the more frequent end clause pauses and decreased mid clause pause. Secondly, that those pauses are evidence of L2 speakers searching for and accessing vocabulary from their mental store, and higher PSTM levels facilitate this process.

The research findings lend further support to Robinson's cognition hypothesis (2007) and the idea that PSTM plays a more important role in the performance of cognitively complex tasks. The role of PSTM, and the individual differences between performances on tasks, provides a possible explanation for the variety of observations recorded in different reports of the impacts of task complexity on L2 oral fluency. The impacts which PSTM has on the articulation stage of speech production may also provide insight into how increases in oral fluency affect pauses at the mid-clause. This is due largely to the difference in the speed in which voiced and unvoiced speech is generated, with the internal speech having to be 'held', or stored while overt speech catches up through articulation (Levelt, 1989). As the internal speech is stored in the articulatory buffer, and retrieved in chunks as needed, higher levels of PSTM may facilitate the storage of larger amounts or the faster retrieval of internal speech (O'Brien et al., 2007). This study has provided evidence that the role which PSTM plays in adult, and more proficient, L2 learner's speech production and linguistic performance is an important one, and it offers an insight in to a greater understanding of PSTM and the way in which it interacts with speech production, and in particular with oral fluency measured at the mid-clause.

8.9.2 Methodological Implications

The distinction of those pauses at the mi and end clause is an important and hopefully this comparison can be employed to gain a more accurate understanding of what is happening to L2 fluency during speech production, and which of the attendant mental processes are affected, and how, by changes in the task input format and design. Task design and how it affects fluency is a hugely important area of research with obvious implications for how tasks themselves are designed, but also for how syllabuses are structured, and individual lessons are delivered. Any investigation and increased understanding of how these various aspects of task design interact would be of great importance to future studies, not just of task design, but oral fluency as well.

It is also possible that these findings could enable future research to investigate the role of mid-clause pauses in fluency development. Whether or not mid-clause pausing frequency decreases with increases in L2 proficiency may be able to provide an indication of L2 levels of proceduralisation, automatization, and that the processing in the formulator is improving and the cognitive load is being handled more efficiently.

The research study design also saw the random allocation of L2 participants from several complete classes in to one of the two groups. Taking learners from complete classes and randomly assigning them to groups, as well as conducting the tasks outside of the classroom setting, helped to counter sample bias due to what Marsden (2007) refers to as ‘inter group correlation’ and what Torgerson and Torgerson (2003) term ‘cluster effects’. The setting in which the class takes place, as well as the composition of the class, individual learner’s experience, can all cause problems and lead to confounding results (Marsden, 2007), but the research design employed by the present study helped to alleviate these potential issues.

A further benefit of the research design employed was the counterbalancing of the tasks, in both task complexity and the input format. The tasks were organised so that the half of the learners would transact the more complex task first, and the other half would perform the less complex task initially. However, the tasks were also balanced for input format, meaning half of the participants completed the written tasks first and then moved on to the pictorial task, while the remaining participants performed the opposite. This benefitted the research by not only neutralising any effects concerning the order in which the tasks were performed, but also by controlling for the effect of any 'practice' or learning impacts which may have taken place after completion of the first task (Marsden & Torgerson, 2012).

Another possible area of methodological interest is data analysis, and obtaining sufficient quality and quantity of data from all of the participants. As all learners are different each one will inevitably perform the task in their own way, interpreting it slightly differently from other participants, despite the clarity of the instructions provided and a precise research plan (Dörnyei & Kormos, 2000). The L2 speakers were generally very good in this regard and understood the task and stayed focused, however some speakers were quite uncomfortable and nervous, feeling pressure to 'perform' in a one on one setting with the researcher. As mentioned previously in the section on the pilot study (5.12) the unfamiliarity of the L1 speakers with language learning tasks was also an issue. They often failed to completely fill the allotted time and occasionally had long pauses, where they were unsure of how to proceed, or what was expected of them. Attempts to alleviate this issue were employed in the main task data collection by providing: further more detailed instructions, explaining the context of the task within its role as language 'practice', and detailing the required outcomes and length of response required. Due to the relatively large number of participants in this study, as well as the more nuanced instructions provided it is felt that the impact of these factors have been reduced as much as is possible.

The use of technology in facilitating the valid and reliable measurement of speech data is another area of methodological importance for this study. Increasingly in studies of oral fluency the use of technology has been employed to increase the accuracy of recording temporal measures of utterance fluency. Integral to the implementation of this research study was the software program PRAAT (Boersma & Weenink, 2008) to automatically record the silences in the participant's speech samples (De Jong & Perfetti, 2011). Employing automatic fluency software is not without drawbacks however, as was detailed in the section on the pilot study (5.12) the importance of securing high quality audio, free from background noise is integral to the success of PRAAT and the analysis of silences. I would argue that employing PRAAT in an active language learning classroom would be highly problematic due to the levels of background noise inherent in a classroom setting (Lambert et al., 2017) and the difficulty of focusing on specific language elements. In more practical terms using automated software can 'remove' the researcher from the process of data analysis, and may result in a reduced opportunity to discern patterns or errors and may inhibit the chance to get a feel for the data and the individual participants.

8.9.3 Pedagogical Implications

The pedagogic implications of this research, and in particular for teaching, are obviously lessened as the research was not collected in a language learning environment, but rather a laboratory. However, the research does still provide some tangible insights which may have valuable contributions to the understanding and formulation of syllabus and task design, speaking tests and interviews, as well as language teaching itself.

One of the primary pedagogical implications derived from this study concerns the importance of the order in which tasks are approached in the classroom, and this is a long established area of detailed research in the TBLT and SLA fields. To briefly summarise; the way in which tasks are sequenced within a TBLT framework syllabus is designed to encourage the acquisition process as students are able to consolidate what they have already learned from previous tasks into the performance of future tasks (Long, 2015). The gradual increase of tasks from less to more complex can provide encouragement and motivation for learners, it helps to stretch their 'interlanguage' and forces learners to utilise more advanced language as they strive to meet the challenge offered by the increasingly complex tasks (Robinson, 2011). Presenting learners with features of language they have failed to use correctly in speech production, and expecting them to practice them, will hinder language acquisition as the learners are not yet ready to incorporate them (Ellis, 2003). Task design and syllabus design then could benefit then from a greater understanding of how to better manipulate tasks along cognitive complexity dimensions, as well as the ability to better sequence tasks to grade more accurately the difficulty of the target task. Obviously, the wide variation in learner characteristics, coupled with the individual sociolinguistic settings need to be taken into account, and make the attempt of task sequencing challenging and one which defies a one size fits all recommendation. However, with a better understanding of *what* makes a task more difficult, and *how* it makes it more difficult course designers and teachers are better informed, and more able to sequence tasks appropriately for their students.

Another pedagogical implication of this research is in regards to the way in which SLA is approached in the syllabus of a Korean university setting. My initial interest in oral fluency was sparked from Korean university students repeated desire to communicate more ‘fluently’ in English, and be able to communicate face to face, in real time with English L1 speakers. In the present format many Korean English programmes do not employ a TBLT approach to English, and while a few may employ a ‘task-supported approach to language teaching’, much of the instruction is still delivered following more traditional teaching methods, focusing on grammar and vocabulary. This is in large part what Korean learners expect and are comfortable with, but it is also driven by the extreme number of tests South Korean learners must endure. The successful negotiation of these tests, which includes, but is not limited to: the Korean ‘sunung’ (high school equivalency test), TOEFL, TOEIC, IELTS, university placement tests, employer specific promotional exams, etc. is often the principal motivation for learning English. These tests all, to a certain extent, have a partial focus on grammatical structures, and on understanding linguistic content. So, for a country whose educational philosophy emphasises the learning of knowledge and the passing of tests, rather than skill development a task-based approach may be problematic to incorporate (Ellis, 2009b). If a Korean student is taking a prerequisite English course, which necessitates the passing of a university test, and relies heavily on assessing the student’s grammatical understanding of the target language, then a high level or communicative competency and oral fluency may not be necessary. The specific sociolinguistic setting and the desired learning outcomes of the students should be central to any course design or syllabus, as well as the individual tasks themselves. However, I believe that a task based approach to syllabus design can still benefit Korean university students. It can enable the learners to better apply the grammatical ‘knowledge’ of a language, while still allowing them to improve their ability to communicate more confidently and fluently.

8.10 Limitations

Perhaps the most obvious limitation of this research project concerns the setting of the study and the fact that it was not carried out in a language learning classroom, but rather in a carefully controlled laboratory setting, primarily for the purposes of recording higher quality audio. The study was undertaken outside of a regular English class learning environment, meaning that student performance was not as authentic as those which occur in the classroom. Bygate et al., (2009) caution that any studies investigating task based learning performed in laboratory settings should be treated as case studies, and while proving valuable contributions to the empirical implications of TBLT they require further widespread and pedagogically contextualised research. However, the current research design was carefully chosen according to the literature, with focus on the accuracy of fluency measures, and the theoretical implications, rather than on the teaching and development of fluency. As such it was felt that a laboratory setting was appropriate for the present study, while acknowledging that further investigation within an intact authentic language learning classroom would be beneficial.

A further limitation which must be acknowledged is that only Korean L1 learners of English participated in the study, the vast majority were of a similar age range, were all enrolled at a specialised language learning university, and all had approximately the same L2 proficiency. This homogeneity amongst participants makes the results valid for the specific learner context, but makes the reliability of expanding conclusions and generalising too widely about the findings more problematic. Investigations exploring different target languages, and employing groups from different sociolinguistic backgrounds would be needed to expand on the reliability of the findings of the present study. The level of L2 language proficiency is of particular import, as with higher levels of proficiency comes greater speed and efficiency of speech production through repeated practice, access to more formulaic language, increased automatization of speech, and the expanded range of linguistic resources available (Tavakoli, 2011; Segalowitz, 2010).

The automatic analysis of speech data with PRAAT it is not without several issues worthy of mention. Firstly, the automatic counting of syllables is a possible area of contention as it can be argued that this is often not entirely accurate for the recording and calculating of L2 spontaneous speech. This is because of the variation in the way L2 speakers pronounce words in the target language. A relevant example would be the way in which an L1 speaker of English would be expected to pronounce the word 'English', using two syllables /ˈɪŋglɪʃ/. However, a Korean L1 speaker of English as an L2, may pronounce the word /ɪŋgɒˈrɪʃiː/ due to the difficulty in the production of consonant blends for many Korean speakers, resulting in a word with three or more syllables. This phenomena of matching the 'expected' number of syllables to the actual syllables uttered is not exclusively a problem for L2 speakers either as L1 speakers are often able to pronounce the same word in a different way, resulting in a word which may be counted as containing a different number of syllables (De Jong and Perfetti, 2011). For example, 'our' can be pronounced in English as /aʊə/ or /ɑː/. While the issue of syllable recognition may not appear to be directly relevant to the automatic recording of pause locations, if PRAAT is unable to identify syllables correctly then it may erroneously label speech as silence resulting in the incorrect recording of pause length or type. Further empirical study would be required to explore the amount and frequency of any discrepancies present between manually and automatically measured syllable counts and pause locations.

While the automatic coding of syllables is not always completely accurate, and some variation in pronunciation is inevitable between participants, both L1 and L2 speakers, it was deemed to be the best choice for this study due to the increased consistency it provides across both sets of speakers. The manual counting of syllables was deemed to be too time consuming for this study and ultimately the gains in accuracy would be insufficient to justify the extra work and possible loss in consistency. As the focus of the present study is primarily on the location and frequency of pauses these were automatically coded at first, and then hand coded as described above. This was still a considerable undertaking, with each audio file containing anywhere between 6 and 25 pauses. However, this was far fewer than the number of syllables produced which were often approaching 200 in each one minute audio file. This decision resulted in combining the use of the automatic measures, to save time and improve consistency, and the hand coding, to check the accuracy of the location, type, and length of any pauses.

