Task complexity as mediated by proficiency, working memory, and attention: Competition for cognitive resources during L2 oral task performance

James William Pownall Graumann

Task complexity as mediated by proficiency, working memory, and attention: Competition for cognitive resources during L2 oral task performance

TESI DOCTORAL
presentada per

James William Pownall Graumann

Com a requeriment per a l'obtenció del títol de
Doctor en Lingüística Aplicada

Director: **Dr. Roger Gilabert Guerrero**

Programa de Doctorat: Estudis Lingüístics, Literaris i Culturals

Departament de Filologia Anglesa i Alemanya
Universitat de Barcelona
Setembre 2016

ACKNOWLEDGEMENTS

First and foremost I would like thank my son, Joshua, and my wife, Mercè, who have supported me in every way imaginable over the last few years. From the time I began the master’s program in applied linguistics at the University of Barcelona up through the writing of this dissertation, they’ve never left my side. And not just my side; they’ve pushed me from behind when I need a push and lifted me up from below whenever I was down. They were always where I needed them. Nothing would be complete if it hadn’t been for them. All the good is dedicated to you.

I want to thank my supervisor Dr. Roger Gilabert for his help throughout the entire process which led up to this dissertation. It was Roger who moved my interest in the study of the cognitive processes which are involved in language acquisition. Apart from my motivation, Roger has always been a source of inspiration; someone to look up to as a great professional who knows the field and the means to get around. He’s guided me the whole time as a leader, a teacher and a provider for whatever it was that I needed to move along with the project. I am grateful and indebted to him for all he’s done.

I also want to thank everyone else that I’ve come to know over the years at the University of Barcelona for their companionship and inspiration. Kerry Brennan, Jessica MacKay, and Aleksandra Malicka who I got to know during the master’s degree course as I and whose camaraderie I have counted on all along. Mayya Levkina for her friendship, help, and expertise whenever it was needed. Every time we met it seems that she was offering a helping hand with something. A special thanks to Lena Vasylets who always gave me fresh ideas and who could always find the time to sit down and talk about things. She introduced me to the measure of idea units and provided me with everything I needed to make them a

key part of my dissertation. A thanks also to Dr. Joan Carles Mora for the extra time afforded me on various occasions, especially as I was putting together my program for the alternating runs paradigm. He gave me wonderful advice and resources which were of great help. And a thanks to all those who I haven’t mentioned but who have all formed a wonderful part of my time here at the university.

I’d like to reserve a special expression of my gratitude to my colleagues at Blanquerna School of Communication Studies at the University of Ramon Llull. It’s a great joy to be able to count on colleagues whose support at all times is beyond what one might expect. Most of the volunteers who participated in my experiments came from their classes, and that came about because of their selfless implication in my own project to no benefit of their own. For this and for much more I am grateful to you all.

A special thanks as well to my colleagues, Steve, Noemi, and Maria Jose at Activ-Lingua School of Languages and ICL Idiomas whose support I was able to count on at all times.

I must certainly thank all those volunteers who participated as a part of my experiment and pilot studies. You are many, and it is impossible to thank each and every one again, but without your help, nothing could have been completed. I am grateful for the time that you dedicated to this project.

Finally, I would like to send a special thanks to my parents and family back in the states who, despite distances, are always with me in spirit, and most specially, Malia, Katie, Stephen, Drew, and Zach.

ABSTRACT

Drawing on known language learning processes, principles behind language course syllabus design, and previous second language acquisition research, Long introduced the concept of task-based language teaching (TBLT) in 1985. Interest in TBLT has prompted research into the effects of task demands on the learning process. Interactive demands promote the negotiation of meaning between interlocutors and provide opportunities for uptake of corrective feedback from a teacher or fellow students. Cognitive demands can direct attention to primary aspects of the language ensuring successful task performance. Increasing cognitive demands builds on previously acquired access and automatization of speaking processes pushing learners to approach complex concepts with suitable speech.

The advent of task-based language teaching included a call for establishing a set of criteria by which tasks could be graded and sequenced within a language learning syllabus. Tasks sequenced from simple to complex are believed to enhance the learning process by recreating ontological learning processes while benefiting from adults' conceptual understanding of the world. This call has motivated research within the field of second language acquisition (SLA) to understand the influence that task design features have on language production and learning processes in order to develop the needed task sequencing criteria. Much research has focused on how cognitive resources are allocated to different aspects of speech within the bounds of two hypotheses which attempt to predict outcomes: The Trade-off Hypothesis (Skehan, 1998; Skehan & Foster, 2001) and the Cognition Hypothesis (Robinson 2001a, 2001b, 2003a, 2005). Conclusive results, however, have been elusive.

The study of individual differences and the role that they play in language production and acquisition processes may provide some direction to researchers. The current dissertation proposes that learners of varying cognitive capacities would demonstrate trade-off effects between measures of linguistic complexity, accuracy and fluency at different points along a continuum of task complexity. In addition, where trade-offs were detected, it was predicted that subjects of higher cognitive capacity would demonstrate them more clearly. Five narrative tasks each representing different levels of cognitive complexity were performed by 47 subjects who were categorized according to high and low levels of proficiency, and attentional and working memory capacity. Repeated measures and correlational analyses did not provide support for the hypotheses although they do indicate that individual differences may play a role in communicative strategies that subjects use in order to meet task demands.

RESUM

L’any 1985, Long va introduir el concepte d’ensenyament basat en les tasques pedagògiques (TBLT) basant-se en els coneixements dels processos d’aprenentatge, en els fonaments del disseny dels plans d’estudis de cursos de llengües estrangeres i en la investigació realitzada prèviament en l’àmbit de l’adquisició d’idiomes. El interès en TBLT ha motivat diverses investigacions sobre els efectes dels requisits de les tasques en els processos d’aprenentatge.

Els requisits interactius fomenten un diàleg sobre aspectes de la llengua entre els interlocutors i proporcionen oportunitats per captar comentaris correctius del professor o altres estudiants. Els requisits cognitius d’una tasca poden orientar l’atenció d’un alumne als aspectes importants de la llengua, assegurant així un resultat satisfactori. Augmentant els requisits cognitius de les tasques permet consolidar processos com són els d’accés i d’automatització de certs aspectes de la llengua. Això incentiva als alumnes a fer servir un llenguatge adequat al enfrontar-se a tasques complexes.

L’aparició de l’ensenyament basat en les tasques pedagògiques va incloure un reclam per establir un conjunt de criteris mitjançant els quals es podia graduar i seqüenciar les tasques en un pla d’estudis. Es creu que seqüenciar les tasques de simple a complexa millora el procés d’aprenentatge al recrear el procés de desenvolupament ontològic. Aquest reclam ha motivat investigacions en el camp de l’adquisició d’idiomes estrangers (SLA) per entendre la influencia del disseny de les tasques sobre l’ús i els processos d’aprenentatge dels idiomes amb la finalitat de desenvolupar els criteris. La investigació s’ha centrat en com els recursos cognitius estan repartits entre els diferents aspectes de la parla dins els paràmetres de dos hipòtesis: La *Trade-off Hypothesis* (Skehan, 1998; Skehan i

Foster, 2001) i la *Cognition Hypothesis* (Robinson 2001a, 2001b, 2003a, 2005). Resultats concloents, malgrat tot, han estat elusius.

L'estudi de les diferencies individuals i el paper que assumeixen en els processos de la parla i de l'adquisició d'idiomes poden orientar als investigadors. La present dissertació proposa que els aprenents amb capacitats cognitives diferents demostren efectes de *trade-off* (compensació) entre mesures de complexitat lingüística, precisió gramatical i fluïdesa amb resultats diferents. Quaranta set participants van completar cinc tasques narratives, cada una d'elles representativa d'un nivell de complexitat cognitiva diferent. Els participants van ser classificats segons el seus nivell de coneixement del idioma i de les seves capacitats d'atenció i de memòria de treball. Un anàlisi estadístic dels resultats no demostra les hipòtesis com a certes malgrat que sí que indica que les diferencies individuals poden influir en les estratègies comunicatives que els participants van utilitzar per completar la tasca.

TABLE OF CONTENTS

Chapter 7 Results 131

List of tables:

List of figures:

CHAPTER 1: OVERVIEW

1.1 Introduction

Background for the present study begins in the area of task-based language teaching and the challenges which have arisen out of the need to determine the best way to sequence tasks in a language learning syllabus. Long, in 1985, while introducing a general framework for task his based teaching approach suggested that a task-based syllabus present pedagogical tasks in a graded order from simple to complex as determined by the intrinsic difficulty that each task represents for the learner. The final goal is to eventually present pedagogic tasks which simulate the highly complex target tasks found in the real world. The challenge of doing so in an effective and empirically sound format has opened up a whole new area within the field of second language acquisition (SLA) research. The objective is to find a way to grade tasks in terms of increasing complexity and to understand how learners perform and benefit from conscious manipulation of task characteristics that designers can incorporate into task-based syllabi.

Various SLA perspectives on learning have been drawn on by researchers in task-based language teaching. Two of these include the interactionist perspective the information-processing perspective. The former involves investigation into tasks which promote negotiation for meaning, pushing second language development. The latter, the information-processing perspective, is closely intertwined with work in the area of cognitive psychology on learning and performance. From this latter perspective, researchers investigate how task performance requires the focus of, and instigates the consequential competition for, limited cognitive resources to an extent which is determined by the characteristics of the task.

Two researchers, Skehan and Robinson, stand out in recent SLA literature concerning the information-processing perspective. In 1998, Skehan presented what has come to be known as the trade-off hypothesis (Skehan, 2009). This hypothesis essentially states that committing attentional resources to one dimension of linguistic production may have a negative impact on the others. He suggests a competition for attentional resources between fluency and form during speech production. Within the construct of form, there is additional competition between grammatical form and meaning. Greater task complexity, while depleting attentional resources, would then not be expected to advantage all three performance areas simultaneously as remaining resources would be allocated to prioritize aspects of speech in accordance with the demands of the task. In contrast, Robinson's Cognition Hypothesis (Robinson 2001a, 2001b, 2003a, 2005), suggests that various pools of attentional resources allow for those resources to be allocated simultaneously toward both meaning and grammatical form without suffering competition between them.

Results of research carried out to determine how cognitive resources compete have provided little clear evidence in favor of either one or the other hypothesis prompting a change of direction in recent investigation. While many of these experiments have focused on language produced as task variables were manipulated to increase cognitive complexity, results have been somewhat inconclusive. Researchers are currently turning to the study of individual differences to help clarify remaining doubts.

Of the various categories of individual differences, working memory capacity has been proven to be a strong predictor of language comprehension (Daneman & Merikle, 1996) as well as of general L2 proficiency (Kormos &

Sáfár, 2009; Guará Tavares, 2009; Finardi, 2008; Finardi, & Weissheimer, 2008; Borges Mota Fortkamp, 2007; Bergsleithner, 2007; Borges Mota, 2003; Fortkamp, 1999; Borges Mota Fortkamp, 1998). However, little investigation has been done that combines the factors of increased cognitive complexity of tasks and measures of working memory or attentional capacity to discover whether individuals who differ in these respects produce language in distinct ways under increasingly demanding conditions.

The present investigation intends to investigate the role that working memory capacity, attentional capacity may play in the way learners perform second language (L2) oral production tasks. Proficiency will also be taken into account as it can be expected to exert an influence results as well (Gilabert, Levkina, & Baron, 2011). Results may provide insight into why research so far has been inconclusive about the effects of manipulating task characteristics along levels of task complexity.

1.2 Relevance

This research is relevant to current trends in SLA research that focus on the efficient sequencing of materials used in Task-Based Language Teaching (TBLT) and Content and Language Integrated Learning (CLIL) classrooms. Established guidelines that dictate how tasks should be created and positioned within a teaching sequence should allow for them to be used with maximum efficacy as a language learning tool. To create these guidelines, researchers are currently investigating task characteristics under varying performance conditions. The goal of this dissertation is to provide researchers with notions about the relevance that working memory capacity, attentional capacity, and proficiency have in oral language production. Such notions may lead to research methods which more accurately are able to predict performance on oral tasks as individuals become more challenged by increases in task complexity. More conclusive results of empirical investigation may ultimately lead to more effective guidelines for creating and sequencing pedagogic tasks for optimal learning effects in TBLT and CLIL environments.

1.3 Structure of the dissertation

The dissertation is divided into eight different chapters. Chapter 2 begins with a discussion of the concept of Task-Based Language Teaching. A general background of TBLT will be provided as well as the issues surrounding sequencing of tasks within a TBLT syllabus. Two perspectives of development, the interactionist perspective and the cognitive processing perspective will also be introduced as means from which researchers have been reviewing issues in SLA to clarify the underlying processes involved in language acquisition. This dissertation will be investigation matters from a cognitive processing perspective.

Chapter 3 describes several prominent speech production models which have been influential in SLA research. Two prominent hypotheses concerning cognitive processes behind language acquisition, the Trade-Off Hypothesis and the Cognition Hypothesis will be described, as well as associated frameworks which are intended to provide rationale by which researchers as well as by educators can predict language performance from characteristics of factors which make up a task.

Chapter 4 delves into the constructs of memory, attention, and other kinds of personal characteristics which make up the individual differences which account for the reasons why everyone reacts differently to instructional methods.

Memory and attention are especially pertinent to the current investigation as these are the principle distinguishing factors by which participants will be compared.

Chapter 5 is a review of relevant literature which inspired to motivation behind the current research project. In this chapter, research questions will be posited and the corresponding hypotheses will be proposed.

Chapter 6 provides a detailed description of the methodology and protocols followed by which the investigation was carried out. An account in support of the methods and tests that were created and used as well as the processes of verification where they were necessary will be provided.

Chapter 7 will describe the results of the statistical analysis of the data. This information will be further discussed in chapter 8 where conclusions will also be drawn.

CHAPTER 2: TASK-BASED LANGUAGE TEACHING

2.1 Introduction

This chapter introduces task-based language teaching methodology and addresses the issue of task sequencing. Earlier teaching methodologies used linguistic units that were broken down and presented to the learner piece by piece. The learner was intended to re-synthesize the language into a functional tool. Task-based learning methodologies, in contrast, focus on tasks, a medium by which a learner draws on known elements of a target language to attain a goal, but which also pushes the learner a bit further, leading to interlanguage improvement (Long & Robinson, 1998).

2.2 Background

Tasks, the primary component of task-based language teaching, are pedagogic tools that provide situations under controlled conditions where learners can develop language abilities that they can use to meet the demands imposed by real world situations. Task performance within a language learning environment develops interlanguage through various means. Negotiation of meaning occurs when learners are pushed to make clarification requests by asking about information that has not been understood, and to make confirmation checks to ensure that they have understood their interlocutor properly or to be certain their interlocutor has understood what they are saying. Task performance also elicits recasts as positive feedback from the teacher, as well as language related episodes during which learners comment on the target language drawing their attention to its features, promoting noticing and aiding acquisition (Long & Crookes, 1992).

Long recognized the potential of tasks as an alternative unit for sequencing language learning syllabuses to substitute synthetic systems and methods which overly focus on forms. In 1985 he introduced a general framework for task-based language teaching (TBLT) methodology which he argued was based on what was already known about the processes of language learning, findings from second language classroom research, and principles behind English as a foreign language course design.

According to Long's framework, language teaching programs are designed beginning with an analysis of learners' needs in order to establish course objectives. These objects become the *target tasks* which learners are intended to master eventually; answering the phone in an office, giving directions to someone on the street, or any other kind or communicative procedures which have a defined outcome. The target task is that task which the learner needs to be able to perform in real life in the target language. Once the course objective is identified, corresponding pedagogical tasks can be chosen always with the target task in mind. The pedagogical tasks are the vehicles used in classroom situations by which target language samples are presented. In addition, pedagogical tasks present opportunities for comprehension and production and are sequenced according to the task syllabus. The syllabus presents tasks from simple to complex as determined by the intrinsic difficulty that each task represents for the learner. The final tasks simulate to the greatest degree the highly complex target tasks found in real world situations.

As an analytic approach to language teaching (Long & Robinson, 1998), TBLT methodology presents a target language in whole chunks at a time with limited intervention or control, relying on learners' abilities to recognize patterns in input and to draw conclusions as to what linguistic forms can or cannot be used to express an idea. Proponents for such non-interventionist approaches to language teaching draw in part on observations made of learners who pass through natural developmental sequences, each marked by specific aspects of a series of interlanguage structures. The passage through these sequences on the way to higher levels of target language proficiency appears to be both unavoidable and unalterable by instruction. So, the actual sequence in which learners assimilate the new linguistic code will not be altered, in spite of a formal synthetic syllabus where chosen language items are presented on a step by step basis in a pre-determined sequence.

The pitfalls of synthetic syllabi notwithstanding, Long credits results of previous studies in SLA that suggest that formal methods of instruction do offer important benefits to learners. In order to tap these benefits, Long’s framework deviates from the core ideals of the non-interventionists and reserves an important role for focus on form in the TBLT classroom. Focus on form refers to situations where students’ attention is drawn to linguistic code features of the target language as a part of task completion, but not as a principal objective. As a result, and in comparison to outcomes of naturalistic, or purely non-interventionist learning approaches, the rate of language learning speeds up, acquisition processes are affected in ways that are beneficial to long term accuracy, and the ultimate level of attainment is higher (Long, 1991).

The learning processes involved in TBLT do not imply step by step acquisition of concrete linguistic elements. However, through application of general cognitive processing capacities of attention and working memory, input will be reshaped and new form-function relationships made evident to the learner (Doughty, 2001). As these relationships are strengthened, made more readily retrievable from long term memory and incorporated into more complex associations, the target language is expected to grow in development (Long and Crookes, 1992).

2.3 Sequencing

The TBLT approach came up against criticism particularly as to how it should be implemented. A key factor to be resolved in order for TBLT to become fully functional is establishing definitive means for determining how to sequence tasks from simple to complex so that learners can be guided through the acquisition process in the most efficient way possible (Sheen, 1994; Robinson & Urwin, 1995; Skehan, 1996).

The sequencing of tasks within a task-based language learning syllabus is intended to reproduce the ontological evolution of both cognitive and linguistic capacities that takes place during a person’s formative years. Evidence of correlations of morphological acquisition orders and developmental sequences between L1 and adult L2 learners indicate that during linguistic expression a general universal sequence of underlying notions shared by children maturing in their conceptual development with adults who are learning a new language becomes apparent; adults have been observed to use simple structures while beginning the learning process of an L2 (Perdue 1993).

The apparent conclusion is that there is a similarity between how children learn an L1 and how adults learn an L2. However, as pointed out by Slobin (1993) and further adapted by Robinson (2003a) for matters concerning task sequencing, there is a large difference that exists between child L1 and adult L2 learners. While a child lacks certain conceptual notions at given points during their linguistic development, adults are fully developed cognitively at the outset of learning an L2. A child must first develop a notion, understand it, and only after gaining an understanding of that notion can the child learn how to express it. An adult, on the other hand, will already have a clear understanding of such conceptual matters and must only be concerned with knowing how to express them during their L2 learning process.

The question has arisen as to why adults have difficulty acquiring adequate expressive abilities in an L2 when the concepts which underlie the related notions are already present. Slobin (1993) gives two reasons for simplified adult language during L2 acquisition despite their fully developed cognitive capacities. First, he suggests that adults may be calling upon a primordial sense of simplicity which they have retained since the time of their own cognitive development, leading them to focus on the simplest notions available to them within the target language to meet task demands. The second reason is twofold: Either simple linguistic aspects are used more frequently in a target language and this is being reflected by learners' interlanguage, or learners are simply not able to access the linguistic means for expressing complex notions in early stages of their interlanguage development.

Whichever the reason may be, sequencing of tasks from simple to complex simulates the L1 learning process of a child who develops in cognitive capacities, learns to understand increasingly complex notions about the surrounding environment, and then learns to express those concepts through increasingly complex language to meet the complex demands of the environment. While the adult already possesses the cognitive machinery, the process becomes one of pushing that machinery to express increasingly complex notions through increasingly complex codes of the target language (Robinson, 2003a).

2.4 Interactionist and information processing perspectives of development

There are two main perspectives from which the issue of L2 development is addressed. The interactionist perspective (Long, 1985; Pica, 2006; Gass and Mackey, 2006) and the information-processing perspective (Skehan, 1996, 1998; Skehan & Foster, 1997, 2001; Robinson, 1995a, 1995b, 2001a, 2001b). Interactionist theory stresses the importance of social interaction for cognitive development in children. The interactionist perspective involves the study of tasks which drive interaction and promote negotiation for meaning as a tool for pushing second language development. The latter, the information-processing perspective, is closely intertwined with how research in the area of cognitive psychology approaches the issue of learning and performance.

The information processing perspective approaches the mind as a limited capacity system set up to process information based on a logical set of rules. Researchers working from this perspective describe and explain cognitive development through an understanding of the mechanisms and strategies applied by an individual while performing an activity. Analogies of the workings of the mind are often made to computer systems, simply in the way information is introduced, stored and processed for an eventual output although the human mind far outperforms even the most complex computer systems and algorithms. Structures and processes involve stimuli perception, attention, awareness, noticing, storage and retrieval, of information which, according to some models, are performed under the regulating instrument of executive function which monitors, selects and organizes the processes by which information is managed.

Understanding these processes is essential for researchers who are developing criteria that can be used for determining how to sequence pedagogic tasks in TBLT syllabi. Task analysis is commonly used to look for evidence in how cognitive development occurs as systems process information efficiently. Task sequencing criteria based on efficient processing of items to be learned should simulate a natural learning process. But task manipulation intended toward more efficient information processing may not benefit everyone universally. Efficiency in processing may depend on various factors; differences in the way information in encoded by information processing systems, or changes in strategies applied by an individual, the degree to which processes have become automatized, how the stimuli is represented, or as a result of individual differences in capacities of cognitive resources. Regardless of the individual, however, capacity for cognitive resources such as those needed for attention and memory is limited. The ensuing competition for these resources by different cognitive processes which need them to function ensures that they can be allocated in the most efficient way possible according to the demands that a task requires (Wickens 2002).

2.5 Summary

This chapter reviewed some of the background behind the development of task-based language teaching methodologies. The issue of task sequencing was also addressed and the role played therein by the information processing perspective on cognitive development. The next chapter will focus on models of speech production and related hypotheses concerning L2 acquisition.

CHAPTER 3: SPEECH PRODUCTION

3.1 Introduction

In the previous chapter, Task-Based Language Teaching was discussed as well as issues concerning the challenges of sequencing of tasks for effective task-based syllabus design. In this chapter, several models of speech production will be described. Speech production models have drawn on information processing approaches to develop a picture of verbal communication with memory and attention playing a vital role in explaining how the processes work. These models tend to focus on L1 production but are commonly used as well in SLA research to understand processes involved in language acquisition. The chapter will also address two hypotheses which draw on the information processing perspective of language acquisition as well as the speech production models to describe the interaction of task characteristics and their influence on language production.

3.2 Background

According to Kormos (2006), speech production researchers agree on the most important components involved in speech. *Conceptualization* is the planning stage during which an individual decides upon the information that they want to convey. *Formulation* involves the means to encode a message grammatically, lexically, and phonologically. *Articulation* involves the speech sounds which finally transmit the end message. *Self-monitoring* is the system by which speech and message are controlled for correctness and appropriateness. Attention in essential to the message planning stage in which the underlying ideas are conceptualized. The formulation and articulation stages are automatic in L1 speech. Processing mechanisms involved in these stages can work in parallel which allows for fluid and fast transmission of the concept.

The spread of activation model is commonly drawn on by speech production theorists. The model is a mathematical construct by which activation is spread over associative networks. This has been adapted to theories in neuropsychology, and memory and semantic processing (Collins and Loftus, 1975; Anderson, 1983). The theories are based on findings that networks are formed between interconnecting neurons which activate one another. Varying factors including externally or internally produced stimuli determine the formation of neural pathways. The frequency with which these same pathways are used determines the efficiency with which the connections perform, creating a substrate upon which memories may be built. Pathways are modified on a continuous basis. Memory becomes an active process of reclassification rather than a matter of retrieval from a fixed or static store as might be implied by the computer systems allegory mentioned earlier. Information retrieval occurs when connections between neural networks are formed or disassembled according to stimulus in the environment or demands of a task at hand until the desired networks are organized or activated to satisfactorily meet one's needs. This forms the basis of *connectionism* and related theories. This is plays an important role in speech processing models which assume that processes of conceptualization, formulation, and articulation depend on the transmission of information between the mechanisms involved and neural networks which form memory stores.

3.3 Levelt's speech production model

Among psycholinguistic models of speech production, Levelt's (1989, 1999) model of L1 speech production is prominent. The model draws heavily on the concept spread of activation theory, but as a modular theory, adheres to the precept that flow of activation is directed in a single direction across modules, or a linear series of specialized components. Errors are perceived only after phonological encoding or actual speech at which point the message must be re-conceptualized and pass through the whole process again. Despite serial processing, this does not mean that only one concept is produced at a time. Any component of the process may be activated by only a fragment of input. So, all components work at the same time on varying aspects of the message allowing for simultaneous organization of several concepts.

Modular theories contrast with spreading activation models of speech which allow for activation to reverse itself immediately upon perception of errors. According to this train of thought, the development of a concept is existent at various times as a semantic, syntactic, morphological and phonological representation. The development at the different levels occurs simultaneously. While one level may depend on the rate of processing of another level in order for its completion, there is a continuous back and forth of information until the final representation is available for articulation (Dell 1986).