As previously explained all of the audio files were trimmed to be one minute to facilitate the comparison of oral fluency measures. However, this meant that a participant would often still be speaking and the time limit of one minute would expire in the middle of a speech run. In all instances where this occurred the analysis moved to the end of the run, to the next pause or breakdown. This was to more accurately compare pauses, and to classify them as either mid-clause or end-clause, and resulted in several files slightly exceeding the one minute, or sixty second, time limit. However, this was deemed preferable to interrupting the speech run and was viewed as being more reflective of a natural speech pattern (De Jong & Perfetti, 2011).

A further issue that was identified with the automatic coding of silence during the initial checking phase was the appearance of non-verbal phenomena such as laughing, coughing, and throat clearing. These sounds were incorrectly recognised by PRAAT as speech and the script identified them as syllables. To combat this problem all examples of these phenomena were removed from the text grids, and were not counted as either silent, or voiced pauses, nor were they recorded as speech. The file 'textgrid' file was manually annotated to task into account these sounds, and they were removed from any calculations performed by the scripts.

Another possible limitation in the area of data analysis is the possibility of an underspecified model arising from the forward selection procedure, (Appendix M) implemented in data analysis with RStudio. Briefly, underspecified models may arise due to the possibility of a better fitting model, which includes a combination of parameters which have not been tested by the model selection procedure (Murakami, 2016). However, this study employed a process whereby variables where carefully considered and tested before fitting them to the model. This enabled a greater number of variables, as well as greater combinations of variables to be inspected. For a more in depth and detailed discussion of model fitting and using mixed effects models in data analysis see Appendix S.

It should also be noted that even though the sample size of the study was quite large (n = 80) these were subdivided into smaller groups, by speaker type, task complexity, and task input format. This means that any analytical conclusions drawn would become less reliable. This may explain why some of the results were inconclusive, particularly concerning the impact of task input format. Furthermore, as was detailed in chapters 5 and 7, some outliers were present in the data which may have skewed the results to a small extent, but were left in due to the need to balance the design of the study and create comparable groups. Despite these factors however, I think it is reasonable to argue that the findings of this study are generalizable for similar populations of second language English learners, particularly those studying at an intermediate level in a Korean university setting.

It is possible that there are issues regarding the implementation of the PSTM memory test. The test in the current study was administered as a digit span, and the methodological choice was supported by previous research and justifications in the literature. However, it is possible that a nonword repetition test (NWRT), delivered in the participants' L1 (Korean), may have been a better way to gauge the participants levels of PSTM. NWRTs have been shown in many research studies (Suzuki, 2021; O'Brien, et al., 2007) to provide accurate levels of PSTM. The NWRT can be deployed in the participants L1 to mitigate the impact of a variation in language proficiency among participants. The number based input of the digit span test may also be confounded by issues related to the numeracy abilities among participants. The language provided in NWRTs removes any numeracy issues amongst participants, and is purely focused on the reptation of the nonwords, to allow for a clearer view of PSTM. However, one major problem with implementing NWRTs in this particular study is a lack of Korean ability. Tests for NWRTs have been created in many languages, and translations are available, but as I am unable to speak Korean well enough to judge the participants pronunciation and accurate recall, it was decided that a digit span test would have to suffice.

A further limitation which may be present in the design of the study, despite the best efforts to guard against it, is the inherent monolingual bias. The current research project, in seeking to explore the differences between L1 and L2 speech production, may have unintendedly positioned L2 speech as deficient, and held L1 speech as the ideal or target. The positioning of the L1 and the monolingual speaker as central to language research often fails to adequately take into account the experience of the specific learners, but also the language users who speak more than one language, which are of course all L2 speakers (Cook, 2002).

This issue is of particular relevance to this research project, with its genesis in the language teaching and learning of South Korea, which often comes with the stated goal of becoming a 'native like speaker' of English. Anecdotally, I have evidenced this positioning of L2 speech as inferior when teaching English in South Korea, with Korean English language students frequently saying they are 'bad' at English, even though they may have high levels of language proficiency, and be able to communicate across a wide range of subjects. Grosjean (1989) points out that L2 speakers will often esteem the native speaker, and as a result be critical of their own language competence in the target language. This notion of deficit is one which continues to persist throughout the English education system in South Korea. However, what the present study is seeking to do is not to point to L2 speech as lacking, or deficient in comparison to L1 speech, but rather to examine and understand the differences between speech production and to highlight them, whilst exploring the links between these differences and the stimulus provided by classroom tasks. If these differences can be better understood then hopefully this will lead to better teaching practices through the creation of more targeted and specific tasks, with more appropriate task design, and ultimately lead to better learning outcomes for second language learners.

8.11 Future Research

The findings of this research project have produced a number of areas for exploration in further research studies. Firstly, to what extent does the cognitive complexity of tasks impact the oral fluency of *beginner* level Korean learners of English? It would be of theoretical and pedagogical interest to learn if the impacts of task complexity measured at the mid and end-clause are different amongst less proficient learners. It could provide added insight into how speech production, and in particular automaticity and cognitive processes develop as learners become more proficient at producing speech in an L2. Correlations with levels of PSM would also be of great interest due to previous studies showing the importance of PSTM to the early stages of SLA.

Another area of interest for investigation would be the way in which task complexity is operationalised. This present study employed +/- information order to increase the cognitive load of the task, but differing operationalisations of task complexity may yield a different impact on oral fluency. In a similar way a variation in the manner in which task input formats were employed could result in a different response. It would be of interest to see if utilising video or audio for task input impacts the way in which learners produce speech, and if the variation in input results in a change in the cognitive load of the task. This would help to more clearly define which attributes of a task are impacting which elements of speech production, and provide some much needed exploration of the variation in task input and the ways in which it affects L2 oral fluency.

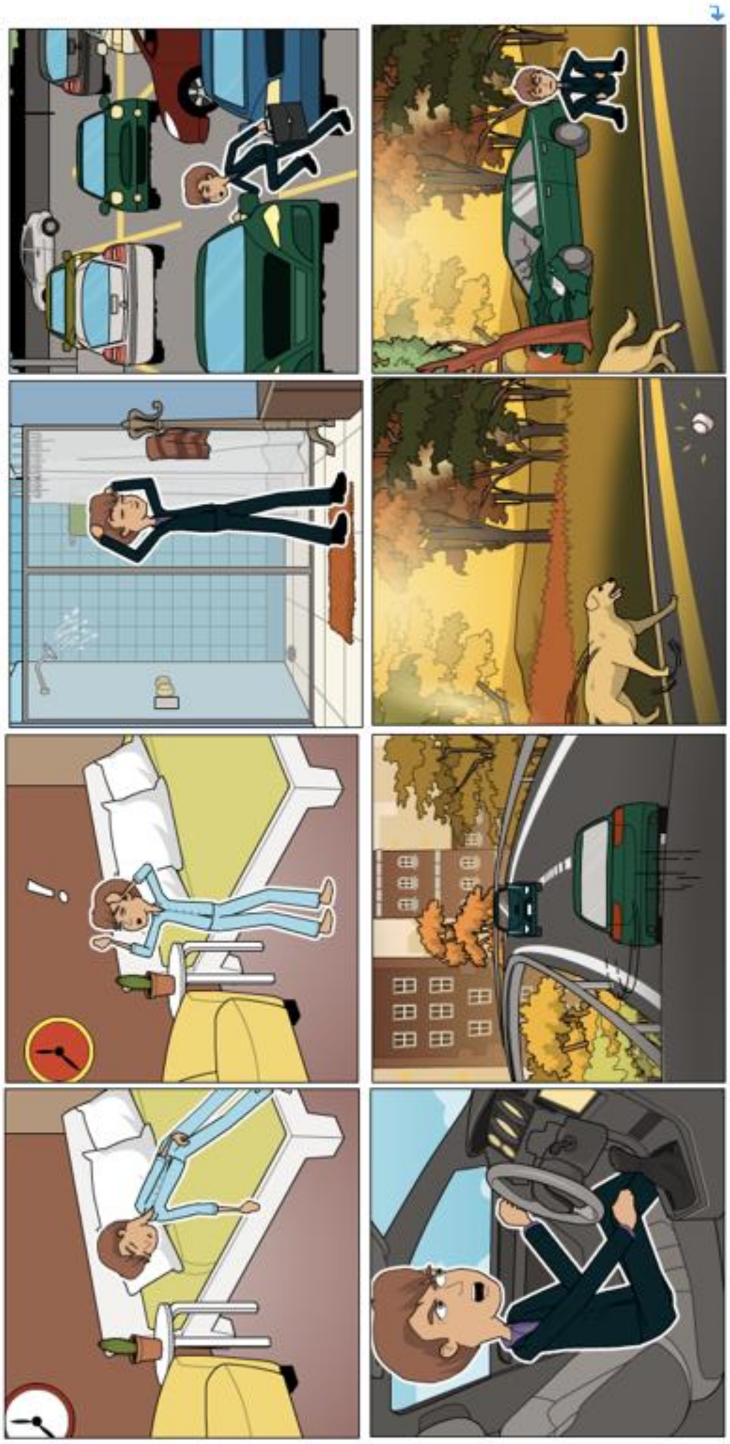
A further way in which future research could build on the findings of the present study would be to investigate the impacts of increased task cognitive complexity on oral fluency in an intact language learning classroom. As was mentioned previously in the limitations section, it is felt that while research in the classroom brings with it a wide range of challenges, not least in the accurate audio recording of speech, it does still offer the opportunity to gain more authentic first-hand knowledge of the way speech is actually produced as L2 speakers learn. A classroom based study would also offer clearer links to language teaching pedagogy, while allowing for the investigation of language interaction, in pair or group work tasks, rather than solely on narrative monologues as was the case for this study.

The final area of research which has been suggested by the findings of this study is concerned with PSTM. Further investigation into which aspects of speech production are impacted and how and why task complexity interacts with PSTM would benefit from further research. Echoing the suggestions made in the previous section it would be worth investigating the specific language produced after an end clause pause, and to see if the increased end clause pause frequencies correlate with decreased pausing at the mid clause position for individual learners.

Appendices

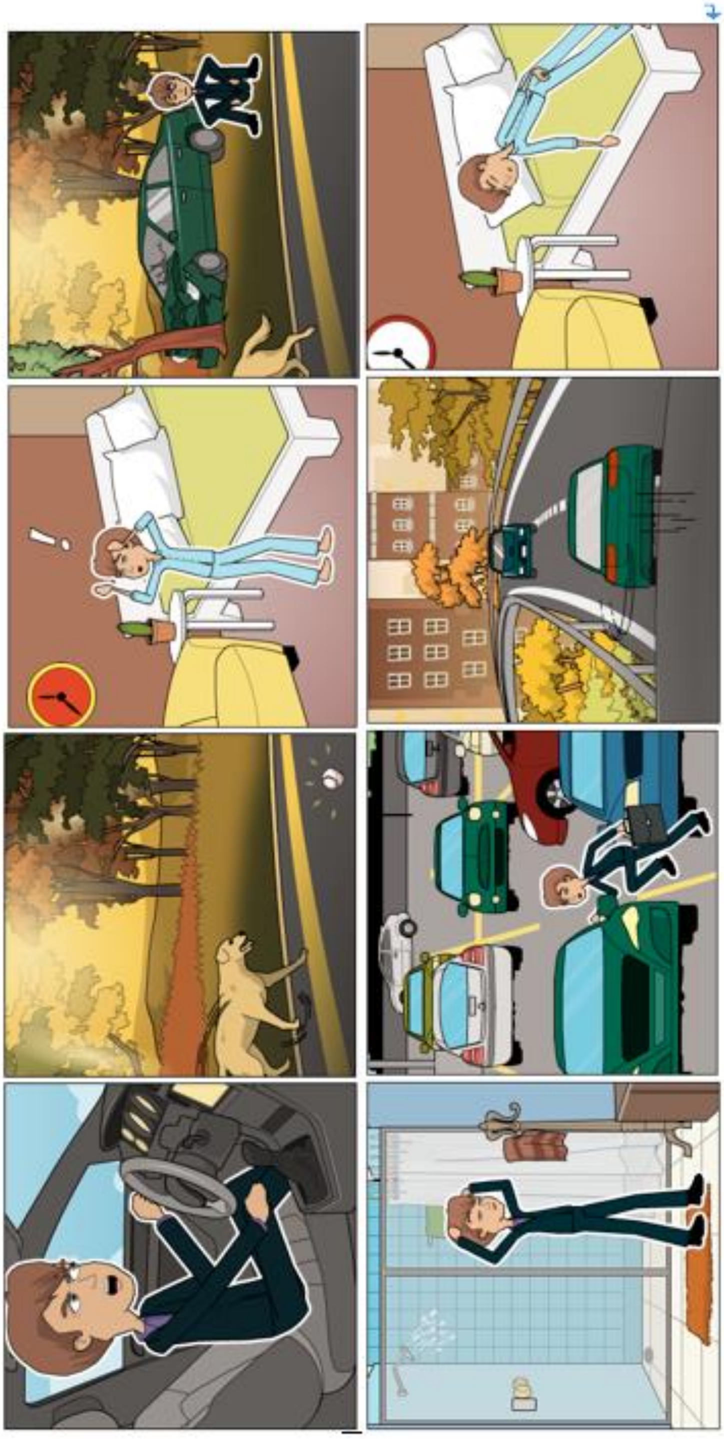
Appendix A – Pictorial Input Task Less Complex

Less complex version showing pictures in chronological order



Appendix B – Pictorial Input Task More Complex

More complex version showing pictures in fixed random order



Appendix C – Written Input Task Less Complex

Less complex version showing bullet points in chronological order

Jonny Depp 13 facts

1: Born, Kentucky, U.S.A., June 9, 1963.

2: Parents’ divorce 1978.

3: Quit School 1979.

4: Move to LA 1979.

5: Marriage Lori Anne Wilson 1983.

6: Divorce 1985.

7: *Edward Scissorhands* movie 1990.

8: *Donnie Brasco* movie released 1997.

9: Lilly-Rose Depp born, May 27, 1999.

10: *Pirates of the Caribbean* movie 2003.

11: Marriage Amber Heard 2015.

12: Divorce 2017.

13: Fifth *Pirates of the Caribbean* movie 2017.

Appendix D – Written Input Task More Complex

More complex version showing bullet points in fixed random order

Jonny Depp 13 facts

Pirates of the Caribbean movie 2003.

Marriage Amber Heard 2015.

Move to LA 1979.

Donnie Brasco movie released 1997.

Divorce 1985.

Born, Kentucky, U.S.A., June 9, 1963.

Edward Scissorhands movie 1990.

Parents' divorce 1978.

Lilly-Rose Depp born, May 27, 1999.

Divorce 2017.

Marriage Lori Anne Wilson 1983.

Fifth *Pirates of the Caribbean* movie 2017.

Quit School 1979.

Appendix E – L1 Participant Post Task Feedback

Participant Feedback

Thank you very much for taking part in this research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

Appendix F – Task Instructions

Instructions to read to students

Biography (Timeline) task cognitively *less* complex version.

Please look at the sentences in front of you.

For this task I want you to tell me about the life of Jonny Depp using the information provided.

Please talk about the major events of his life in chronological order, i.e.; in the order they happened to him, starting with his birth.

Please speak in full sentences and give as many details as you can, don't just read.

You may look at and use the sentences during the task to help you.

Try to talk for at least two minutes.

Please don't ask me any questions once you have begun, do your best, and don't worry about any mistakes that you may make.

You have two minutes to prepare and then we will begin.

Biography (Timeline) task cognitively *more* complex version.

Please look at the sentences in front of you.

For this task I want you to tell me all about the life of Jonny Depp using the information provided.

Please talk about the major events of his life in chronological order, i.e.; in the order they happened to him, starting with his birth.

Please speak in full sentences and give as many details as you can, don't just read.

You may look at and use the sentences during the task to help you.

Try to talk for at least two minutes.

Please don’t ask me any questions once you have begun, do your best, and don’t worry about any mistakes that you may make.

You have two minutes to prepare and then we will begin.

Pictures (Accident) task cognitively *less* complex version.

Please look at the pictures in front of you.

For this task I want you to tell me all about the events portrayed in the comic strip using the pictures provided.

Please talk about each picture in chronological order, i.e.; in the order they happened to him, starting with the earliest.

Please speak in full sentences and give as many details as you can, don’t just describe the picture.

You must make at least two sentences about each picture.

You may look at and use the pictures during the task to help you.

Please don’t ask me any questions once you have begun, do your best, and don’t worry about any mistakes that you may make.

You have two minutes to prepare and then we will begin.

Pictures (Accident) task cognitively *more* complex version.

Please look at the pictures in front of you.

For this task I want you to tell me all about the events portrayed in the comic strip using the pictures provided.

Please talk about each picture in chronological order, i.e.; in the order they happened to him, starting with the earliest.

Please speak in full sentences and give as many details as you can, don’t just describe the picture.

You must make at least two sentences about each picture.

You may look at and use the pictures during the task to help you.

Please don’t ask me any questions once you have begun, do your best, and don’t worry about any mistakes that you may make.

You have two minutes to prepare and then we will begin.

Appendix G – PSTM Digit Span Test

Number Repetition for Ages 5–16

Ages 5–16

Materials Needed

Record Form 1, pages 18–19

Record Form 2, page 22

Start Point

Item 1 for all ages

Repetitions

Repeat the directions if necessary. Do not repeat the items.

Discontinue Rule

Discontinue testing after 0 scores on both parts (*a* and *b*) of an item; however, administer both the Numbers Forwards and the Numbers Backwards sections of this subtest.

Objective

To evaluate the examinee's ability to repeat random number sequences of graduated length forwards and backwards. This task places a heavy demand upon attention and concentration and auditory/verbal working memory (Cohen, 1997).

Relationship to Curriculum and Classroom Activities

In general, learning can be viewed as the process of acquiring new information, and memory as the consolidation and retention of that acquired knowledge. Learning and memory are based on a foundation of attention. If an examinee has difficulty directing his or her attention to the task at hand or sustaining attention for an appropriate amount of time, learning cannot occur (Cohen, 1997). CELF–4[UK] provides two short tests that assess the ability to sustain and direct attention and/or concentration, processing speed, and working memory. These measures and the index score that they form can be used as an initial step in examining a student's memory skills. An examinee's performance on the memory measures and the Working Memory index score can be compared with an examinee's overall performance on CELF–4[UK] to gain information about an examinee's language skills in comparison to his or her memory abilities and as a springboard to consider further referral for memory evaluation.

Administration Directions

In this subtest, the examinee is asked to repeat digit sequences of increasing length. For the Numbers Forwards portion, the examinee repeats the numbers in the same order as you say them. For the Numbers Backwards portion, the examinee repeats the numbers in the reverse order of how you say them. Administer both portions of this subtest. Each item in this subtest has two parts (*a* and *b*). Administer both parts and record the examinee's responses before proceeding to the next item.

Numbers Forwards

Say, **I'm going to say some numbers. Listen carefully, and when I'm finished, you say them in the same order as I said them. For example, if I say 1, 2, you say 1, 2.** Read the digits at the rate of 1 per second from the Record Form, then ask the examinee to repeat the sequence. When reading the digits, let the pitch of your voice drop on the last digit of each sequence. Administer and record both parts of each item even if the examinee passes part *a* of each item. Continue testing until the student scores 0 on both parts of an item, and then proceed to the Numbers Backwards section. Record the examinee's response to each item VERBATIM in the space provided in the Record Form.

Numbers Backwards

Say, **Now I'm going to say some more numbers, but this time when I stop, I want you to say them *backwards*. For example, if I say 3, 4, you say** _______. [Pause for the examinee to respond.] If the examinee responds correctly (4, 3), say, **That's right,** and proceed to part *a* of Item 1. If the examinee responds incorrectly, say, **That's not quite right. I said 3, 4, so you would say 4, 3. Do you understand?** Repeat the example if necessary. This time, whether the examinee answers correctly or incorrectly, proceed to part *a* of Item 1. Do not help the examinee on any of the items. Read the digits at a rate of 1 per second from the Record Form. Administer and record both parts of each item (*a* and *b*), even if the examinee passes only part *a* of each item. Continue testing until the examinee scores 0 on both parts of an item. Record the student's response to each item VERBATIM in the space provided in the Record Form.

Scoring the Items

For each part (*a* and *b*) score 1 point for each correctly recalled sequence or 0 points for each incorrectly recalled sequence. Sum the scores of Numbers Forwards Items 1–8. Sum the scores of Numbers Backwards Items 1–7. To get a Total Raw Score, sum the Numbers Forwards Raw Score and the Numbers Backwards Raw Score. Record the subtest raw scores in the Raw Score boxes at the end of the Number Repetition section in the Record Form

Number Repetition (NR, Ages 5–16) *continued*

Forwards

Item			Response	Score
1.	a.	3–5	3 5	1 0
	b.	7–2	7 2	1 0
2.	a.	2–8–6	2 8 6	1 0
	b.	6–3–4	6 3 4	1 0
3.	a.	6–2–5–8	6 5 2 8	1 0
	b.	2–4–1–7	4 1 7	1 0
4.	a.	9–5–1–4–8		1 0
	b.	5–8–2–1–6		1 0
5.	a.	4–7–8–1–6–3		1 0
	b.	7–3–9–8–6–4		1 0
6.	a.	6–1–7–4–2–3–8		1 0
	b.	9–3–8–6–5–1–2		1 0
7.	a.	5–3–8–7–2–1–6–4		1 0
	b.	2–4–9–5–7–1–6–3		1 0
8.	a.	1–6–4–5–9–7–2–8–3		1 0
	b.	4–5–2–3–6–8–9–7–1		1 0
			Forwards Raw Score	4

Backwards

Item			Correct Response	Response	Score
1.	a.	3–8	(8–3)	3 8	1 0
	b.	7–4	(4–7)	7 4	1 0
2.	a.	4–8–3	(3–8–4)		1 0
	b.	3–6–8	(8–6–3)		1 0
3.	a.	5–2–9–6	(6–9–2–5)		1 0
	b.	8–3–4–9	(9–4–3–8)		1 0
4.	a.	4–7–1–5–3	(3–5–1–7–4)		1 0
	b.	9–2–7–5–8	(8–5–7–2–9)		1 0
5.	a.	1–8–6–9–5–2	(2–5–9–6–8–1)		1 0
	b.	3–4–6–9–7–1	(1–7–9–6–4–3)		1 0
6.	a.	8–2–5–4–9–3–2	(2–3–9–4–5–2–8)		1 0
	b.	4–1–5–8–7–2–9	(9–2–7–8–5–1–4)		1 0
7.	a.	6–8–9–5–1–2–6–3	(3–6–2–1–5–9–8–6)		1 0
	b.	3–2–1–8–7–5–9–4	(4–9–5–7–8–1–2–3)		1 0
				Backwards Raw Score	0
				NR–Total Raw Score	4

Appendix H – L1 Participant Post Task Questionnaire Response Examples

Order of information increased difficulty

1 /

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

Not being in order made it tricky to get the whole picture, ~~wh~~ which made it harder to interpret the images

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

The events were in order and reading from text is very easy

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

Pretty good at interpreting pictures and what they mean

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

Kept reading same bits of info twice

Was difficult to put into chronology in head.