Levelt (1989, 1999) developed a ‘blueprint of the speaker’, a speaking model structured on top of two underlying principles: A semantic/syntactic system on which one maps a conceptualized idea into a system of lexical items, and a phonological/phonetic system by which a speaker can orally transmit the conceptualized message. Levelt’s model recreates the processes of speech production through the workings of a series of autonomous components that make up the system; the Conceptualizer, the Formulator, and the Articulator.

The process begins with conceptual preparation, a stage during which the creation of a message comes out of internal reflection or interaction with an interlocutor. The message is generated through processes of macroplanning and microplanning. Macroplanning entails the development of the intention toward communication, taking into consideration important external and internal factors which influence the conceptualization of the message. Theory of mind, or social competence, one’s consciousness about beliefs, wishes and fears as well as understanding of intentions of one’s interlocutor and the ability to predict behavior takes on an important role during the macroplanning phase. Microplanning consists in determining how the message is to be conveyed in light of the situational context. Expressed otherwise, microplanning uses aspects as spatial positioning, whether a concept has been dealt with at some point previously, etc., to influence the linguistic structure of the utterance. Consequently, the mode of discourse must be decided upon for the message to be conveyed appropriately. An understanding of shared knowledge of the world will regulate what information can affectively be communicated and upon which a conversation may be built and become fruitful. A final negotiation between what can and cannot be expressed develops into a preverbal message upon which the final message will be formed.

Once the preverbal message is in place, the Formulator takes the conceptual structure of the message and gives it linguistic structure through grammatical encoding. This stage is the point at which lemmas are activated in the mental lexicon, the mental store of available vocabulary. The framework of the message is built up and the surface structure or the message is prepared. Upon formation of the surface structure the next substrate of the blueprint is accessed: The Phonological/phonetic system.

Morpho-phonological encoding ensues when a lemma is selected and, based on the way the way the expression has been encoded grammatically, is syllabified and intonation patters are established. The resulting ‘phonological score’ drives the phonetic encoding stage at which point the message is matched with appropriate articulatory gestures which are necessary for when articulation finally emerges.

Throughout the production process, monitoring allows the speaker to control self-generated speech either before or after utterance. A series of three monitor loops revert back to the monitor located in the Conceptualizer; the first loop allows comparison of the preverbal plan with the intended message before this message arrives to the Formulator stage to ensure that the plan corresponds to the information which must be conveyed as per the given context. The second loop controls the phonological score or internal speech during which the speaker is concerned with ensuring that the message is properly encoded before it is actually spoken. The final loop occurs after actual speech when the speaker is aurally aware of the final output of the process and becomes conscious of error in what has been spoken, calling for repair or other pragmatically determined action as deemed necessary.

3.4 Debot’s speech production model

Levelt’s model has served as a basis for the development of other models that have been created as well to illustrate the speech production process as pertains to speakers of more than one language including those by De Bot (1992, 2004) and Kormos (2006). The model proposed by De Bot (2004) incorporates a language node which regulates the different processing components in regards to the language which had been chosen during the conceptual stage. Upon conceptualization, the communicative intention and information about the language which is to be used is transmitted both to the system which generates lexical concepts, and to the language node. During following stages of the process, the language node regulates information drawn from language specific subsets of stores of conceptual features, syntactic procedures, and form elements. Once a particular language subset is activated, elements from that same language will be activated at the other levels as well. The language node will also regulate subsets which overlap, acting as a monitoring device to compare the language which has been intended for use with the language which is actually used.

3.5 Kormos' speech production model

Kormos' speech production model (Kormos 2006) is similar to that proposed by Levelt with several distinctions. Kormos identifies her model with the modularity of Levelt's model by the serial manner in which it functions along the various specialized processing modules. However, between the lexical and phonological level, she allows for cascading activation. Here, lexical selection and phonological encoding run in parallel. This results in competition between L1 and L2 lexemes in phonological encoding as the target lexeme as well as related lexemes, including those in a competing language, are also activated. Kormos' model includes knowledge stores positioned within long term memory. These stores of the four main memory systems include episodic memory, semantic memory, the syllabary in which automatized gestural scores for syllable production are stored, and a separate store for declarative knowledge of L2 syntactic and phonological rules. Where such rules for the L1 are automatized, rules for the L2 are not part of the encoding system and must be retrieved separately. The control of separation of language is done by a language cue tagged onto concepts during the conceptualization phase of the model. The cue is used then to match concepts with language appropriate information in the knowledge store to complete the speech process successfully.

3.6 The influence of task characteristics on L2 production

The goal of developing feasible sequencing criteria for tasks in the TBLT classroom has stimulated an area of study intent on identifying how individual task characteristics or combinations of them can be manipulated to obtain predictable results in language output. Investigations have contemplated models of the cognitive processes involved with speech production, the interaction between aspects of fluency, accuracy and complexity on language, and the manipulation of aspects of task complexity, task conditions and task difficulty, and the resulting effects on dimensions of speech.

Speaking is information processing subject to limited cognitive resources that are available to fulfill the needs of three primary characteristics that have been identified among linguists as dimensions by which to evaluate language performance: Accuracy, complexity, and fluency. While using an unfamiliar L2, there is a concern for speaking accurately to avoid confusion on part of interlocutor, and doing so with a sufficient degree of linguistic and structural complexity, as well as fluency to ensure that the message is conveyed with the proper nuance.

3.7 Skehan’s framework for task-based instruction

Skehan (1996) created a task sequencing framework. Within this framework, importance is place on the content of the message, and concern about content is attentional resource depleting. The system he created allows for tasks to be analyzed and compared, and then sequenced according to the amount of attentional resources that they would require. Properly sequenced tasks would balance fluency and accuracy, directing attentional resources to content or form while spare attentional capacity would allow for restructuring of language in use.

The scheme contrasts *code complexity* with *cognitive complexity* and *communicative stress*. *Code complexity* refers to syntactic and lexical difficulty, and range. Skehan refers to these as formal factors of language. *Cognitive complexity* refers to content of what is being communicated which Skehan links to the conceptualization stage of Levelt's (1989) speaking model. Cognitive complexity is further divided into *processing* and *familiarity*. Processing is how much on-line effort is needed in order to perform a task, and familiarity is the degree to which task demands can draw on automatized processes, or on information that is already available in memory. C*ommunicative stress* is made up of various factors that are not directly linked to code or meaning, but which influence communicative success nonetheless. Factors are *time pressure*, *modality*, *scale*, *stakes*, and *control*. Time pressure refers to the urgency of a task. Modality refers to how communication is performed; Skehan contrasts speaking with writing, and listening with reading with the former of each pair imposing more pressure than the latter. Scale refers to task-based teaching method factors such as the number of participants in a task, or the kinds of relationships involved win the task. Stakes refers to the importance of the outcome of the task. If stakes are low, it is because the consequences of a poor performance are few. But if consequences of poor performance are negative, then stakes are high. Control refers to whether or not the participant can control task performance by negotiating task goals or making clarification requests. More control implies less difficulty.

Research which resulted from the development of the sequencing scheme drove home an accounting of the influence that task characteristics exert on language production. Skehan (1996) proposed three aspects of language for measure; complexity, accuracy, and fluency (CAF) which come into competition with each other for attentional resources during task performance. Skehan and Foster (1999) define fluency as capacity to use language in real time, accuracy as the ability to avoid error during performance, and complexity as the capacity to use more advanced language. These have become commonly used measures whose variations may be compared either between them or with other independent variables. But understanding the interplay between these factors has spawned several theories and frameworks by which investigation into tasks and the influence of task characteristics on L2 performance.

3.8 Skehan's Trade-Off Hypothesis

Skehan(1998) and Skehan & Foster (2001) developed the Trade-off Hypothesis based on his earlier findings and especially in light of Level's model of speaking. It is based on the precept that people have a limited amount of attentional capacity, a concept borrowed from the view in psychology that limited capacity is a primary characteristic of attention (Broadbent, 1958; Kahneman, 1973). This view stipulates that attentional resources are located in a single store, or pool, within working memory from where its limited resources are allocated toward competing task demands. The degree of difficulty that a task entails is determined by its capacity consumption (Kahneman, 1973).

From a language processing perspective, limited attentional capacity leads to competition for attentional resources that are available, forcing them to be allocated between content and form of language during production. Under conditions where an increase in the cognitive complexity within a given task depletes any surplus of those resources, either content or form will benefit while the other loses quality. Lacking sufficient attentional resources to attend to both form and meaning, the latter tends to be prioritized to ensure that the intended message is properly conveyed. More cognitively complex tasks requiring resources to focus on message content will therefore draw attention away from language form resulting in a decrease in fluency, complexity and accuracy during language production when these are still in need of controlled processing as will be found during the acquisition of an L2. However, results in Skehan's research have indicated that this may not always be the case.

Skehan (2009) suggests that both accuracy and complexity may rise during oral task production, but that this is the result of task type or task manipulation, not because of how attention is intrinsically allocated by the executive system. Foster and Skehan (1999) investigated the effect of manipulating sources of planning (teacher-led, solitary, and group-based). Results showed that teacher-led planning generated greater accuracy, but not at the expense of complexity or fluency as predicted by the trade-off hypothesis. Solitary planning increased complexity, fluency and turn length while neither group-planning nor focus of planning on either content or language produced any effect on speech. But it was the teacher-led planning that resulted in an increase in measures of both accuracy and complexity at the same time. Skehan (2009) suggests that this is due to the effects of two kinds of planning that were involved in the task. One is rehearsal planning which leads to better more accurate speech. The other is complexification planning leading to more complex speech. The effects of the teacher led planning were cumulative, leading to simultaneous increases in both aspects.

Simultaneous increases in both accuracy and complexity were also reported in Tavakoli and Skehan (2005). In this case, a more tightly structured task led to greater fluency and accuracy while the fact that the need to carry over background information into the narration of the task increased complexity. These results were replicated in studies reported in Tavakoli and Foster (2008, 2011) and Foster and Tavakoli (2009) which used cartoon images which depicted tight and loose, as well as single or multiple storylines. Tight storylines promoted greater accuracy while multiple storylines produced more complexity. Both accuracy and complexity were increased in instances in which the cartoons depicted both tight and multiple storylines within a single narration. Ahmadian (2015) used two tasks with structured and unstructured storylines performed under pressured and careful on-line planning conditions. Results showed that participants who performed the structured and careful online task paradigm demonstrated increases in measures of more complex, accurate and fluent speech while those who performed the unstructured task under pressured planning conditions were those who obtained the lowest scores all around.

The consensus is that attentional resources are divided between focus on content or on form, but task design features can induce subjects to perform tasks using more fluent speech while simultaneously increasing accuracy and complexity of language.

In 2009, Skehan illustrated how certain task factors are linked with parts of Levelt’s speaking model to account for observations of speaker performance under specific task conditions (table 1).

Table 1. Levelt model linked to influences on L2 performance (Skehan, 2009)

Complexifying/Pressuring		Easing/Focusing
• Planning: extending • More complex cognitive operations • Abstract, dynamic information	Conceptualizer	• Concrete, static information • Less information • Less complex cognitive operations
• Need for less frequent lexis • Non-negotiability of task	Formulator: Lemma Retrieval	• Planning: organizing ideas • Dialogic
• Time pressure • Heavy input presence • Monologic	Formulator: Syntactic Encoding	• Planning: rehearing • Structured tasks • Dialogic • Post-task condition

Task characteristics listed in the column on the left side of the table will complexify the task, or put pressure on the speaker during performance. Those on the right will make the task easier, or focus attention on a specific area. At the Conceptualizer stage, factors in the column on the left are more demanding of working memory resources. Planning for extending or manipulating and transforming information are demanding cognitive processes and will result in more complex speech. This contrasts to the factors in the right hand column where a lesser need to manipulate or retrieve information will lessen the burden on cognitive resources, freeing them to be focused on other aspects of the language.

At the Formulator stage, factors in the column on the left will pressure a speaker during task performance. A need to access less common vocabulary and the pressure of being required to perform within non-negotiable limits force the learner to perform difficult functions which slow lemma retrieval as a result of limited capacity of the L2 speaker to maintain parallel processing functions. Factors influencing syntactic encoding include pressures exerted by limits in the amount of time allowed to perform a task, as well as the amount of input that must be handled. A monologic task also requires greater processing capacity as successful completion is dependent on the speaker alone.

In the right hand column at the lemma retrieval stage in the Formulator, the ability to plan and rehearse, what is to be said allows for lexical elements to be primed and ready for retrieval when they are needed. Dialogic operations also benefit lemma retrieval as lemmas again are primed by interaction between interlocutors as they support one another and share the task’s burdens. At the syntactic encoding level of the Formulator, planning allows for rehearsal of the message focusing resources on more accurate or complex form. Structured tasks allow for more time available as learners can avoid having to piece together a story before performance allowing resources to focus on other aspects of the message. Dialogic performance, in addition to benefiting the lemma retrieval stage of message formulation, also benefits the syntactic encoding stage by focusing attention on accuracy in order that a message is properly transmitted between interlocutors. This is similar as well to the influence of a post-task condition which pushes focus onto accuracy in order that the message will be properly conveyed.