Appendix I – L1 Participant Post Task Questionnaire Response Examples

Order of information eased difficulty

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

The pictures were clear to understand and it definitely helped being in order.

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

The ordering made it very difficult to follow a logical timeline, would be helpful to have a pen to mark the order or something

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

Pictures were easy to interpret, chronologically made sense, which made the description easier

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

Not in order; tended to miss out events and backtrack.

Appendix J – L1 Participant Post Task Questionnaire Response Examples

Task input format eased difficult

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

hard version

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

It was easy to understand the story and explain what I saw.
Slightly more difficult to work out all the details in each picture

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

easy version

4: Why do you think it was difficult/easy? Please give details below

The iformation was in bullet paints which made it easy to read.

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

easy version

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

Trying to look at all the pictures at the same time to make the story but easier to look at pictures than words

3: How difficult did you find the Depp timeline task? (Circle one)

Hard Version

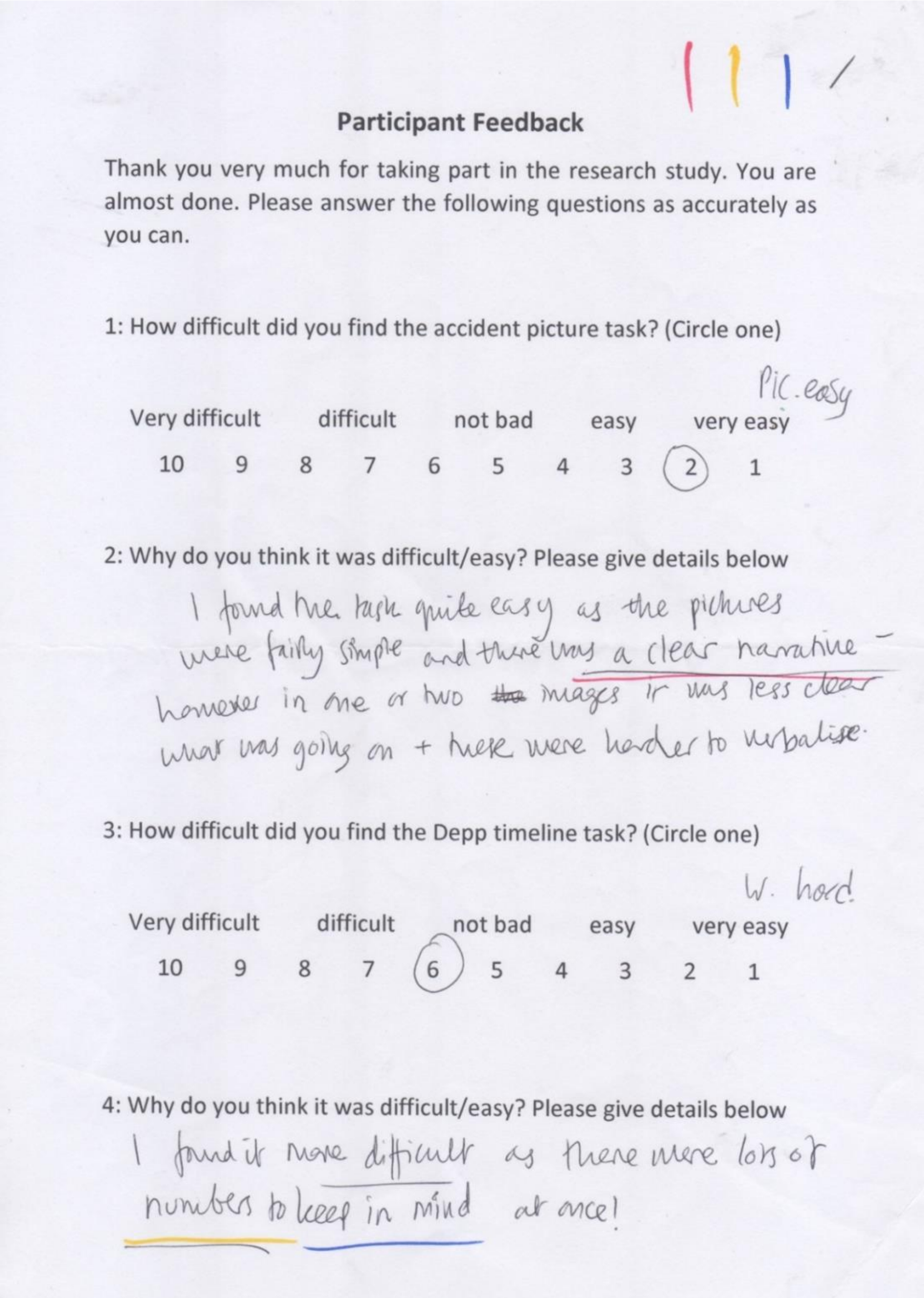
Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

~~Ha~~ A lot harder to look down a list of sentences and order dates, seemed to be in different formats as well. some italic for example.

Appendix K – L1 Participant Post Task Questionnaire Response Examples

Task input format increased difficulty



Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Pic. easy

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

I found the task quite easy as the pictures were fairly simple and there was a clear narrative – however in one or two ~~the~~ images it was less clear what was going on + these were harder to verbalise.

3: How difficult did you find the Depp timeline task? (Circle one)

W. hard

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

I found it more difficult as there were lots of numbers to keep in mind at once!

Appendix L – L1 Participant Post Task Questionnaire Response Examples

Task difficulty in general

Participant Feedback

Thank you very much for taking part in the research study. You are almost done. Please answer the following questions as accurately as you can.

1: How difficult did you find the accident picture task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

2: Why do you think it was difficult/easy? Please give details below

It was quite difficult as it required thinking on my feet. PH

3: How difficult did you find the Depp timeline task? (Circle one)

Very difficult		difficult		not bad		easy		very easy	
10	9	8	7	6	5	4	3	2	1

4: Why do you think it was difficult/easy? Please give details below

It was easier as you could repeat the sentance to get thinking V.E.

Appendix M - Analysis: Statistical Model Creation and Justification

Pause location was first analysed in R (R Core Team, 2018) using mixed-effects linear regression (Baayen, Davidson, & Bates, 2008) using the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core team, 2018). The three categorical independent variables were sum coded to allow for better interpretation analysis (Field, et al., 2012; Winter, 2019). Sum coding involves converting a categorical predictor into sum-codes, for example with the current data set for speaker type the L1 category was assigned the value of -1 and the L2 speaker category was assigned the value of +1. The benefit of employing this coding scheme is that the categorical predictor becomes 'centred', and the intercept is now halfway in between, or in the middle of the two categories (Winter, 2019). As the data collected is balanced, with an exactly equal number of participants (n = 40) L1 speakers and (n = 40) L2 speakers, the intercept is also the overall mean of the dataset. This means that the coefficient table produced by statistical analysis in R now lists the 'main effects' of the interaction rather than the 'simple effects' (Field et al., 2012).

As each participant delivered two responses, one for each task complexity type, dependencies are introduced (Field et al., 2012). The data collected is not independent and therefore a mixed model is appropriate to analyse and perform analysis (Winter, 2019; Field et al., 2012). Mixed models avoid violating the assumption of independence and instead model relationships between observations, unlike analyses of variance (Field et al., 2012). Mixed effects models also confer several additional advantages when compared to analyses of variance as they allow for increased robustness against violating sphericity and homoscedasticity, as well as incorporating the flexibility required of binomial variables in the distribution of data (Marsden & McManus, 2018).

So far to date the literature on statistical analysis has failed to reach a consensus on how best to perform model selection in mixed effects modelling, either within the linguistics community or amongst the wider statistical research community (Gries, 2016). There has been a tendency amongst linguists to fit 'overly complex models' (Winter, 2019) with unnecessarily complex random effects structures, largely based on the advice of Barr et al. (2013) who cautioned against ignoring important factors. The Barr et al., (2013) article led to many researchers within linguistics incorporating random and varying slopes for every critical variable within a study and 'over-fitting' statistical models (Winter, 2019). More recent research has pointed out that a more balanced and nuanced approach to model fitting is advisable and the "keep it maximal" credo of Barr et al., (2013) often results in difficulties with statistical estimations (Matuschek, et al., 2017), and may also reduce the statistical power of the model (Seedorff, Olseon, & McMurray, 2019).

Following research conducted by Marsden and McManus (2018) and following the suggestions of Field et al. (2012) and Murakami (2016) multiple models were constructed in a stepwise 'approach' to find the most plausible through comparison. The models were constructed in order of complexity, beginning with the baseline model containing only the simplest elements and then adding additional parameters one by one (Murakami, 2016; Field et al., 2012). The individual models were then compared as they were constructed using the AIC, in combination with the *anova* function, with the additional elements only retained if they improved the model by recording a significant decrease in the AIC value (Winter, 2019; Field et al., 2012).

Mid pause frequency Model Construction and Comparisons

The first model created was a baseline model to explore the main effects of task complexity and speaker type on mid-clause pause frequency, which included no predictors, only the intercept (Murakami, 2016). As the first research question is whether task complexity impacts L1 and L2 speaker oral fluency measured at the mid-clause level, the initial model was constructed with a formula exploring mid pause frequency as a function of task complexity. The model created was named MidFq1 with the code used in R displayed below.

```
MidFq1 <- lme (mid_pf ~ 1, random = ~ 1 | speaker_id, data = all_data, method = "ML")
```

The model is specified as predicting the outcome only from the intercept mid pause frequency (mid_pf ~ 1), and is using the maximum likelihood to estimate the model (method = "ML"). The random part of the model includes only the variable of speaker id, a label given for each individual participant. This enables the model to compensate for any individual differences among speech patterns and allows for the fact that some participants may pause more than others in general, and not solely as a result of the impacts of task complexity (Field et al., 2012).

To see the impact of each main effect and/or random effect and the interaction they may have they were individually added to the model one by one in a 'step-wise' fashion (Field et al., 2012; Murakami, 2016; Marsden & McManus, 2018). The models were then compared against each other using the *anova* function in R, with the resultant decrease in the Akaike Information Criterion (AIC) value taken as confirmation of a better fit and the associated *p* value confirming if the improvement in fit was statistically significant. The table below shows a summary of model comparisons and shows the results displaying the AIC of each model, as well as the difference in the AIC score. A lower value for the AIC and a negative value in the difference of AIC expresses higher predictive accuracy for the model which precedes it (Field et al., 2012; Murakami, 2016; Marsden & McManus, 2018). Table A-1 below also includes the likelihood ratio test results in comparison to the model directly preceding it.

Table A-1. *Summary of generalised linear mixed-effects model comparisons for mid-clause pause frequency*

Model	Fixed-effects	Random-effects	AIC	Δ AIC	-2LL statistic	p
MidFqMod 1	None	By-participant	-410.67			
MidFqMod 2	None	MidFqMod 1 + by-complexity	-408.67	2	$\chi2(1) = 0.00$	1
MidFqMod 3	None	MidFqMod 2 + by-format	-406.67	2	$\chi2(1) = 0.00$	1
MidFqMod 4	MidFqMod 3 + Task complexity	Same as MidFqMod 1	-423.16	-12.5	$\chi2(1) = 14.49$	<.001
MidFqMod 5	MidFqMod 4 + Speaker type	Same as MidFqMod 1	-488.50	-65.34	$\chi2(1) = 67.34$	<.001

MidFqMod 6	**MidFqMod 5 + Task format**	**Same as MidFqMod 1**	**-488.26**	**0.24**	**χ2(1) = 1.75**	**0.19**
MidFqMod 7	**MidFqMod 5 + Speaker type x Task complexity interaction**	**Same as MidFqMod 1**	**-489.38**	**-1.12**	**χ2(1) = 3.12**	**0.08**
MidFqMod 8	**MidFqMod 5 + Speaker type x Task format interaction**	**Same as MidFqMod 1**	**-485.38**	**2**	**χ2(1) = 0.00**	**0.97**
MidFqMod 9	**MidFqMod 5 + Task format x Task complexity interaction**	**Same as MidFqMod 1**	**-483.38**	**2**	**χ2(1) = 0.00**	**0.95**
MidFqMod 10	**MidFqMod 5 + Speaker type x Task complexity x Task format interaction**	**Same as MidFqMod 1**	**-483.71**	**1.67**	**χ2(1) = 0.33**	**0.57**

The first model compared with the baseline model was MidFq2 which added the random factor of task complexity as a repeated measures predictor to the original baseline model. Task complexity was 'nested' within the previous variable of speaker id (participant) as each participant performed both levels of task complexity (Field et al., 2012). A comparison of these two models shows the fit of the model is actually decreased, as can be seen in the table above with a slight increase in the AIC value. MidFq2 shows that adding task complexity as a nested random factor within the participant variable did not have any effect on mid pause frequency, $\chi2(1) = 0.00$, p =1.

The next model to be compared with the previous one was MidFq3 which again added a nested random factor, this time for the format of the task. Again, the variable was nested within participant as each participant completed both kinds of task format (Field et al., 2012) and again the fit of the model was decreased as a comparison of the models displays an increase in the AIC value. This model shows that adding the nested random variable of task format did not improve the model, $\chi2(1) = 0.00$, $p = 1$.

The next stage of the model fitting process moved from adding random factors to adding fixed effects factors. The first fixed effect to be added was that of task complexity and was added to MidFq1, as both the previous models had failed to increase the fit these additional elements were not included in the construction of the model and only the by-participant factor was included. Adding the main effect of task complexity can be seen to significantly improve the model with a reduction in the AIC, which was statistically significant $\chi2(1) = 14.49$, $p < .0001$. This model shows that the complexity of the task did have a significant impact on the frequency of mid-clause pauses.

MidFqMod 5 was then created to assess the addition of speaker type as a main effect, and shows the impact in comparison to the previous model that contained only the fixed effect of task complexity. The table shows the fit of the model is significantly improved with a considerable reduction in the AIC value. This confirms that the type of speaker did have a significant effect on mid-clause pause frequency, $\chi2(1) = 67.34$, $p < .0001$. It is important to note that this model contains the predictors for both speaker type *and* task complexity, in comparison to the previous model which contained *only* task complexity as a main effect. The step wise production of the model includes all of the elements contained within the previous model with the extra additional elements.

The next model constructed, MidFqMod 6, was created with the further addition of the main effect predictor of task format, and shows that including task format in addition to task complexity and speaker type did not improve the fit of the model. The AIC value actually increases very slightly with the addition of task format as a fixed effect, indicating that task format did not have a significant overall effect on the frequency of mid-clause pauses $\chi2(1) = 1.752$, $p = .185$.

The next group of models added were the two-way interactions between pairs of variables. These models were constructed adding to MidFqMod 6, containing all of the fixed main effect factors mentioned above. Again, these interactions were added one at a time with the addition of task complexity x speaker type interaction to create MidFqMod 7. The table summarises the results of the comparison with the previous model, MidFqMod 6, and shows that there is a slight improvement in the fit of the model, but it is not a significant improvement on the previous model, $\chi 2(1) = 3.12$, p =.077. The results of the interaction term mean that although mid pause frequency was affected by whether or not the task was more or less complex, the way in which mid pause frequency was affected by complexity was different in L1 and L2 speakers, though importantly this difference was not shown to be statistically significant.

The next model was MidFqMod 8 and incorporated the speaker type x task format interaction into the model. The summarised results show that the speaker type × task format interaction decreased the fit of the model with an increase in the AIC value of 2, and was not significant, $\chi 2(1) = 0.001$, p = .97. This indicates that although the speaker type did have an impact on the frequency of mid-clause pauses, this did not differ across the different task formats.

MidFqMod 9 was next constructed and added the task complexity x task format interaction to the model. The results show that the task complexity × task format interaction was also not significant, $\chi 2(1) = 0.003$, p = .95. This none significant interaction means that although mid pause frequency was impacted by whether the task was more or less complex, the way in which pauses were affected by complexity was not different in a comparison of the written and picture format tasks.

The last model constructed was MidFqMod 10, and includes all of the main effects and interactions from previous models whilst also accounting for the addition of the three-way interaction of speaker type x task complexity x task format. The results are summarised in the table and show that the speaker type × task complexity x task format interaction does not have a significant impact on the frequency of mid-clause pauses, $\chi 2(1) = 00.33$, p = .57. This means that the task complexity × speaker type interaction was not significantly different across the different task formats.

The same step wise approach in model building was incorporated in the analysis of pauses frequency at the end of clauses, with each consecutive model tested to see if it was an improvement on the previous model and changes added only if they provided a significant increase in fit (Appendix N).

Appendix N – Generalised Linear Mixed-Effects Model End-Clause

Summary of generalised linear mixed-effects model comparisons for end clause pause frequency

Model	Fixed-effects	Random-effects	AIC	Δ AIC	-2LL statistic	p
EndFqMod 1	**None**	**By-participant**	**-391.62**			
EndFqMod 2	**None**	**EndFqMod 1 + by-complexity**	**-389.62**	**2**	**$\chi2(1) = 0.00$**	**1**
EndFqMod 3	**None**	**EndFqMod 2 + by-format**	**-387.62**	**2**	**$\chi2(1) = 0.00$**	**1**
EndFqMod 4	**EndFqMod 3 + Task complexity**	**Same as EndFqMod 1**	**-399.23**	**-11.61**	**$\chi2(1) = 9.61$**	**<.001**
EndFqMod 5	**EndFqMod 4 + Speaker type**	**Same as EndFqMod 1**	**-427.98**	**-28.75**	**$\chi2(1) = 30.75$**	**<.001**
EndFqMod 6	**EndFqMod 5 + Task format**	**Same as EndFqMod 1**	**-448.25**	**-20.27**	**$\chi2(1) = 22.27$**	**<.001**
EndFqMod 7	**EndFqMod 6 + Speaker type x Task complexity interaction**	**Same as EndFqMod 1**	**-446.40**	**1.85**	**$\chi2(1) = 0.15$**	**0.70**
EndFqMod 8	**EndFqMod 7 + Speaker type x Task format interaction**	**Same as EndFqMod 1**	**-452.41**	**6.01**	**$\chi2(1) = 8.02$**	**<.005**
EndFqMod 9	**EndFqMod 8 + Task format x Task complexity interaction**	**Same as EndFqMod 1**	**-450.29**	**2.12**	**$\chi2(1) = 0.48$**	**0.49**

EndFqMod 10	**EndFqMod 9 + Speaker type x Task complexity x Task format interaction**	**Same as EndFqMod 1**	**-452.29**	**2**	**χ2(1) = 3.40**	**0.07**

Appendix O - De Jong and Wempe (2009) PRAAT Script used for automatic coding of silence

```
###########################################################################
#                                                                         #
#  Praat Script Syllable Nuclei                                           #
#  Copyright (C) 2008  Nivja de Jong and Ton Wempe                        #
#                                                                         #
#    This program is free software: you can redistribute it and/or modify #
#    it under the terms of the GNU General Public License as published by #
#    the Free Software Foundation, either version 3 of the License, or    #
#    (at your option) any later version.                                  #
#                                                                         #
#    This program is distributed in the hope that it will be useful,      #
#    but WITHOUT ANY WARRANTY; without even the implied warranty of       #
#    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the        #
#    GNU General Public License for more details.                         #
#                                                                         #
#    You should have received a copy of the GNU General Public License    #
#    along with this program.  If not, see http://www.gnu.org/licenses/   #
#                                                                         #
###########################################################################
#
# modified 2010.09.17 by Hugo Quené, Ingrid Persoon, & Nivja de Jong
# Overview of changes:
# + change threshold-calculator: rather than using median, use the almost maximum
#     minus 25dB. (25 dB is in line with the standard setting to detect silence
#     in the "To TextGrid (silences)" function.
#     Almost maximum (.99 quantile) is used rather than maximum to avoid using
#     irrelevant non-speech sound-bursts.
# + add silence-information to calculate articulation rate and ASD (average syllable
#     duration.
#     NB: speech rate = number of syllables / total time
#         articulation rate = number of syllables / phonation time
# + remove max number of syllable nuclei
# + refer to objects by unique identifier, not by name
# + keep track of all created intermediate objects, select these explicitly,
#     then Remove
# + provide summary output in Info window
# + do not save TextGrid-file but leave it in Object-window for inspection
#     (if requested in startup-form)
# + allow Sound to have starting time different from zero
#      for Sound objects created with Extract (preserve times)
# + programming of checking loop for mindip adjusted
#      in the orig version, precedingtime was not modified if the peak was rejected !!
#      var precedingtime and precedingint renamed to currenttime and currentint
#
# + bug fixed concerning summing total pause, feb 28th 2011
###########################################################################


# counts syllables of all sound utterances in a directory
# NB unstressed syllables are sometimes overlooked
# NB filter sounds that are quite noisy beforehand
# NB use Silence threshold (dB) = -25 (or -20?)
# NB use Minimum dip between peaks (dB) = between 2-4 (you can first try;
#                                          For clean and filtered: 4)


form Counting Syllables in Sound Utterances
   real Silence_threshold_(dB) -20
   real Minimum_dip_between_peaks_(dB) 4
   real Minimum_pause_duration_(s) 0.25
   boolean Keep_Soundfiles_and_Textgrids yes
   sentence directory \directory
endform
```

```
# shorten variables
silencedb = 'silence_threshold'
mindip = 'minimum_dip_between_peaks'
showtext = 'keep_Soundfiles_and_Textgrids'
minpause = 'minimum_pause_duration'

# print a single header line with column names and units
printline soundname, nsyll, npause, dur (s), phonationtime (s), speechrate (nsyll/dur), articulation rate (nsyll / phonationtime), ASD (speakingtime/nsyll)

# read files
Create Strings as file list... list 'directory$'/*.wav
numberOfFiles = Get number of strings
for ifile to numberOfFiles
   select Strings list
   fileName$ = Get string... ifile
   Read from file... 'directory$'/'fileName$'

# use object ID
   soundname$ = selected$("Sound")
   soundid = selected("Sound")

   originaldur = Get total duration
   # allow non-zero starting time
   bt = Get starting time

   # Use intensity to get threshold
   To Intensity... 50 0 yes
   intid = selected("Intensity")
   start = Get time from frame number... 1
   nframes = Get number of frames
   end = Get time from frame number... 'nframes'

   # estimate noise floor
   minint = Get minimum... 0 0 Parabolic
   # estimate noise max
   maxint = Get maximum... 0 0 Parabolic
   #get .99 quantile to get maximum (without influence of non-speech sound bursts)
   max99int = Get quantile... 0 0 0.99

   # estimate Intensity threshold
   threshold = max99int + silencedb
   threshold2 = maxint - max99int
   threshold3 = silencedb - threshold2
   if threshold < minint
      threshold = minint
   endif

  # get pauses (silences) and speakingtime
   To TextGrid (silences)... threshold3 minpause 0.1 silent sounding
   textgridid = selected("TextGrid")
   silencetierid = Extract tier... 1
   silencetableid = Down to TableOfReal... sounding
   nsounding = Get number of rows
   npauses = 'nsounding'
   speakingtot = 0
   for ipause from 1 to npauses
      beginsound = Get value... 'ipause' 1
      endsound = Get value... 'ipause' 2
      speakingdur = 'endsound' - 'beginsound'
      speakingtot = 'speakingdur' + 'speakingtot'
   endfor

   select 'intid'
   Down to Matrix
   matid = selected("Matrix")
   # Convert intensity to sound
   To Sound (slice)... 1
```

```
sndintid = selected("Sound")

# use total duration, not end time, to find out duration of intdur
# in order to allow nonzero starting times.
intdur = Get total duration
intmax = Get maximum... 0 0 Parabolic

# estimate peak positions (all peaks)
To PointProcess (extrema)... Left yes no Sinc70
ppid = selected("PointProcess")

numpeaks = Get number of points

# fill array with time points
for i from 1 to numpeaks
   t'i' = Get time from index... 'i'
endfor


# fill array with intensity values
select 'sndintid'
peakcount = 0
for i from 1 to numpeaks
    value = Get value at time... t'i' Cubic
    if value > threshold
          peakcount += 1
          int'peakcount' = value
          timepeaks'peakcount' = t'i'
    endif
endfor


# fill array with valid peaks: only intensity values if preceding
# dip in intensity is greater than mindip
select 'intid'
validpeakcount = 0
currenttime = timepeaks1
currentint = int1

for p to peakcount-1
   following = p + 1
   followingtime = timepeaks'following'
   dip = Get minimum... 'currenttime' 'followingtime' None
   diffint = abs(currentint - dip)

   if diffint > mindip
      validpeakcount += 1
      validtime'validpeakcount' = timepeaks'p'
   endif
      currenttime = timepeaks'following'
      currentint = Get value at time... timepeaks'following' Cubic
endfor


# Look for only voiced parts
select 'soundid'
To Pitch (ac)... 0.02 30 4 no 0.03 0.25 0.01 0.35 0.25 450
# keep track of id of Pitch
pitchid = selected("Pitch")

voicedcount = 0
for i from 1 to validpeakcount
   querytime = validtime'i'

   select 'textgridid'
   whichinterval = Get interval at time... 1 'querytime'
   whichlabel$ = Get label of interval... 1 'whichinterval'

   select 'pitchid'
```

```
    value = Get value at time... 'querytime' Hertz Linear

    if value <> undefined
      if whichlabel$ = "sounding"
          voicedcount = voicedcount + 1
          voicedpeak'voicedcount' = validtime'i'
      endif
    endif
  endfor


  # calculate time correction due to shift in time for Sound object versus
  # intensity object
  timecorrection = originaldur/intdur

  # Insert voiced peaks in TextGrid
  if showtext > 0
    select 'textgridid'
    Insert point tier... 1 syllables

    for i from 1 to voicedcount
        position = voicedpeak'i' * timecorrection
        Insert point... 1 position 'i'
    endfor
  endif

  # write textgrid to file
#select textgridid
Write to text file... 'directory$'/'soundname$'_250.TextGrid

  # clean up before next sound file is opened
   select 'intid'
   plus 'matid'
   plus 'sndintid'
   plus 'ppid'
   plus 'pitchid'
   plus 'silencetierid'
   plus 'silencetableid'

   Remove
   if showtext < 1
     select 'soundid'
     plus 'textgridid'
     Remove
   endif

# summarize results in Info window
  speakingrate = 'voicedcount'/'originaldur'
  articulationrate = 'voicedcount'/'speakingtot'
  npause = 'npauses'-1
  asd = 'speakingtot'/'voicedcount'

  printline 'soundname$', 'voicedcount', 'npause', 'originaldur:2', 'speakingtot:2', 'speakingrate:2', 'articulationrate:2', 'asd:3'


                endfor
```