In summary, complexifying is primarily linked to the Conceptualizer first and foremost. Structural and lexical complexity will be influenced. Pressuring, easing and focusing carry more relevance during the Formulator stage of speech production thereby influencing accuracy and fluency.

3.9 The Cognition Hypothesis

The Cognition Hypothesis’ creation was inspired by a need to discover a means by which to sequence tasks for optimal effectivity (Robinson 2001a, 2001b, 2003a, 2005). Researchers such as Candlin, Crookes, Long, Prabhu, and

Skehan, had proposed using tasks as a valid alternative unit to synthetic type syllabi which divided a language up into grammar based units each taught separately in a step by step fashion. The learner was intended to acquire the language through a gradual process of accumulation. Robinson looked toward studies in developmental psychology for support in developing an empirically sound rationale behind new sequencing options (Robinson 2005).

3.9.1 Development of the Cognition Hypothesis

Cromer's (1974) cognition hypothesis for first language acquisition claims that cognition is what determines language acquisition. Conceptual meanings are made available for expression as these concepts become evident to an individual while through various stages of development. Robinson draws both on Cromer's idea of first language acquisition where linguistic development occurs as notions become available and from Slobin's (1985) conclusions drawn from observations of parallels between adult and child language acquisition. While there are parallels, Slobin claims that the underlying factors to which the parallels are attributable must differ. He makes an observation which forms one of the premises behind the Cognition Hypothesis; adults, despite full cognitive development, retain a scale of conceptual complexity upon which the process of second language acquisition is based. But the two language learning processes are fundamentally different. Child learners map linguistic elements directly onto concepts drawn from their surroundings. In contrast, adult second language learners, who already have a clear understand of contextually driven concepts and who have a linguistic foundation previously mapped onto those concepts in their L1, must learn to remap those concepts into the L2 (Slobin, 1985). Turning to the scale of conceptual complexity that adults maintain, they will revert to simple constructions under simplified task conditions. Then under complex conditions, according to Perdue (1993), they are pushed from simple language variety to more complex forms in order to expresses themselves better (Robinson, 2003a).

That adults revert to a sense of complexity scaling in their approach to linguistic acquisition hints at the existence a natural order of sequencing for tasks along levels of cognitive complexity. Along these lines, the Cognition Hypothesis claims that it is possible simulate the ontogenetic process of language development by sequencing tasks in a language syllabus according to levels of cognitive demands that each task requires of the participants. In doing so, a sequence of tasks which follow a natural tendency by which language capacity is developed would provide an optimal context for student to make the form-function mappings needed to learn a second language (Robinson, 2003a).

Increased task complexity will induce more attention to language production as well as to how well input is processed. This should lead faster and more effective learning processes as learners attend to and notice input that is presented. Greater focus on form as well as on communicative content will direct allocation of attentional and memory resources to elements in complex tasks facilitating uptake resulting from noticing input and interaction from feedback. Another assumption by the Cognition Hypothesis is that adult L2 learners are different in their capacity of cognitive resources which affects the rate at which a second language is learned. This will result in differences in performance success, becoming more pronounced between learners as task demands increase. (Robinson, 2003a).

3.9.2 Claims of the Cognition Hypothesis

In summary, Robinson’s Cognition Hypothesis claims that increasing the cognitive demands of tasks along specific dimensions will (a) push learners toward greater accuracy and linguistic complexity in order to meet demands imposed by the task; (b) promote more interaction and negotiation during task performance, and will focus attention and noticing onto forms that are made salient in the input resulting in longer term retention; (c) there will be a greater effect of individual differences such as in working memory and attentional capacity, among others, on task performance and subsequent learning as tasks increase in their level of cognitive complexity (Robinson, 2005). Robinson proposes that the Cognition Hypothesis be used as basis for operationalizing task complexity in the form of a framework for syllabus design for task-based learning. In doing so, sequencing tasks from simple to complex could create optimal conditions for practicing language, speeding automaticity by presenting input in such a way as to facilitate the functions of executive processes. In addition, such a framework may be useful for designers of language tests as well as those in the research community who my need a means to calibrate data collecting materials according to complexity level.

3.10 The triadic componential framework

Robinson (2001a, 2003a, Robinson and Gilabert, 2007) proposes a triadic framework as an operational taxonomy of task characteristics for examining the implications of the Cognition Hypothesis toward L2 to answer a call for a theoretically motivated, means for applying empirically based findings about language learning to task-based approaches to syllabus design. Based on premises exposed in the Cognition Hypothesis he created the framework as a base from which to distinguish the relative complexity of tasks intended for syllabus and test designers, instructors, and researchers. In this framework he distinguishes between task complexity, task difficulty and task conditions, three groups of factors which interact together with an influence on task performance and learning.

3.10.1 Task difficulty

Task Difficulty entails learner factors, or characteristics of the learner which may induce a differentiation of the perception of difficulty of a task from one learner to another. This is due to the availability of cognitive resources which differs from person to person. Task difficulty characteristics will account for between-subject variation when two individuals performing the same task may perceive the difficulty of the task in different ways, perhaps because of differences between them in attentional and working memory capacities, or because one individual may have a greater aptitude for language learning than the other. These characteristics are divided into *affective variables* and *ability variables*.

3.10.1.1 Affective variables

Affective variables are changeable by nature. Resources available to these factors may change on a temporary basis and are susceptible to teaching methodologies. Levels of motivation or openness to experience may affect how a learner reacts to specific kinds of tasks making them important factors to take into consideration while making decisions about classroom teaching methods and means.

3.10.1.2 Ability variables

Ability variables are relatively permanent variables and can generally be diagnosed before learners are designated to a specific syllabus. Further research in addition to work described in the current dissertation, may provide insight into the interactive effects of some of these factors with task characteristics that will provide answers to how learners can benefit from improved syllabus design. It is important to view research with a clear understanding of whether cognitive abilities or characteristics such as different levels of working memory or attentional capacity will result in different levels of performance.

3.10.2 Task conditions

Task Conditions are characteristics that contribute to the demands that interactive factors impose on learners by affecting how information flows between participants during a task. These conditions are divided into *participation* actors and *participant* factors.

3.10.2.1 Participation factors

Participation factors include whether a task is monologic or dialogic, how participants are grouped during a task, or whether the solution is open or convergent. These factors determine the intrinsic design of the task and will affect strategies used by the participants as they perform.

3.10.2.2 Participant factors

Participant factors are those which affect demands imposed by aspects of the participants involved. So, as an example, whether or not participants share the same level of proficiency may affect the perceived difficulty of the task. These are important considerations in task planning as they may affect perceptions of a learner’s role or status during in a learning activity with may influence the degree of participation during interactive tasks (Robinson, 2001b).

3.10.3 Task complexity

Task Complexity includes cognitive factors of a task which can be manipulated to increase the cognitive demands that a task makes on learners during performance (Robinson 2001a; 2001b). According to Robinson (2003a), of the three factors, task difficulty, task condition, and task complexity, the latter factor is that which is most appropriate for task sequencing considerations in syllabus design. Robinson (2003a) argues that conditions of task difficulty, as they are dependent on individual factors of participants, are difficult to control and may be affected by varying task conditions. Task conditions, although controllable, may best be determined a priori according to the needs of the particular situation and held constant while cognitive complexity is increased along factors of task complexity. Robinson distinguishes between attentional resource dispersing and attentional resource directing dimensions of task complexity (Robinson 2003a). The division depends on how specific task characteristics affect the focus of attentional resources.

3.10.3.1 Resource dispersing dimensions

Resource dispersing dimensions of complexity differentiate between task characteristics which create performative or procedural demands such as allowing or not for planning time or providing or not previous knowledge about a task situation. Although these factors place demands on attentional and memory resources, they do not direct these resources to any particular area of the language production system. Manipulation of these variables disperses resources, simulating real-world situations in which a speaker must perform under varying circumstances such as handling new or unexpected circumstances that would have to be reacted to spontaneously, and thereby promoting access to and control of already established interlanguage knowledge within an existing L2 knowledge base.

3.10.3.2 Resource directing dimensions

The resource directing dimension of complexity differentiates between task characteristics in terms of conceptual or linguistic demands. As these characteristics are increased in level of complexity, there is a potential to link cognitive resources, such as attention and memory, to effort at the conceptualization stage of Levelt's speaking model (Levelt, 1989) so that the concept may be well created and proper elements of the target language will be primed for formulation. Such demands may be met through specific aspects of the linguistic system such as by distinguishing between necessity to refer to the past or present verb form to compensate for the state of temporality imposed through manipulation of the +/- here and now variable of a task. These demands may also be met through the use of subordination when a speaker must justify actions or support reasons for interpreting a situation in a particular way when task demands are increased along dimensions of reasoning. Manipulating resource directing dimensions of cognitive complexity within a task, directs learners' attentional and memory resources toward the aspects of the language production system, promoting greater syntacticization and grammaticization so that the message is more efficiently expressed. Manipulation of the complexity of the task along resource directing variables also promotes development of an L2 during task performance by extending the L2 repertoire and increasing the demands of the conceptual or linguistic requirements for expression of spatial location, temporality or causality as the learners meet gaps in their knowledge that they are pushed to fill in order to complete the task.

3.10.4 Predictions of the Cognition Hypothesis

Predictions of the Cognition Hypothesis for the effects of changes in tasks complexity on language performance and learning are based on a multiple resource view of attention proposed in Wickens' (1989) model of dual task performance which was founded heavily on multiple resource theory.

Multiple resource theory was born out of study of attention as it is related to the performance of complex tasks, although outside of the area of SLA research. The theory was originally intended to address practical solutions to operators working in high work load environments in industry, aviation navigation, or other fields where greater efficiency in task performance procedures could lead to fewer operator induced errors. The risk of errors could be minimized by means of a multiple resource model capable of predicting performance based on changes in task design (Wickens 2002). Designers would be provided with a tool that would allow them to manipulate tasks toward a predicted performance outcome.

Robinson (1995a) observed that up to time of his writing, such understandings of the allocation of attentional resources as applied to other areas of cognitive psychology were not commonly referred to by SLA researchers.

Wicken's model advocates that attentional resources are drawn from multiple pools. This is in contrast to Kahneman's model (Kahneman, 1973) which establishes attentional resources as coming from a single pool. Wickens suggests that such multiple pools provide attentional resources to varying classes of activity that require them in order to be carried out. Any number of tasks will share resources from a single pool while others will require resources from a different pool. As long as two tasks being performed simultaneously draw on distinct pools, neither will interfere with the other. However, in the event that activities sharing resources from a single pool are carried out simultaneously, there will be competition for the limited resources. These must be allocated by a central executive in order that the demands of the tasks at hand may be carried out as efficiently as possible. If one task demands more resources, this will lead to poorer performance of the other task for lack of resources that have been allocated to the other.

3.10.4.1 Predictions along resource-directing dimensions

Along resource-directing task dimensions, the Cognition Hypothesis predicts that as tasks increase in their level of cognitive complexity, both accuracy and linguistic complexity will benefit, but at the detriment of fluency. The increase in task demands along these dimensions will direct attention to forms needed to meet specific aspects of the target language code so that demands are met. These aspects of the target language may or may not be known. In the event that the learner does not know these forms, the gap will become noticed and a change in the interlanguage will be available provoking learning.

3.10.4.2 Predictions along resource-dispersing dimensions

Along resource-dispersing task dimensions, the prediction is that as tasks increase in cognitive complexity, accuracy and linguistic complexity will decrease. By increasing task demands along these dimensions, attention is not directed to any aspect of the target language, rather it is dispersed along other dimensions including linguistic or other features in order to ensure that the task is performed properly. Practice with situations in which resource-dispersing dimensions are increased during task performance should promote faster and more automatic access to the target language (Robinson, 2011).

Lastly, the Cognition Hypothesis also predicts *synergetic* effects on speech production. When tasks are complexified along both resource-directing and resource-dispersing dimensions as in real world situations, the benefits that might be observed because of increased resource-directing aspects of the task may be offset by increases in resource-dispersing aspects (Robinson and Gilabert, 2007).

3.11 Summary

This chapter reviewed speech production from an information processing perspective and described various related models of speech production. Two principal hypotheses which address issues in SLA concerned with cognitive processes involved in speech production and language acquisition were also described as well as frameworks for task sequencing which draw from the theory behind these hypotheses. The following chapter will discuss individual differences and how they affect research in the area of SLA.

CHAPTER 4: MEMORY, ATTENTION, AND INDIVIDUAL DIFFERENCES

4.1 Introduction

In the last chapter, speech production models as well as hypotheses toward language acquisition were discussed. In addition, frameworks were presented which are built on the precepts behind the hypotheses. Coming from an information processing perspective of SLA theory, the hypotheses and frameworks discussed, and especially research which concerns them, rely heavily on an understanding of the way that individuals may differ in cognitive resource capacity and the affect that it will have on language output.

Individual differences are traits or characteristics which are assumed to be common to all people but which vary among them in the degree with which the trait may distinguish one individual from another. In learning contexts, identification and understanding of individual differences play an important role as they impact how each learner filters or focuses information affecting the extent that they assimilate information. In addition, awareness of individual differences increases the sensitivity of instructors towards learners’ needs as they affect aptitudes for learning, willingness to learn, or preferences for styles of learning. An understanding of personality traits and characteristics which may reflect learning abilities is an essential tool both to the educator as well as to the investigator.

This chapter will focus on aspects of memory and attention and their role in L2 acquisition. Both of these constructs form a principle part of this dissertation. Individual differences between these two learner characteristics will be taken into consideration as linguistic performance by individuals who differ in these aspects are analyzed.

4.2 Memory

Memory, as it is understood today, is a dynamic, ongoing process of reclassification of information, the result of constant changes in the brain's neural networks and parallel processing activity. Apart from being discussed in general terms, memory is commonly divided into several sub-categories: Sensory memory, short-term memory, working memory, and long-term memory (Engle, 2002; Engle, *et al* 1999; Kane, et al 2001; Kane and Engle, 2003).

Sensory memory is the arousal of sensory organs by stimuli. This arousal lasts very little time and further processing depends on whether or not the stimuli is eventually attended to. Short term memory allows for storage of a limited number of items for a limited period of time and long term memory is considered to be that in which memory of both recent and older events have become consolidated and available for retrieval as required. As concerns working memory, a universal definition has remained a bit elusive, although researchers have offered a wealth of descriptions of its functions. Nevertheless, there seems to be some consensus that, in broad terms, working memory acts as an interface between perception, long and short-term memory, and actions directed toward the achievement of task related goals.

Learning takes place when items are encoded in memory by altering neural networks in the brain which are made up of a series of neurons communicated chemically or electrically by junctions known as synapses. During learning processes, information is stored in long term memory through a process of Long-Term Potentiation whereby synapses joining the activated neurons increase in efficiency facilitating the passage of the neural message along the circuit. This process establishes new neural networks and strengthens associations with other stored items ultimately forming the memory of the newly learned item. The efficiency of synaptic transmission will determine how affectively information is stored in long-term memory, while the frequency or the intensity with which neural networks are activated establishes the stability of the synaptic contacts which make up the memory item. More recent information will be fragile while older information will become crystallized, or consolidated, although not immune from being forgotten (Bliss and Lømo, 1973; Lømo, 2003, Lynch 2004).

Retrieval of information previously stored in long-term memory uses encoding indexes where stimuli activate associated memories. This process is divided into two sub-processes known as recall and recognition. Recall is considered more demanding of resources as it consists in activating all associated neurons while reconstructing information. Recognition, on the other hand, requires a simple decision as to whether a particular item, among others, has been encountered at some point previously, necessitating only partial activation of a network.

Access to information in short-term memory is carried out by retrieval operations using neural machinery similar to that used in long-term memory according to findings by Nee and Jonides (2008). In fact, short-term memory is commonly regarded as a temporarily active portion of long-term memory, the principal difference being duration and capacity of the memory stores (Cowan, 2008), having been demonstrated that information in short-term memory stores appears to decay with time and that capacity is limited; Miller (1956) instituted the number seven as the average number of items or chunks of information that can be stored.

4.2.1 Working memory

Working memory, central to the study of the cognitive underpinnings behind language acquisition and production processes, is defined by Engle e*t al.* (1999) as a system consisting of those long-term memory traces active above threshold, the procedures and skills necessary to both achieve and maintain that activation, and finally, controlled attention. Dehn (2008), on the other hand, suggests that working memory is a distinct and independent memory store responsible for the manipulation, management and transformation of information that is taken either from short or long-term memory. Finally, Cowan (1993: 166) describes working memory as: the "interface between everything we know and everything we perceive or do"; it is the mechanism by which information is stored and retrieved.

Nonetheless, a short and simple definition of working memory may not do it complete justice as precisely what it is and how it should be defined has been a topic of debated in the fields of cognitive psychology and cognitive neuroscience. Miyake and Shah (1999) compiled a series of opinions on the subject from nine different teams of researchers by posing each with the same set of questions concerning working memory. Their intention was to identify at which points researchers' hypotheses converge and where they show the greatest deviation. In the end, they composed what they assert to be a definition generated out of six points of general consensus among the researchers interviewed for their work. These points are first, that working memory is not a structurally separate box or place in the mind or brain. Possible misinterpretations of models may have given way to traditional concepts of short-term memory or working memory as being a specific ‘place’ or ‘box’ where memory is stored. This view, however, is not postulated by the researchers who contributed to the survey. Secondly, working memory’s maintenance function is in the service of complex cognition. A strong consensus among the researchers is that the role of working memory is not for memorizing specific items in and of themselves, but works in the service of complex cognitive activities such as the processing of language, visuospatial thinking, decision making, and reasoning and problem solving. Thirdly, a completely unitary, domain-general view of working memory does not hold. As a result, working memory capacity and performance, as a sum of its parts, is dependent on the capacity and performance limitations of it parts. So, for example, attention, which is a construct of working memory identified as a focused subset of the information within the activated neural networks which make it up (Nee & Jonides, 2008), is limited in its capacity to be controlled and sustained under stress when there is interference or distraction. Consequently, the degree of affectivity with which working memory can function in the storage and retrieval of information is dependent on these limitations as well. Nevertheless, and leading to the fourth point of consensus identified by Miyake and Shah, capacity limits reflect multiple factors and may even be an emergent property of the cognitive system. Working memory, consisting of an array of constituent processes, finds its capacity limited, not as a result of limitations of any one of its constituents, rather as the result of any variety of factors that are involved in its functions. Fifthly, executive control, a system by which constituent parts or processes are intercommunicated, is integral to working memory functions playing a key part in the control and regulation of cognitive activity. Finally, long-term knowledge plays an integral role in working memory performance. The definition that Miyake and Shah (1999:450) propose is the following:

> "Working memory is those mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition, including novel as well as familiar, skilled tasks. It consists of a set of processes and mechanisms and is not a fixed "place" or "box" in the cognitive architecture. It is not a completely unitary system in the sense that it involves multiple representational codes and/or different subsystems. Its capacity limits reflect multiple factors and may even be and emergent property of the multiple processes and mechanisms involved. Working memory is closely linked to long term memory, and its contents consist primarily of currently activated long term memory representation, but can also extend to long term memory representations that are closely linked to activated retrieval cues and, hence, can be quickly reactivated."

4.2.2 Baddely's model of working memory

A widely cited model of working memory created by Baddeley and Hitch (1974) and Baddeley (1986) is based on a multicomponent memory system distinct from that of short-term memory storage (figure 1). Previously, the role of working-memory was generally attributed to short-term memory. Nevertheless, questions arose from situations in which evidence from aphasiac patients with damaged short-term memory, understood to perform a crucial working-memory function in the performance of complex tasks, did not demonstrate that the damage had any affect on tasks identified as such (Baddeley and Hitch, 1974).

The researchers chose to investigate whether tasks of reasoning, comprehension, and learning all shared a common working-memory mechanism, and to understand what relationship existed between this mechanism and short-term storage. Their research resulted in the development of a model intended to provide a solid, empirically substantiated scheme for the working-memory hypothesis. The original model consisted of an attentional controller referred to as the central executive aided by two slave systems; the visuo-spatial sketchpad and the phonological loop. These were to work together to form a unified working memory useful for the performance of complex tasks.

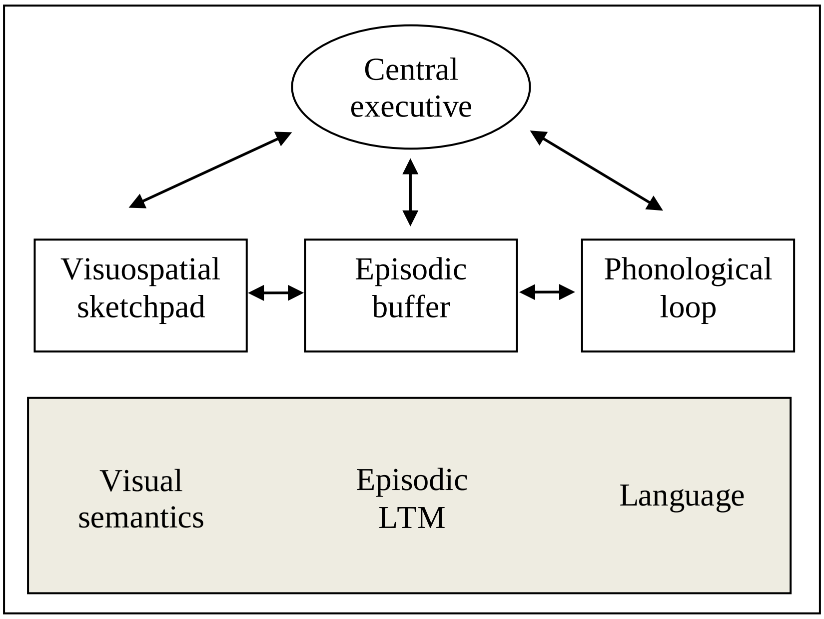
4.2.2.1 The phonological loop

The phonological loop is that part of memory which is concerned with aurally received stimuli or with stimuli which produces vocal or sub-vocal speech. It consists of two sub-components, the first of which acts as temporary store capable of maintaining memory items for a few seconds until refreshed by the second sub-component which involves a sub-vocal rehearsal system. Accordingly, the phonological loop is not used solely for storing aurally received data, but for storing visual stimulus as well, as long as the visual item can be named and the name rehearsed sub-vocally.

4.2.2.2 The visuo-spatial sketchpad

The visuo-spatial sketchpad is described as a memory store for integrating spatial, visual, and possibly kinesthetic, or motor, information into a unified representation. This includes information received through any form of stimuli, so long as it can be represented visually, spatially, or, again, possibly kinesthetically.

Figure 1: Multicomponent model of working memory (Baddely, 2003)



4.2.2.3 The central executive

The central executive, the operation of which is referred to as executive processes, is responsible for the attentional control of working memory. Attentional control is the means whereby the flow of information through the system is directed, maintained according to the task at hand, or by which task-irrelevant information is suppressed (Baddely, 2003; Conway and Engle, 1994). During speech production, it is the executive processes which direct attentional resources, a limited reserve of energy allocated for mental effort, toward greater focus on either fluency, grammatical accuracy, or linguistic complexity depending on the specific needs of a particular task.

4.2.2.4 The episodic buffer

The episodic buffer was added to the model later in order to answer for how information from a single stimulus or event that is stored in the separate sub-systems of working memory could be integrated into a single representation. It also provides for a mechanism by which this integration of information could involve access to long-term memory (Baddeley 2000). Although separate from the central executive, access to the episodic buffer between subsystems and long term memory happens through the central executive. It is a temporary store and is limited in capacity. As per Miller's (1956) hypothesis, it can hold seven items, or chunks, on average (Baddeley & Hitch, 2006).

4.3 Attention

As a construct of working memory, attention is a cognitive resource that aids in the flow of information through the executive system. The roles that attention plays are multiple as it is responsible for the conscious control of mechanisms which are not fully automated, as in the case of L2 speech processing (de Bot, 1992). By various accounts, it is considered to play a vital role in the learning process since it is necessary for attending to input so that the input can be moved into memory storage allowing for hypothesis forming and testing by the learner (Schmidt, 2001; Ellis, 2001; Doughty, 2001, Segalowitz & Frenkiel-Fishman, 2005). In the case of L2 learners, attention is what pushes them to become aware of the gaps in their interlanguage while they identify differences between the language that they produce and what other proficient speakers produce. Attentional agility also allows a learner to distinguish between stimuli that are relevant to a particular context, and which are not relevant resulting in the inhibition of reactions which would otherwise be inappropriate under the circumstances at hand.

Attention is described as a multicomponential faculty composed of various distinguishable elements (Eviatar, 1998; Tomlin &Villa, 1994). One of these elements, an orienting/focusing element, is the process by which the body reacts reflexively to focus on external stimuli. A subsystem of orientation, alertness attends selectively to stimuli so that high priority information is filtered for further processing according to current needs. A higher state of alertness increases the rate at which stimuli can be attended to, albeit to the detriment of the quality of the information attended, which may result in poor task performance (Posner & Peterson, 1990). Secondly, there is a detecting/encoding element by which stimuli is registered cognitively. The stimuli may be registered with or without awareness. Registration without awareness is when attentional focus is not centered on the incoming stimuli which remains peripheral and stored briefly in short term memory where it may either dissipate or move into long term memory. Conversely, registration may be made with awareness by which the stimuli are the focus of attention. Schmidt (1990, 1993, 2001) refers to the latter as noticing, or conscious perception, a crucial concept in the area of second language acquisition studies. A stimulus which is noticed is pushed into long term memory through the rehearsal processes of working memory.