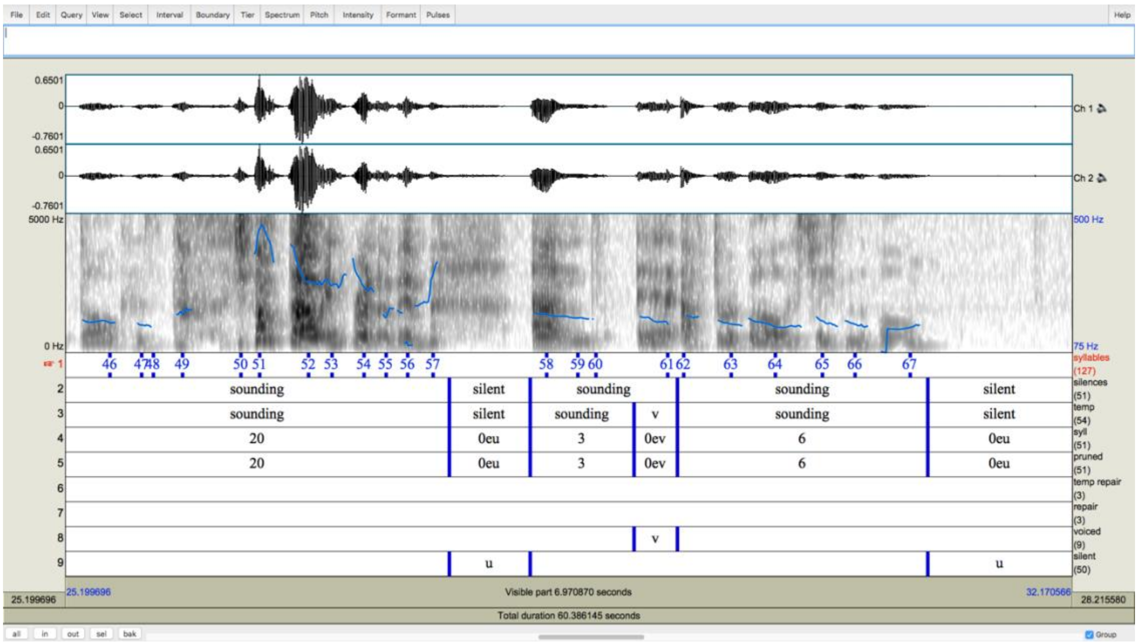
Appendix P - Handley and Haiping (2018) PRAAT Script used for automatic calculation of pause frequencies

```
# Helpful hints
# use GUI to find hints to what the commands might be, then Google Praat and the command
# remember variable names must start with a lower case letter

# create form
form Extracting Syllables tier
   sentence directory \directory
endform

# create header for output file
# note rates are all per minute
printline x, speed fluency, x, x, x, x, x, breakdown fluency, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x, x
printline x, working, x, x, x, measures, x, working, x, x, x, x, x, x, x, x, x, x, x, measures, x, x, x, x, x, x, x, x, x, x, x
printline x, x, x, x, x, x, x, counts, x, x, x, x, x, total duration, x, x, x, x, x, pause rate, x, x, x, x, x, average duration, x, x, x, x, x,
printline filename, n. syllables, duration, silence, phonation time, speech rate, articulation rate, unfilled end, filled end, complex end, unfilled mid, filled mid, complex mid, unfilled end, filled end, complex end, unfilled mid, filled mid, complex mid, unfilled end, filled end, complex end, unfilled mid, filled mid, complex mid, unfilled end, filled end, complex end, unfilled mid, filled mid, complex mid

# read files
Create Strings as file list... list 'directory$'/*.TextGrid
numberOfFiles = Get number of strings
for ifile to numberOfFiles
   select Strings list
   fileName$ = Get string... ifile
   Read from file... 'directory$'/'fileName$'

# select the textgrid
    textgridid = selected("TextGrid")

# initiate variables
# silence variable initiation
   silentTime = 0
   unfilledendTot = 0
   filledendTot = 0
   complexendTot = 0
   unfilledmidTot = 0
   filledmidTot = 0
   complexmidTot = 0

# syllable count initiation
   nSyll = 0

# extract sample duration
    endTime = Get end time

#calculate silent time
   silentNumberofIntervals = Get number of intervals: 9
   for silentIntervalNumber from 1 to silentNumberofIntervals
   nUnfilledInterval$ = Get label of interval: 9, silentIntervalNumber
      if startsWith(nUnfilledInterval$, "u")
         unfilledStart = Get start point: 9, silentIntervalNumber
         unfilledEnd = Get end point: 9, silentIntervalNumber
         unfilledDuration = unfilledEnd - unfilledStart
         silentTime = 'unfilledDuration' + 'silentTime'
       endif
     endfor

#calculate phonation time
    phonationTime = 'endTime' - 'silentTime'

#calculate total duration of different types of silence
   syllNumberofIntervals = Get number of intervals: 4
   for syllIntervalNumber from 1 to syllNumberofIntervals
   nSyllInterval$ = Get label of interval: 4, syllIntervalNumber
      if startsWith(nSyllInterval$, "0eu")
```