Attention therefore plays an important role in the language acquisition process since once input is attended to and selected, then awareness, or noticing, takes place, as per Schmidt (1990, 1993, 2001). According to Schmidt’s noticing hypothesis what learners notice in input is what becomes intake for learning, indistinctively of whether that noticing is done deliberately or unintentionally. Noticing, according to Schmidt, is a necessary condition for the acquisition of a second language so that it can then be rehearsed, in the case of language, in the phonological loop according to Robinson (2005).Noticing is promoted when awareness of language features is strengthened by attention being focused on them through instruction (Skehan, 1998).

Attentional capacity is limited according to the currently accepted view in psychology. For reason of economization, attention is also selective. Control of attention is the process by which attentional resources can be directed to specific aspects of input. This faculty is critical when input is multiple, simultaneous and must be prioritized. Effective and efficient inhibition of irrelevant stimuli prevents attentional focus from being drawn away from the principle task demands. By excluding irrelevant information, attentional resources, and thereby, working memory resources are economized freeing up those resources for further processing functions (Bialystok & Martin 2004; Kane & Engle, 2003). The general current view is that there are tasks which are attention-demanding and require attentional resource depleting effort to perform. Correspondingly, there are tasks which are automatic and require fewer attentional resources to be performed.

4.3.1 Kahneman's model of attention

In 1973, Kahneman's capacity model of attention described how a limited amount of available attentional resources originating from a single resource pool are allocated to various stimuli. The distribution of these resources depends on the state of arousal of a person, enduring dispositions, or demands required by automatic processes, as well as current needs required by the present task. As long as arousal is sufficient and task demands do not exceed the amount of attentional resources that are available, task performance can be carried out without negative effects on the various aspects which define the quality of the performance. But once available attentional resources begin to run low, selective processes must prioritize stimuli to be attended to according to the demands of the task. The result would be a trade-off of allocated resources between various aspects of performance where some of those aspects would benefit from prioritization while others would suffer.

4.3.2 Wickens' model of dual task performance

Wickens' model of dual task performance (Wickens 1989, 1992, 2007), suggests that attentional resources flow out of multiple resource pools, in contrast to Kahneman's single source. So, two activities that are not very similar in nature but which are carried out simultaneously, as in the case of a person driving a car while talking to a passenger, will not interfere with one another as resources needed to perform both activities will be drawn out of several different attentional pools. Nonetheless, since the overall store of attention will be lessened as it is dispensed to both activities, poor performance on one or both activities may result, determined in part by the attentional capacity of the individual. On the other hand, when various tasks which are performed simultaneously draw on the same pool of attentional resources, their relative difficulty increases and it becomes impossible to carry them out. A person is then forced to handle each task one after the other. As an example, the case of a person trying to take part in two conversations at once is a task requiring a degree of attention that an ordinary person would find extremely difficult to manage. In contrast, when one of the various tasks is automatized, then both may be performed simultaneously without interference of one on the other. Wickens stakes the claim that individual differences come into play concerning ability to perform two tasks simultaneously because of the differences in capacity of attentional resources that each individual possesses.

Cowan (2008) observes that the efficiency of the attentional system and its use in working memory seem to differ substantially across individuals. One's working memory capacity is primarily a matter of ability for that individual to control attention in order to keep information in an active and quickly retrievable state, a reflection of the efficiency of synaptic transmissions and capability to economize memory and attentional resources allowing for more or less efficient cognitive activity (Engle, 2002). Those who are better habilitated to control attention should be able to allocate it more efficiently under stress.

4.3.3 Attention switching

Goal related activities which make up real life tasks performed every day are rife with a variety of activities. Each of these demand attention whose focus must switch as choices are imposed and decisions must be made. With each decision, one executes executive control to ensure that the decision taken is the best to achieve the goals of the moment. At the same time one must resist any temptation to perform another task which, while achieving a different goal, will result in a poorer performance of the goal at hand.

There are many tasks which through experience or instruction have been acquired and are stored in memory. If that task is common place, or practised, or if it is a recently acquired and practised task, it is easier for us to re-enact that task when need calls for it. But one may also find themselves performing a task under a circumstance where stimuli that we associate with that task is perceived, even though one's intentions are not to do so.

Endogenous control is the ability to control attentional focus and to allocate attentional resources to aspects of a task at hand to ensure its successful completion. It is internally driven by what Baddeley (1986) calls the central executive and what Norman and Shallice (1986) call the supervisory attention system. This is what allows the capacity to anticipate and prepare for a switch to a new task paradigm even before the switch occurs based on previous experience and practice.

In order to effectively execute endogenous control, exogenous control must be overcome. Exogenous control is exerted by stimuli that distract attention from the task at hand provoking a behavior habitually associated with those stimuli, even if that behavior differs from the individual’s intended behavior. To illustrate this concept, Rogers &Monsell (1995) use the example of a man who goes to his bedroom intending to dress for dinner, but soon finds himself lying in bed with his pajamas on. The man demonstrates a loss of endogenous control over his behavior by failing to complete his intended task which was to dress for dinner. Instead, the exogenous control exerted by the familiar environment of the bedroom dominated and he ended up getting dressed for bed, which is how he normally dresses when he goes in there.

Of two simultaneously activated task paradigms and their corresponding sets of rules driving behavior, one may dominate over the other (Norman & Shallice, 1986). This depends on the strengths of association between the competing paradigms and the stimuli available in the environment, or any remnant of performance rules stored in memory which determine a specific kind of behavior according to the circumstances. When a dominant paradigm of a current task must be inhibited in order to switch behavior to comply with new task demands, endogenous control of attentional resources, driven by the executive system, allocates resources to effectively adjust activation levels of the competing sets of rules so that the individual’s behavior corresponds to task demands.

In the current experiment, external stimuli in the form of pictures which the subjects use to narrate their stories as well as internal stimuli drawn from long term memory store of what the subject already knows of what is depicted in the images are all be weighed against one another in order to arrive at a successful completion of the task. Narrators are required to a greater or lesser degree to modulate between exogenous and endogenous control of their creative narrative behavior as they interpret the depicted situations and conceptualize, formulate, and finally tell their stories. The more efficiently and effectively the narrator is able to switch attentional focus among various internal and external stimuli, as well as balance resources between exogenous and endogenous control, the fewer the attentional resources that will be needed to complete the task successfully. As more cognitive resources are available during task performance, the threshold at which trade-offs between measures of performance become observable as tasks increase in complexity can be expected to rise.

In language, where attentional focus forms part of the cognitive processes which are involved in speech, task switching is an integral function of bilingual ability. The capacity to do so efficiently and effectively has an effect on linguistic performance (Luo *et al*, 2010; Wickens, 2007; Weissberger *et al*, 2015; Robinson, 2003b).One might speculate that a more efficient allocation of resources would presuppose a higher limitation threshold of attentional resources. Greater attentional capacity would correspond to greater resistance against suffering trade-offs between aspects of linguistic complexity, accuracy and fluency during language task performance. However, once the limitation threshold is crossed and resources are forced to be allocated to meet the needs of the task, where measures of working memory capacity imply optimum cognitive processing capabilities, greater efficiency in allocation of those resources along aspects of CAF could be expected and measures depicting that allocation should be more dynamic than in cases where working memory capacity is lower. This forms the basis of my hypotheses which will be explained in the chapter on methodology later in the dissertation.

4.4 Other key individual differences

Dörnyei (2006) surveyed literature in the field of study of individual differences. He came up with a list of five learner characteristics which he considers most important to SLA research: Personality, language aptitude, motivation, learning styles, and learning strategies.

4.4.1 Personality

Personality, according to Dörnyei's research, is one of the factors which plays a lesser role in determining success at language learning. The research into personality traits seems to have inconclusive results. One reason for this, he suggests, may be in the way that personality taxonomies have been used in research, and that differences in how traits are understood may be at fault in the way that researchers are approaching the issue. Within the field of second language acquisition, the most frequently researched personality traits are extroversion and introversion but that as of the writing, no clear results have come out of the investigation.

4.4.2 Language aptitude

Language aptitude is understood as the capacity of an individual to easily pick up a new language. Of course, the ability of one to learn a language with a relative degree of ease is the focus of much SLA research. Language aptitude is a construct of multiple components each of which are considered individual differences in their own right. HI-LAB, as a matter of example, is a composite test battery designed by the Center for Advanced Study of Language at the University of Maryland intended to diagnose candidates who exhibited language learning aptitude (Doughty, *et al*, 2010). Motivated by experts in second language acquisition, the test taps into eight different constructs which the authors say underlie high level language aptitude. The constructs include memory, perceptual acuity, speed, primability, induction, pragmatic sensitivity, and fluency.

The construct of memory has already been discussed in detail above. Acuity refers to perceptual acuity, or the ability to see or hear cues that are presented during testing. Speed refers to processing speed, or how quickly one responds to stimuli. Primability is the degree to which stimuli presented to an individual prompts subsequent processing. Induction is the capacity to draw conclusions based on presented patterns which may be presented either implicitly (implicit induction) or explicitly (explicit induction). Pragmatic sensitivity is the ability to make connections between contextual cues and detecting errors. Fluency measures automaticity in planning and producing speech.

The end of the matter is that language aptitude in and of itself is a composition of various cognitive factors which can be used jointly to define or predict an ability to learn language with a greater or lesser degree of facility. Dörnyei (2006) even questions the usefulness of a the term ‘language aptitude’, but that standard measures of the construct continue to function with a relatively good degree of success and the term remains in general use.

4.4.3 Motivation

Research into language learning motivation focuses on various dimensions. One’s disposition to an L2 group and attraction to that community, or how language learning experience affects a person’s disposition toward an L2 are both factors of motivation. Motivation is sensitive to time as it changes with time, passing through peaks and valleys. The way one views oneself as a language learner has also been a fairly recent dimension of study. This focus is on the way a learner perceives themselves as an ideal L2 learner or how they perceive the way that they ought to be in terms of their L2.

4.4.4 Language learning styles

Everyone has a preference to how they approach learning. These are often the product of early learning experiences which condition behavior. A style which gave positive results earlier on is a style one tends to maintain. Learning styles are important to language acquisition not only in that the learner must find the most efficient and effective way to acquire and L2, but an understanding of different styles of learning will be an advantage to an instructor who can adapt their own teaching styles to accommodate the way their students learn.

However, as a subject of study, according to Dornyei (2006), measures and means have been the object of some controversy, and results of formal study have not been especially rewarding in a general sense. Nevertheless, there has been some extensive work done in the past from which several theories or models have emerged (see Wong, Dubey-Jhaveri, and Wong, 2015 for a review).

4.4.5 Language learning strategies

There is no doubt as to what a strategy is, nor if one speaks of a language learning strategy is any doubt assumed as to what is being referred to. But according to Dörnyei (2006) few good definitions have arisen in concerned literature. Hardan (2013) provides an interesting overview of research with a summary of taxonomies of various researchers who have identified trends in strategies. As a general rule, strategies tend to be dichotomies of direct and indirect nature. The direct strategies refer to cognitive strategies which imply specific learning tasks and conscious study of material. Indirect strategies employ meta-cognitive strategies which employ planning processes, goal setting, self-management and general logistic issues which allow a learning process to develop. Indirect strategies may also include social strategies where interaction is involved.

4.4.6 Other factors

Researchers take other factors into consideration in addition to those listed above. Two of these which are applicable to the current investigation because of potential effects on results are briefly discussed below.

4.4.6.1 Age

It is generally understood that under normal conditions everyone will attain native L1 speech during their development from childhood to adulthood. Adults, however, are much less likely to acquire a native like levels of an L2 after beginning to learn a new language. The Critical Period Hypothesis (CPH) provides an explanation for why this happens to adults. The CPH proposes that there is a limited period of time during which a person is able to acquire languages and achieve native levels of proficiency. Once that period of time has passed, the potential to develop native like proficiency decreases due to the loss of neural plasticity which allowed the brain to develop along with input from the environment. Once cognitive maturity sets in, language learning becomes a process of unlearning parameters set by the L1 and developing a new paradigm within which both the L1 and subsequently learned languages are able to coexist (Birdsong, 1999).