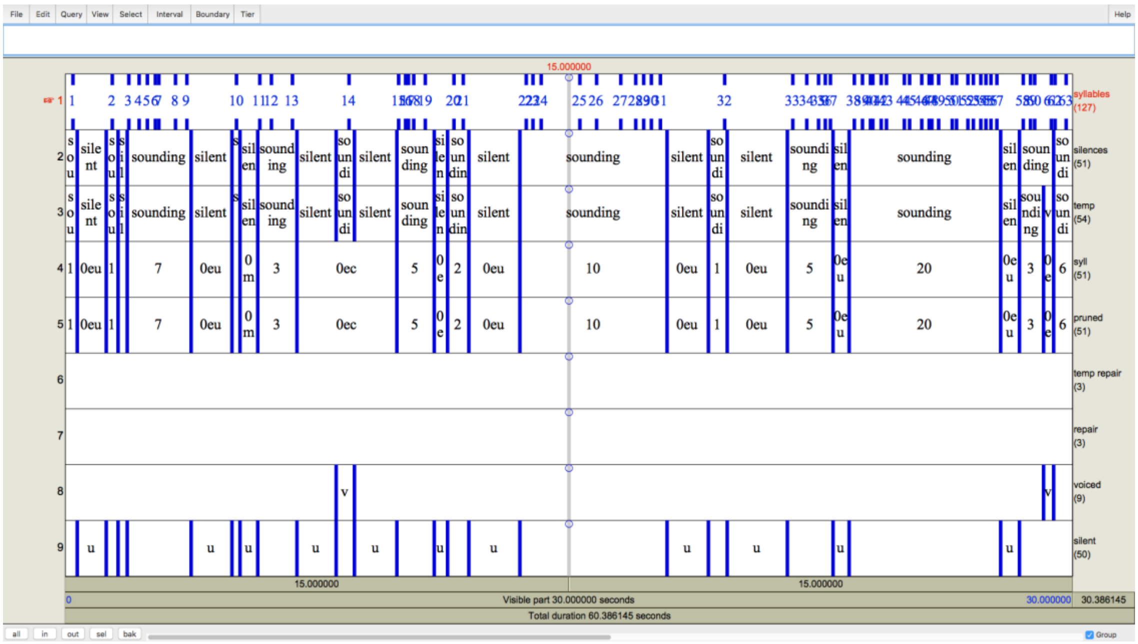
```
        unfilledendStart = Get start point: 4, syllIntervalNumber
        unfilledendEnd = Get end point: 4, syllIntervalNumber
        unfilledendDuration = 'unfilledendEnd' - 'unfilledendStart'
        unfilledendTot = 'unfilledendDuration' + 'unfilledendTot'
    elsif startsWith(nSyllInterval$, "0ev")
        filledendStart = Get start point: 4, syllIntervalNumber
        filledendEnd = Get end point: 4, syllIntervalNumber
        filledendDuration = 'filledendEnd' - 'filledendStart'
        filledendTot = 'filledendDuration' + 'filledendTot'
    elsif startsWith(nSyllInterval$, "0ec")
        complexendStart = Get start point: 4, syllIntervalNumber
        complexendEnd = Get end point: 4, syllIntervalNumber
        complexendDuration = 'complexendEnd' - 'complexendStart'
        complexendTot = 'complexendDuration' + 'complexendTot'
    elsif startsWith(nSyllInterval$, "0mu")
        unfilledmidStart = Get start point: 4, syllIntervalNumber
        unfilledmidEnd = Get end point: 4, syllIntervalNumber
        unfilledmidDuration = 'unfilledmidEnd' - 'unfilledmidStart'
        unfilledmidTot = 'unfilledmidDuration' + 'unfilledmidTot'
    elsif startsWith(nSyllInterval$, "0mv")
        filledmidStart = Get start point: 4, syllIntervalNumber
        filledmidEnd = Get end point: 4, syllIntervalNumber
        filledmidDuration = 'filledmidEnd' - 'filledmidStart'
        filledmidTot = 'filledmidDuration' + 'filledmidTot'
    elsif startsWith(nSyllInterval$, "0mc")
        complexmidStart = Get start point: 4, syllIntervalNumber
        complexmidEnd = Get end point: 4, syllIntervalNumber
        complexmidDuration = 'complexmidEnd' - 'complexmidStart'
        complexmidTot = 'complexmidDuration' + 'complexmidTot'

#calculate total length of different types of silence
    else
        nSyllInt = number(nSyllInterval$)
        nSyll = nSyllInt + nSyll
     endif
    endfor

# calculate measures of speed fluency
    speechRate = nSyll / 'endTime' * 60
    articRate = nSyll / 'phonationTime' *60

  # extract no. pauses from syllable tier
    nunfilledend = Count labels: 4, "0eu"
    nfilledend = Count labels: 4, "0ev"
    ncomplexend = Count labels: 4, "0ec"
    nunfilledmid = Count labels: 4, "0mu"
    nfilledmid = Count labels: 4, "0mv"
    ncomplexmid = Count labels: 4, "0mc"

# calculate no. pauses per minute
    unfilledEndPausePerMin = 'nunfilledend' / 'endTime' * 60
    filledEndPausePerMin = 'nfilledend' / 'endTime' * 60
    complexEndPausePerMin = 'ncomplexend' / 'endTime' * 60
    unfilledMidPausePerMin = 'nunfilledmid' / 'endTime' * 60
    filledMidPausePerMin = 'nfilledmid' / 'endTime' * 60
    complexMidPausePerMin = 'ncomplexmid' / 'endTime' * 60

# calculate average pause duration
    avgUnfilledEndPause = 'unfilledendTot' / 'nunfilledend'
    avgFilledEndPause = 'filledendTot' / 'nfilledend'
    avgComplexEndPause = 'complexendTot' / 'ncomplexend'
    avgUnfilledMidPause = 'unfilledmidTot' / 'nunfilledmid'
    avgFilledMidPause = 'filledmidTot' / 'nfilledmid'
    avgComplexMidPause = 'complexmidTot' / 'ncomplexmid'

# print results
printline 'fileName$', 'nSyll', 'endTime', 'silentTime', 'phonationTime', 'speechRate', 'articRate', 'nunfilledend', 'nfilledend', 'ncomplexend', 'nunfilledmid', 'nfilledmid', 'ncomplexmid',
'unfilledendTot', 'filledendTot', 'complexendTot', 'unfilledmidTot', 'filledmidTot', 'complexmidTot', 'unfilledEndPausePerMin', 'filledEndPausePerMin', 'complexEndPausePerMin',
'unfilledMidPausePerMin', 'filledMidPausePerMin', 'complexMidPausePerMin', 'avgUnfilledEndPause', 'avgFilledEndPause', 'avgComplexEndPause', 'avgUnfilledMidPause',
```

```
'avgFilledMidPause', 'avgComplexMidPause'
endfor
```

Appendix Q – Example of PRAAT Spectrogram Detail



Appendix R – Example of PRAAT Textgrid Detail



Appendix S – Limitation of Model Fitting

The step-by-step forward-selection procedure (Appendix M) employed for this analysis can possibly lead to underspecified models. These underspecified models may arise due to the possibility of a better fitting model, which includes a combination of parameters which have not been tested by the model selection procedure (Murakami, 2016). To guard against this possibility and mitigate the impact the model was constructed in a "2-in-1-out" fashion. This procedure involved adding two variables at a time which resulted in a decrease in the AIC, and then deleting the variable which resulted in the minimum decrease in AIC (Murakami, 2016). This approach allowed for a wider scope of inspection of the interaction effects of various different parameters by testing a greater combination of factors than a simple forward-selection procedure (Murakami, 2016).

When considering which specific elements to add to each model it is important that each factor must be carefully considered, and a one size fits all approach to model fitting is cautioned against by both Winter (2019) and Field et al., (2012). Instead, a careful consideration of which elements are to be explored should drive model creation rather than the more-the-merrier approach to model fitting. One problem which often arises in model fitting is whether or not to try and add random effects to the model structure, be they random intercepts or random slopes. In the case of this particular research project random intercepts are clearly needed, as it is plausible that some participants may pause more frequently when speaking in general than others so, at a bare minimum, the mixed model needs to account for this variation with the addition of participants as a random factor (Winter, 2019). However, the case for random slopes is often more complicated as the next section discusses.

One of the benefits of using mixed effects models is the ability to fit random slopes to account for the variation within participants who each provide multiple responses. Random slopes for complexity and format could be fitted as they will both vary within individuals as each participant responded to both levels of task complexity and both task formats. This means that it is theoretically possible, and supported by the literature (Winter, 2019; Field et al., 2012), to fit random slopes for the factors of task complexity and task format. It would be unreasonable to assume that all of the participants are affected by the differentiation in task complexity, and the differentiation in task format, in the same way. This difference can be seen by the quantitative data collected in the post task questionnaire (Appendices H, I, J, K, L), and detailed in section (7.2), reason to fit random slopes. Thus, it seems like a clear case that requires the inclusion of by-participant varying slopes, for the condition effects of task complexity and task format (Winter, 2019).

However, the mixed effects model employed in this research project will not be able to fit unrelated, separate linear regression models for each participant, it fits only those models which are 'related' and whose intercepts and/or slopes deviate randomly about a typical intercept and/or slope (Winter, 2019). The random deviations from the typical intercept and/or typical slope follow a normal distribution with a mean of zero and some unknown standard deviation (Field et al., 2012). As a result, when random slopes were added to the model for either task complexity or task format, as well as for both together, the model failed to converge, reporting an error for over fitting:

"number of observations (=160) <= number of random effects (=160) for term: *(1 + task_format | speaker_id);* the random-effects parameters and the residual variance (or scale parameter) are probably unidentifiable."

The mixed effects model here is struggling due to the small number of observations, with only two recorded responses for each participant. Kreft and de Leeuw (1998) point out that the more data you have the better it is for fitting a statistical model successfully, and Twisk (2006) agrees that the greater number of individual responses within each context is very important. If each subject had contributed 5 or 6 observations instead of only 2, possibly 3 responses for each condition of complexity and format, it would be much more likely to accommodate all of the random slopes. The current study design does not support the complex modelling of random slopes, and to support the integration of random slopes within the current model more data would be needed under each condition and for each subject (Field, et al., 2012; Winter, 2019).

References

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research*, *38*(2), 403-414.

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. International Journal of Language & Communication Disorders, 35(1), 95-116.

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

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Anderson, J. R. (2000). *Learning and memory: An integrated approach*. John Wiley & Sons Inc.

Awwad, A., & Tavakoli, P. (2019). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. International Review of Applied Linguistics in Language Teaching, 1(ahead-of-print).

Awwad, A., Tavakoli, P., & Wright, C. (2017). “I think that's what he's doing”: Effects of intentional reasoning on second language (L2) speech performance. System, 67, 158-169.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*(3), 189-208.

Baddeley, A. D. (1986). Working memory oxford. *England: Oxford Uni*.

Baddeley, A. D. (2000). Short-term and working memory. *The Oxford handbook of memory, 4*, 77-92.

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

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model.

Baddeley, A. D., Grant, S., Wight, E., & Thomson, N. (1975). Imagery and visual working memory. *Attention and performance V*, 205-217.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575-589.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. Psychological review, 105(1), 158.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233-252.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(4), 689-725.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. *URL http://www. praat. org/. Retrieved March*, *14*, 2018.

Bolibaugh, C., & Foster, P. (2013). Memory-based aptitude for nativelike selection: The role of phonological short-term memory. *Sensitive periods, language aptitude, and ultimate L2 attainment*, *35*, 205.

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

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of memory and language*, *75*, 104-116.

Breen, M. (1989). The evaluation cycle for language learning tasks. The second language curriculum, 187206.

Bruton, A. (2002). From tasking purposes to purposing tasks. *ELT journal*, *56*(3), 280-288.

Burrows, C. (2008). An evaluation of task-based learning (TBL) in the Japanese classroom. *English Today*, *24*(4), 11.

Bygate, M. (1998). Theoretical perspectives on speaking. *Annual Review of Applied Linguistics*, *18*, 20-42.

Bygate, M. (2009). 23 Teaching and Testing Speaking. *The handbook of language teaching*, 412.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology*, *32*(5), 867.

Christenfeld, N. (1994). Options and ums. *Journal of Language and Social Psychology*, *13*(2), 192-199.

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73-111.

Cohen, L. et al. (2013). Research methods in education. Routledge.

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.

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

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

De Jong, N. H. (2016). 13 Fluency in second language assessment.

De Jong, N. H. & Bosker. H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In Eklund, R. (ed.), Proceedings of dysfluency in spontaneous speech diss, 17-20. Stockholm: KTH.

De Jong, N. H. et al. (2012b). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performance of native and non-native speakers. In Housen, A. Kuiken, F. & Vedder, I. (Eds.), *Dimesnsions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121-142). Amsterdam: John Benjamins.

De Jong, N. H. et al. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(05), 893-916.

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

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

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.

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. (1998). Beyond focus on form. Focus on form in classroom second language acquisition, 42-63.

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

DeKeyser, R. (2008). 11 implicit and explicit learning. *The handbook of second language acquisition*, *27*, 313.

DeKeyser, R. (2017). Knowledge and skill in ISLA. The Routledge handbook of instructed second language acquisition, 15-32.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, *104*(4), 801.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189-1199.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language learning*, *54*(4), 655-679.

Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual review of applied linguistics*, *21*(1), 43.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in second language acquisition*, 349-385.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language teaching research*, *4*(3), 275-300.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition*, 305-352.

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

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

Ellis, R. (2017a). Position paper: Moving task-based language teaching forward. *Language Teaching*, *50*(4), 507-526.

Ellis, R. (2017b). Task-based language teaching. The Routledge handbook of instructed second language acquisition, 108, 125.

Ellis, R. (2020). A short history of SLA: Where have we come from and where are we going? *Language Teaching*, 1-16.

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General, 128*(3), 309.

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.