Age alone is not the only determining factor. The environment in which young people acquire a second language is important. In a naturalistic setting where a child is exposed to a second language for long periods of time immersed in the target language there is a notable advantage for acquisition. However, under conditions where exposure to the target language is limited, then age does not offer such an advantage. In fact, results from the Barcelona Age Factor Project suggest that older learners' greater cognitive development serves as an advantage over age in that their rate of acquisition early on in the language learning process may result in an equal degree of acquisition compared with learners who began language studies earlier in instructed language settings (Muñoz, 2010). Age of acquisition is a good predictor of ultimate attainment of a second language, and age of initial learning at school is a common variable considered in studies that are concerned with the effect of age on language learning. But Muñoz argues that where language exposure has been under limitations, the range of time that language learning takes place within a language learning setting is a more effective measure as this reflects the entire learning process more effectively (Muñoz, 2010). Muñoz (2014) also found that there were other factors such as informal contact with native speakers as well as time spent in immersion study abroad that were better predictors of learner's L2 oral performance than starting age.

4.4.6.2 Proficiency

Proficiency plays an important role as it is what determines the degree of automaticity a speaker enjoys. Greater automaticity will free up attentional resources which can be made available to the most important aspects of speech in order for the most appropriate message to be conveyed according to task demands. A person with higher proficiency will be able to cope with situations better or more efficiently than someone with a lower proficiency level when task demands increase because of the greater amount of attentional resources available to deal with the task (Declerk and Kormos, 2012; Gilabert and Muñoz, 2010; Gilabert, 2007; Kormos, 2000). The availability or lack of attentional resources is what is expected to determine how changes in task complexity are reflected in observed patterns in speech once the results produced by the current investigation are analyzed. This will be discussed in a later section of this dissertation.

4.5 Summary

This chapter summarized current understandings of the functions of memory and attention and their relationship with language acquisition. Other kinds of individual differences were also briefly described. The next chapter will present a brief review of motivational research behind the present dissertation. Research questions and hypotheses based on doubts which have not been fully addressed as of yet in current literature will be posited. The questions stem from doubts about how individual differences in attentional and working memory capacity may play a role in variations of output as individuals perform a series of tasks which increase in their level of cognitive complexity. The experimental process and the underlying methodology will be explained in detail.

CHAPTER 5: EMPIRICAL RESEARCH ON TRADE-OFF EFFECTS

5.1 Introduction

An increasing interest in the task as the means of choice for developing TBLT syllabi and the need for empirically sound rationale for determining task sequence within syllabi has drawn the attention of researchers in the field of SLA. Much of the focus of research has been on the effects of tasks on language production as tasks are manipulated to modify the level of conceptual complexity of the demands that they impose on learners. The ability to predict effects of modifications would give task and syllabus designers an invaluable tool to create efficient and effective language courses. Results of research has centered on the two hypothesis described earlier in this dissertation which form the substrate of the models currently being used make such predictions. These are Robinson's Cognition Hypothesis and Skehan's Trade-Off Hypothesis.

5.2 Research on effects of task manipulation

At present, a clear consensus favoring either the Cognition Hypothesis or the Trade-Off Hypothesis has not been reached as results of investigation have varied. It is, to date, unclear how manipulating tasks along their levels of complexity affect output as measured by dimensions of CAF. It is also unclear how to predict resulting trade-offs between these dimension as attentional resources are allocated to meet task demands. Some prominent studies which have studied trade-off effects between measures of CAF are listed in table 2 according to the cognitive dimensions which were investigated. The list is not exhaustive, and while most of the studies in the table deal with oral production, as in the current study, some do not. Ishikawa (2007), Kuiken & Vedder (2008), Kuiken & Vedder (2007), and Kuiken, Mos & Vedder (2005) all deal with written production. Although these studies differ from the current study in this aspect, they are, nonetheless, representative of current investigation that has contemplated models of the cognitive processes involved with language production, the interaction between measures of linguistic performance, and the manipulation of variables that determine task complexity, task conditions and task difficulty. Studies are described briefly with respect to the resulting trade-off effects observed by the investigators.

Table 2 Studies in SLA on the effects of task demands on measures of CAF

+/- Planning Time	+/- Here and Now	+/- Reasoning Demands	+/- Few Elements	+/- Prior Knowledge	+/- Task Stucture
• Gilabert (2005)	• Gilabert (2007)	• Révész (2011)	• Kuiken & Vedder (2008)	• Robinson (2001a)	• Tavakoli & Skehan (2005)
• Yuan & Ellis (2003)	• Ishikawa (2007)	• Gilabert (2007)	• Gilabert (2007)	• Bygate, et al (2001)	• Tavakoli & Foster (2008, 2011)
• Menhert (1998)	• Gilabert (2005)	• Robinson (2001b, 2007)	• Michel, Kuiken, & Vedder (2007)	• Tavakoli & Foster (2008)	• Ahmadian (2015)
• Ortega (1999)	• Iwashita (2001)	• Nuevo (2006)	• Kuiken & Vedder (2007)		
• Skehan & Foster (1997)	• Rahimpour (1997)	• Niwa (2000)	• Kuiken, Mos & Vedder (2005)		
• Crookes (1989)	• Robinson (2001a, 1995b)	• Fukata & Yamashita (2015)	• Révész (2011)		
• Ting (1996)			• Michel (2011, 2013)		

5.2.1 Planning time studies

Results of planning time studies tend toward support of the Trade-Off Hypothesis which purports that while attentional capacity is limited, attending to one language performance area may take attention away from others. Also greater task difficulty is associated with lowered performance in some areas with complexity and accuracy in competition for resources. Increased planning time tended toward either greater complexity or greater accuracy but not both simultaneously (Crookes, 1989; Ting, 1996; Skehan and Foster,1997; Mehnert, 1998; Ortega, 1999; Yuan & Ellis, 2003; Gilabert, 2005).

5.2.2 Here and Now studies

Here and now studies have shown results indicating simultaneous increases in lexical complexity and accuracy (Robinson, 1995b; Rahimpour, 1997: Gilabert, 2005) and studies showing increase in measures of accuracy (Iwashita, 2001) or greater lexical complexity and fluency Robinson (2001a) without detriment to other aspects. These results are supportive of the Cognition Hypothesis.

5.2.3 Few Elements studies

Studies of +/- few elements also lean toward the Cognition Hypothesis, at times demonstrating simultaneous increase in performance in terms of various aspects of the language (Kuiken, Moss and Vedder, 2005; Kuiken and Vedder, 2007). Some studies reported increased performance in one aspect without other aspects being negatively affected as predicted by the Skehan’s Trade-Off Hypothesis (Kuiken and Vedder, 2008). Michel, Kuiken and Vedder (2007) reported results contradictory to the Cognition Hypothesis, however, in studying monologic and dialogic oral tasks which were manipulated along demands of +/- few elements for task complexity. Monologic tasks gave results predicted by the Cognition

Hypothesis, but the dialogic task showed a trade-off between greater complexity and decreased accuracy as predicted by the Trade-Off Hypothesis. In a study by Michel in 2011, subjects who performed a dialogic argumentative task showed greater lexical diversity in the complex task, but no other measures demonstrated changes, and no trade-offs were detected. In 2013 she revisited the data and performed further analysis with more task specific measures but found little more effect for increased task complexity.

5.2.4 Task Structure studies

As mentioned earlier in this dissertation, Tavakoli and Skehan (2005) complexity was manipulated along demands of +/- task structure as operationalized by +/- loose structure and +/- background events. Tighter structure generated greater accuracy while more background information resulted in more syntactic complexity. In this case, a more tightly structured task led to greater fluency and accuracy while the fact that the need to carry over background information into the narration of the task increased complexity. Results were replicated in Tavakoli and Foster (2008, 2011) as well as in Foster and Tavakoli (2009). Their results offer support for the Cognition Hypothesis although Skehan (2009) suggests that dual increase of measures of accuracy and syntactic complexity is due to cumulative effects of task demands.

Ahmadian (2015) also manipulated tasks along resource dispersing elements of +/- task structure as well as +/- planning time. Simplifying tasks along resource dispersing dimensions resulted in increases in measures of more complex, accurate and fluent speech while those who performed the unstructured task under pressured planning conditions were those who obtained the lowest scores all around.

In Robinson (2001a) a map task where cognitive demands were increased along +/- few elements and +/- previous knowledge were used to investigate effects of task complexity along these two dimensions simultaneously. As in Robinson (1995a), task complexity affected lexical complexity. Fluency was positively affected as well. Syntactic complexity was not significantly affected by task complexity. Results are supportive of the Cognition Hypothesis.

5.2.5 Reasoning demands studies

With respect to studies which placed a focus on +/- reasoning demands. Niwa (2000), investigated effects of task complexity along +/- reasoning demands on language production in a monologic narrative task. Results indicated that as task complexity increased, structural complexity also increased. Niwa also observed that fluency was differentiated, to a large degree, by individual differences. Higher working memory capacity and aptitude were associated with less fluency, as those learners with higher abilities allocated resources toward greater accuracy and syntactical complexity.

However, Nuevo (2006), in investigating learning opportunities and development of the L2 under varying conditions of task complexity which included the use of narrative tasks contrasting along dimensions of reasoning demands, did not report significant effects of task complexity on accuracy. Révész (2011) used an argumentative group discussion task manipulated in cognitive complexity along +/- reasoning demands and +/- few elements. Results confirmed that as task complexity increased, participants' language increased in lexical complexity and accuracy but with syntactically less complex language. Finally, Gilabert (2007) focuses on the use of self-repairs in L2 speech as a measure accuracy as complexity is manipulated along dimensions of +/- Here-and-Now ,

+/- few elements and +/- reasoning demands in three different tasks: narrative, map task, and decision making task, respectively. Results indicated and effect of increased task complexity on accuracy although differently for the varying task types. Overall, it has been shown that increasing demands along resource-directing dimensions may draw attention toward the way the message is encoded.

5.2.6 Intentional Reasoning studies

In the present study, task complexity is operationalized through the resource directing dimension of +/- intentional reasoning as per Robinson's Triadic Componential Framework. There seems to be relatively few studies which use this dimension as a means to manipulate task complexity. Robinson (2007), used a picture arrangement task, but from the Japanese version of the Wechsler Adult Intelligence Scale-Revised. Participants worked in dyads. One was asked to place pictures in order as their partner related a story. There were three levels of task complexity; simple medium and complex. Robinson reports greater structural complexity with the increased reasoning demands while reporting increased accuracy as well.

Another study by Ishikawa (2008) showed similar results where tasks manipulated along dimensions of +/- intentional reasoning also resulted in increases in both syntactical and lexical complexity, as well as in accuracy at the expense of fluency.

One other study, however, by Fukuta and Yamashita (2015) which also included a task where complexity was manipulated through +/- intentional reasoning demands resulted in increased accuracy at the expense of fluency, but with no affect on any measures of linguistic complexity.

The varying results of the effects on dimensions of linguistic aspects in each of these studies, whether they deal with written or oral production, result in conclusions that do not point toward a clear consensus favoring either the Cognition Hypothesis or the Trade-Off Hypothesis. However, as per Dörnyei (2006), one could consider that a closer look at how individual differences between subjects could reveal that otherwise seemingly like subjects, may, in fact perform the same task in quite different ways. Variation in the measures between subjects could possibly diminish observable effects that the results might otherwise demonstrate if those same measures were studied in light of individual differences inherent in subjects. As mentioned earlier in this dissertation, working memory capacity has been demonstrated to be a strong predictor of language proficiency. Several studies have investigated how differences in working memory capacity affect speech production. A review of some of these studies follows.

5.3 Studies on effects of working memory capacity on speech production

In a doctoral dissertation, Bergsleithner (2007) described a study intended to determine the relationship between working memory capacity, noticing and speech production of the L2 grammatical structure of indirect speech. Pre-test and post-tests were given between which learners were subjected to a process of instruction on the target structure. The pre-task consisted in an oral task to assess performance of the use of the grammatical structure of indirect questions. A series of indirect questions were to be elicited from the students using two pictures. Two post-tests were administered; one immediately after the instruction process during which students participated in an oral production task. To learn whether explicit rules were noticed as a result of the instruction process, questions about the rules were posed orally to the students as well. A second delayed post-test was administered two weeks after instruction to test whether instruction resulted in acquisition of the structure. Results show significant relationships between working memory capacity, noticing of L2 forms, and grammatical accuracy. It is also suggested that individuals with a larger working memory capacity noticed more aspects of the target structure and performed better in the L2 oral tasks as concerns the accurate use of the acquired structure.

Carpenter and Just (1989) describe two experiments intending to discover what role working memory capacity plays in reading comprehension. Focus is placed on two principal points. The first is on the computational and storage demands required during comprehension where working memory capacity is viewed as having both storage and operational capacity, necessary in order to relate early parts of language items presented in sequence with language items which are presented at a later point in the sequence. The other focus is on individual differences in the ability to maintain information in working memory during comprehension.

The first experimental procedure used a reading span test where subjects were required to read a series of sentences and remember the final word of each sentence. Sets varied from two to seven sentences. Eye fixation on each word was measured in milliseconds to determine gaze duration according to word length measured in number of letters, or a word frequency index. Gaze duration on a word as related to its length is attributed to the encoding process involved for visual recognition. Gaze duration resulting from the word's normative frequency in the language was attributed to the process of accessing the word's meaning in the mental lexicon. Analyses were focused on the contrast between six low-span and six high-span subjects. The second experiment followed the same procedure, but the final words which were to be recalled by the subjects formed a sentence.