Feng, Y., & Webb, S. (2019). Learning vocabulary through reading, listening, and viewing: which mode of input is most effective? *Studies in Second Language Acquisition*, 1-25.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

Fillmore, C. J. (1979). On fluency. In *Individual differences in language ability and language behavior* (pp. 85-101). Academic Press.

Foster, P. (1999). Task-based learning and pedagogy.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In Bygate, M. Skehan, P. and Swain, M. (eds.), *Researching Pedagogic Tasks: Second language learning, teaching, and testing*, 75–93. London, Longman.

Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In *Vocabulary studies in first and second language acquisition* (pp. 91-106). Palgrave Macmillan, London.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, *18*(3), 299-323.

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

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

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language*, *69*(3), 196-227.

French, L. M. (2006). Phonological working memory and second language acquisition: A developmental study of francophone children learning English in Quebec. Edwin Mellen Press.

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

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. *Theories in second language acquisition: An introduction*, *175199*.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied psycholinguistics*, *27*(4), 513.

Gathercole, S. E., & Baddeley, A. D. (1989). The role of phonological memory in normal and disordered language development. In *Brain and reading* (pp. 245-255). Macmillan Education UK.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*(4), 439-454.

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.

Gathercole, S. E., & Thorn, A. S. (1998). Phonological short-term memory and foreign language learning. *Foreign language learning: Psycholinguistic studies on training and retention*, 141-158.

Gathercole, S. E., Hitch, G. J., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental psychology*, *33*(6), 966.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental psychology*, *28*(5), 887.

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.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*(3-4), 367-395.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech.

Greene, J. O., & Cappella, J. N. (1986). Cognition and talk: The relationship of semantic units to temporal patterns of fluency in spontaneous speech. *Language and Speech*, *29*(2), 141-157.

Gries, S. T. (2016). Quantitative corpus linguistics with R: A practical introduction. Taylor & Francis.

Handley, Z., & Haiping, W. (2018). What is the impact of study abroad on oral fluency development? A comparison of study abroad and study at home. ELT Research Papers, 17(5).

Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental psychology*.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*(2), 153-166.

Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. *Measuring L2 proficiency: Perspectives from SLA*, *27*, 53.

Hoey, M. (2012). Lexical priming. The encyclopaedia of applied linguistics.

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

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *Journal of Asia TEFL*, *3*(4).

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct?. *Applied linguistics*, *29*(1), 24-49.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing?. *Trends in cognitive sciences*, *10*(11), 480-486.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122.

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.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569-591.

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*(2), 457-480.

Kempe, V., & Brooks, P. J. (2011). Individual differences in adult second language learning: A cognitive perspective. Scottish Languages Review, 23, 15-22.

Kim, Y. (2009). The effects of task complexity on learner–learner interaction. *System*, *37*(2), 254-268.

Kim, Y. (2013). Promoting attention to form through task repetition in. *Second language interaction in diverse educational contexts*, *34*(3).

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

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.

Kohonen, V., & Service, E. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?. *Applied Psycholinguistics*, *16*(02), 155-172.

Kormos, J. (2003). Attention and monitoring in a second language. *Mündliche Produktion in der Fremdsprache*, *32*, 116.

Kormos, J. (2006). Speech production and second language acquisition. Mahwah, NJ: Lawrence Erlbaum.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*(2), 148-161.

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

Kormos, J., & Trebits, A. (2011). Verbal Working Memory and Language Production. Second language task complexity. *Researching the Cognition Hypothesis of Language Learning and Performance; Robinson, P., Ed*, 267-285.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, *62*(2), 439-472.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. The modern language journal, 73(4), 440-464.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Sage.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of second language writing*, *17*(1), 48-60.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, *14*(4), 389-433.

Lahmann, C. et al. (2015). Linguistic skills and speaking fluency in a second language. *International Journal of Bilingualism*, 1-15.

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

Lay, C. H., & Paivio, A. (1969). The effects of task difficulty and anxiety on hesitations in speech. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *1*(1), 25.

Lee, J. F. (2000). Tasks and communicating in language classrooms. McGraw-Hill.

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

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

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

Levelt, W. J. & Speaking, M. (1989). From intention to articulation. Cambridge, Mass. Institute of Technology.

Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language teaching*, *40*(3), 243.

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

Long, M. H. (1985). Input and second language acquisition theory. *Input in second language acquisition*, *377*, 393.

Long, M. H. (2015). Second language acquisition and task-based language teaching. Oxford, UK: Wiley-Blackwell.

Long, M. H. & Norris, J. (2009). Task-based teaching and assessment. In Van den Branden, K., Bygate, M., & Norris, M. J (Eds.), *Task-based language teaching: A reader*. Amsterdam, the Netherlands: Benjamins.

Mackey, A. & Gass, S. M., (2015). Second language research: methodology and design. Routledge.

Malicka, A., & Levkina, M. (2012). Measuring task complexity Does EFL proficiency matter?. *Task-based language teaching in foreign language contexts: Research and implementation*, *4*, 43.

Marsden, E. (2007). Can educational experiments both test a theory and inform practice?. *British Educational Research Journal*, *33*(4), 565-588.

Marsden, E., & Torgerson, C. J. (2012). Single group, pre-and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*(5), 583-616.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in second language acquisition*, *34*(3), 379-413.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McManus, K., & Marsden, E. J. (2018). Online and offline effects of L1 practice in L2 grammar learning: A partial replication. *Studies in Second Language Acquisition*, 459-475.

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

Moser, J., Harris, J., & Carle, J. (2012). Improving teacher talk through a task-based approach. *ELT journal*, *66*(1), 81-88.

Murakami, A. (2016). Modelling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*(4), 834-871.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33.

Nuevo, A. M. (2006). *Task complexity and interaction: L2 learning opportunities and development* (Doctoral dissertation, Georgetown University).

Nunan, D. (2006). Task-based language teaching in the Asia context: Defining 'task'. *Asian EFL journal*, *8*(3).

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in second language acquisition*, 557-581.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. Studies in second language acquisition, 109-148.

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

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98-107.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, *191*, 225.

Pawley, A., & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In *Perspectives on fluency*. University of Michigan Press.

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

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2018). R Core Team. 2018. nlme: linear and nonlinear mixed effects models. R package version 3.1-137. *R Found. Stat. Comput. Retrieved from https://CRAN. R-project. org/package= nlme (accessed 19 Jul. 2018)*.

Portney, L. G., & Watkins, M. P. (2000). Foundations of clinical research: application to practice. Upper Saddle River, NJ: Prentice Hall.

Prabhu, N. S. (1987). Second language pedagogy. Oxford: Oxford University Press.

Préfontaine, Y. (2013). Perceptions of French fluency in second language speech production. *Canadian modern language review*, *69*(3), 324-348.

Préfontaine, Y., & Kormos, J. (2015). The relationship between task difficulty and second language fluency in French: A mixed methods approach. The Modern Language Journal, 99(1), 96-112.

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

Rast, R. (2008). Foreign Language Input: Initial Processing (vol.28). Multilingual Matters.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*, 162-181.

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

Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes*, *15*(2), 129-157.

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

Robinson, P. (2003a). The cognition hypothesis, task design, and adult task based language learning. *Second Language Studies*, 21(2). pp. 45-105.

Robinson, P. (2003b). Attention and memory during SLA. *The handbook of second language acquisition*, 631-678.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 1-32.

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

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

Robinson, P. (2011a). Task-based language learning: a review of issues. In P. Robinson (Ed.) *Task-based language learning* (pp. 1-36). Oxford, UK: Wiley-Blackwell.

Robinson, P. (2011b). 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, the Netherlands: Benjamins.

Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In *Domains and directions in TBLT: Plenaries from a decade of the international conference* (pp. 87-122). John Benjamins Amsterdam.

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

Sakui, K. (2007). Classroom management in Japanese EFL classrooms. *JALT JOURNAL*, *29*(1), 41.

Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. Researching pedagogic tasks: Second language learning, teaching and testing, 8(3), 119-140.

Samuda, V., & Bygate, M. (2008). Tasks in second language learning. New York: Palgrave Macmillan.

Sasayama, S. (2016). Is a ‘complex’ task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, *100*(1), 231-254.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).

Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling Type I error.

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

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. *Handbook of bilingualism: Psycholinguistic approaches*, *371*, 388.

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

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21-50.

Service, E., Maury, S., & Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Memory & cognition*, *35*(5), 1122-1135.

Shehadeh, A. (2005). Task-based language learning and teaching: Theories and applications. In Teachers exploring tasks in English language teaching (pp. 13-30). Palgrave Macmillan, London.

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

Skehan, P. (1992). Second language acquisition strategies and task-based learning. *Thames Valley University Working Papers in English Language Teaching*, *1*, 178-208.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

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

Skehan, P. (2003). Task based instruction. *Language Teaching*, 36, 1-14. London, UK: Cambridge University Press.

Skehan, P. (2006). Second and foreign language learning and teaching.

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

Skehan, P. (Ed.). (2014). *Processing perspectives on task performance* (Vol. 5). John Benjamins Publishing Company.

Skehan, P. & Shum, S. (2014). Structure and processing condition in video-based narrative retelling. In Skehan, P. (ed.) *Processing perspectives on task performance,* 187-210. Amsterdam, John Benjamins.

Skehan, P. et al. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, 36. de Gryuter Mouten.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. Language teaching research, 1(3), 185-211.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language learning*, *49*(1), 93-120.

Skehan, P., & Foster, P. (2001). Cognition and tasks. *Cognition and second language instruction*, *183*, 205.

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. *Complexity, accuracy, and fluency in second language use, learning, and teaching*, *207*, 226.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied psycholinguistics*, *25*(2), 293.

Suzuki, S., & Kormos, J. (2019). The effects of read-aloud assistance on second language oral fluency in text summary speech.

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143-167.

Swan, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied linguistics*, *26*(3), 376-401.

Swan, M. (2005). *Practical English Usage*. Oxford Univ. Press.

Sweller, J. (1999). Instructional design. In *Australian educational review*.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast?. *Memory & cognition*, *37*(4), 529-540.

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System, 37*(3), 482-495.

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. & Skehan, P. (2005). Strategic planning, task structure, and performance testing. Planning and task performance in a second language, 11, 239-273.

Tavakoli, P., & Foster, P. (2008). Task design and L2 performance. *Language Learning, 58*(2), 429-473.

Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *61*, 37-72.

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

Tavakoli, P., & Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. In *Planning and task performance in a second language* (pp. 239-273). John Benjamins.

Torgerson, D. (2008). Designing randomized trials in health, education and the social sciences: an introduction. Springer.

Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British journal of educational studies*, *51*(1), 36-45.

Twisk, J. W. (2006). Applied multilevel analysis: a practical guide for medical researchers. Cambridge university press.

Van den Branden, K. (2016). Task-based language teaching. *The Routledge handbook of English language teaching*, 238-251.

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289-296.

Wen, Z., Mota, M. B., & McNeill, A. (Eds.). (2015). *Working memory in second language acquisition and processing* (Vol. 87). Multilingual Matters.

Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied linguistics*, *10*(2), 128-137.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. Language Testing, 14(1), 85-106.

Williams, J. N. (2012). Working memory and SLA. In S. Gass & A. Mackey (Eds.), Handbook of second language acquisition (pp. 427–441).

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language learning*, *53*(1), 67-121.

Williams, J. N., & Lovatt, P. (2005). Phonological memory and rule learning. *Language learning*, *55*(S1), 177-233.

Willis, J. (1996). *A framework for task-based learning* (Vol. 60). Harlow: Longman.

Willis, J., & Willis, D. (2013). Doing task-based teaching-Oxford handbooks for language teachers. Oxford University Press.

Winter, B. (2019). Statistics for linguists: An introduction using R. Routledge.

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

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

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