The main result of the study shows that when memory load is increased, the lexical access process and not word encoding is affected and only for the high-span readers. A trade-off was detected between comprehension and storage. Subjects with less working memory capacity manifested trade-offs earlier. Learners with greater working memory capacity allocated resources dynamically when greater demands were imposed suggesting greater effectiveness with which they are able to allocate working memory resources according to the authors.

Borges Mota Fortkamp (1998) investigated whether there is a correlation between individuals’ working memory capacity and their oral fluency in English as an L2. A picture description task and a narrative task were used. However, in this case, no significant correlations were found between measure of working memory capacity and measures of fluency. It is suggested that the measures of working memory capacity used for the study may not be appropriate as predictors of L2 fluency as they may not be sensitive enough to cognitive process involved in L2 oral production.

Borges Mota (2003) describes a study of the relationship between working memory capacity and L2 speech production as per measures of fluency, accuracy, complexity, and weighted lexical density. Oral data was elicited with a picture description task for which subjects were constrained to a time limit of 2 minutes, and a narrative task which required subjects to describe a movie of their choice. No time limit was established for subjects to complete the task. An indeterminate pre-task planning time was allowed before each task. Data was analyzed using measures of fluency, accuracy, structural complexity and weighted lexical density.

Results indicate that a higher working memory capacity produced greater fluency, accuracy, and complexity, but less weighted lexical density. There seemed to be a trade-off effect between weighted lexical density versus fluency, accuracy, and complexity. Increases in task complexity, however, were not contemplated except that the tasks were considered complex enough in their own right to warrant the trade-off. The speaking span test proved to be a significant predictor of fluency, accuracy, and complexity in L2 speech and the author suggests that working memory capacity partially accounts for variation in L2 oral performance. The author also suggests that grammatical encoding in the L2, as a complex subtask of L2 speech production that requires the control and regulation of attention, may explain the relationship that exists between working memory capacity and the measures of L2 speech production.

Borges Mota Fortkamp (2007) examined the relationship between individual differences in working memory capacity, noticing, and L2 speech production. It is suggested that working memory capacity is related to accuracy in L2 speech production but not to the ability to notice L2 formal aspects in the input.

A pre-testing phase included the speaking span test and a picture description task which required the use of the verb ‘need’. A treatment phase was then implemented which included instruction on the use of ‘need’ with either a gerund or and infinitive. Two post test sessions were done; one immediately after the treatment session and another one week later. The post-test included an oral protocol collection about the target structure and a rule description and production of two example sentences. Oral data was transcribed and measures of accuracy were collected as well as measures of noticing measured by accuracy of students’ example sentences during the oral protocol of the immediate post-test. Results indicated that while subjects with greater working memory capacity produced significantly more accurate speech, those subjects demonstrating lower working memory capacity produced more errors although not significantly so. Measures of noticing did not differ between the groups.

Finardi and Weissheimer (2008) investigated whether increased proficiency resulted in greater L2 working memory capacity being made available to subjects. Results were compared and it was found that higher proficiency level students also scored higher on the L2 speaking span test implying that as proficiency level increases, more L2 production processes become automatized which, in turn, frees up working memory resources that are required for controlled executive processes.

Fortkamp (1999) investigated whether working memory capacity correlates with fluent L2 speech production. Task types included picture description task, an oral reading task, and an oral slip task during which subjects read cued pairs of words shown on a computer screen. The author claims that the results of the study support the view that working memory capacity is task-specific. The efficiency of an individual to perform a particular task may vary as a function of that individual’s working memory capacity as it correlates to the cognitive processes required by the task at hand.

Trebits and Kormos (2008) describe a study of how working memory capacity affects the oral performance of learners’ L2 as during narrative tasks of differing levels of cognitive complexity. Cognitive complexity was manipulated in that where one task consisted in a story narration following a picture sequence which followed a clear storyline, the other task consisted in a series of unrelated pictures from which a story had to be invented, thus increasing cognitive demands on the learner. Increased task complexity did not result in differences between measures. Compared with other studies, it seems that these students had too low a level for differences to appear between measures. The simple task seems to have elicited more lexical sophistication than the complex task which was contrary to the hypothesis. Working memory capacity took a certain, though limited, role in the output. Differences between low and high capacity learners were manifested in the complex task. There was a tendency toward more fluency for learners with greater working memory capacity leading the authors to speculate that these learners superior ability to control attention may aid them in conceptualizing and formulating their speech under complex conditions. In addition, lexical complexity showed to be significantly higher for learners with a greater working memory capacity. However, no difference was found between the groups for accuracy or syntactic complexity. This shows that although more attention was allocated to lexical richness, other linguistic aspects were not diminished. The authors suggest that those learners with a greater working memory are able to regulate and control attentional processes more efficiently, leaving a sufficient amount of resources for retrieving appropriate vocabulary as well as encoding their message accurately and expressing themselves fluently.

Gilabert and Muñoz (2010) investigated the role of working memory capacity in attainment and performance of English as an L2. The task that was used was a film retelling task. They found that the high working memory capacity group was more fluent and used greater variety of vocabulary. The researchers suggest that higher working memory capacity is associated with faster lexical access and retrieval which could aid fluency. However, they also suggest that greater vocabulary knowledge due to subjects' higher proficiency may also have had an effect on greater fluency. They found proficiency to be a better predictor of lexical complexity than working memory capacity. The researchers also found that working memory only correlated with performance of high proficiency learners leading them to conclude that working memory most benefits L2 learners at a later stage of acquisition.

5.4 Focus of the current investigation

This chapter provided a brief summary of investigation which has been carried out within the field of SLA in relation to task manipulation and its effects on language production, especially in light of claims made by both the Cognition Hypothesis and the Trade-Off Hypothesis. While the list of studies and their results is not complete, it is illustrative of the inconsistencies of results which have lacked consistent concordance with the predictions made by either one or the other of the two hypotheses.

In addition to a description of research done on task manipulation and its effects, this chapter has also provided a selective description of several reports on investigation of the effects that working memory capacity has on language production. In these studies, it has been demonstrated that subjects who differ in working memory capacity can produce different results in measures of their linguistic output.

The evidence that points toward a link in individual differences in working memory as it affects linguistic output together with inconsistent results in how changes in task complexity result in trade-offs between dimensions of CAF led to questions in regards to how investigations into the effects of task manipulation might show more consistent results if individual differences were taken into consideration.

5.4.1 Issues under investigation

Tasks as pedagogic procedures which rely not only on their communicative characteristics, but on cognitive aspects as well, guide language learners step by step as they work toward achieving capacity to perform real life activities in the target language. As a part of a task-based syllabus, their presentation from simple to complex is intended to simulate the ontologic development of language skills while at the same time tapping into conceptual knowledge of the world as adults connect these concepts with linguistic aspects of a target language. So, pedagogic tasks should be designed and sequenced according to the increases in their complexity, or the cognitive demands that they place on learners so that in a classroom setting, they approximate task demands that one may find in the real world (Robinson & Gilabert, 2007).

Task complexity is the result of the attentional, memory, reasoning and other information processing demands that are imposed on a language learner by the structure of a task (Robinson, 2001a). Research into task complexity and its effects on language production is necessary in order to make decisions about the grading and sequencing of pedagogic tasks for use in syllabus design (Robinson & Gilabert, 2007). However, investigation has been inconclusive as to

Investigation to date has demonstrated little clear evidence of how increases in task complexity affect allocation of cognitive resources during task performance affecting language production. Results have been varied and conclusions have been indeterminate. A large part of the variation is a result, either directly or indirectly, of individual differences in learner characteristics (Dörnyei, 2006).

There is a need for a more meticulous categorization of study subjects based on an understanding of the effect of individual differences on observable variation in language production so that more definitive results can be found during data analysis. In order to carry this out, however, it is first necessary to fill the gap in the current literature where there is little understanding of how individual differences may affect performance results, and ultimately, learning outcomes. With this knowledge, instructional options can be more effective within language learning contexts and research carried out more efficiently. (Robinson, 2001a, 2002).

The research described in this dissertation intends to demonstrate whether trade-offs between measures of fluency, accuracy and complexity are manifested differently along a continuum of complexity of tasks performed by subjects who differ in their levels of working memory and attentional capacity while taking into consideration affects for proficiency.

5.4.2 Research questions and hypotheses

Review of the SLA literature has raised two principle questions in regards to individual differences in attentional and working memory capacity and the way that they may affect L2 language production as they perform tasks under increasingly demanding conditions.

5.4.2.1 Research questions

1. As learners perform a series of oral tasks that vary in level of cognitive complexity along a continuum from simple to more complex, what influence does attentional and working memory capacity have in

determining the point along the continuum at which they begin to demonstrate competition for attentional resources between dimensions of CAF?

2. What relationship does working memory capacity have with the degree to which trade-off effects appear in speech samples produced under increasingly complex conditions?

5.4.2.2 Hypotheses

1. As tasks increase in complexity, subjects who have lower working memory capacity will manifest competition for attentional resources sooner along the complexity continuum than those subjects who have greater working memory capacity.
2. During completion of an oral task under complex conditions, there will be a difference in the clarity with which trade-offs are manifested between dimensions of fluency, accuracy and complexity in the speech samples of those subjects who differ in working memory capacity.

Hypothesis 1 is based on findings by Carpenter and Just (1989). During a study involving reading comprehension, these researchers showed that a trade relation was manifested once subjects' working memory capacities reached their limits. These limits were attained at different loads of complexity as operationalized by the task and corresponded to the subjects' working memory capacities. Trebits and Kormos (2008) in their study of the relationship between working memory capacity and performance on narrative tasks, also suggest that learners with a greater working memory capacity may have more efficient control over the allocation of attentional and memory resources. In their study, this explanation allows for high capacity learners being left with enough resources for retrieving appropriate vocabulary and for encoding their messages accurately while expressing themselves fluently under increased task demands.

Hypothesis 2 stems from various sources in the literature. Drawing from Dreary *et al* (1996) and their Ability Differentiation Hypothesis, Robinson (2001a) holds that learners with higher level cognitive capacities should demonstrate more differentiation in abilities than learners with lower cognitive capacities. Carpenter and Just (1989) suggest that subjects with high working memory capacity, in contrast to those with less working memory capacity, dynamically reallocated memory resources when new task demands are imposed. Although differences are slight when comparing learners with high and low working memory capacity, where task demands are greater, differences between the two types of learners are more pronounced. The authors suggest that this may result from their ability to allocate resources more effectively to prioritized dimensions of speech once capacity is taxed. In their study of the role of working memory capacity on syntactic processing during reading tasks, King and Just (1991) also observe that the supply and efficient use of cognitive resources may differ from person to person. Finally, as mentioned above, Fortkamp (1999) suggested that individuals may vary in the efficiency with which they perform tasks depending on their working memory capacity as it correlates to the cognitive processes required by the task at hand.

5.5 Summary

This chapter provided a list of research behind the motivation for the present dissertation, and research questions and hypotheses were posited. The next chapter will describe the protocol that was followed to address the questions and to test the null hypotheses that working memory and attentional capacity do not play a role in determining differences in performance between subjects as they perform oral tasks at different levels of task complexity.

CHAPTER 6: METHODOLOGY

6.1 Introduction

Speech production models described above and theories of language acquisition rely heavily on an understanding of workings of memory and attention. Where memory plays a key role in SLA, attentional control determines the processing efficiency of memory (Engle, 2002; Engle, *et al* 1999; Kane, et al 2001; Kane & Engle, 2003) as it affects the quality of perception of stimuli, essential to noticing and awareness of input and subsequent acquisition (Schmidt 2001, 2010).

Proficiency can be expected to have an effect on measures of fluency accuracy and complexity as demonstrated in Gilabert, Levkina, and Baron (2011). The study controls for this by investigating performance at different levels of proficiency to compare and contrast the data collected from each group.

In order to investigate the research questions proposed, measures of working memory, attentional control, and English language proficiency were used to control for each of these factors during the study as measures of oral performance were collected and analyzed. Descriptions of each of the steps of the process follow below.

6.2 Test of working memory capacity

Working memory was tested by means of an automated reading span task (Unsworth *et al*, 2005) adapted to Spanish and Catalan by the Language Acquisition Research Group at the University of Barcelona. The test consists of three parts. During the first part of the test, subjects perform a letter span test in which they must recall fourteen series of between three and nine letters. For the second part of the test, subjects read a series of sentences presented to them one at a time in the language which they previously indicated as their L1, and they must decide whether the sentence makes sense or not. The final part consists in a combination of the first two where they are presented with between three to seven series of sentences and letters. After presentation of each series, subjects are asked to recall the letters that they saw in the order in which they appeared. An adjusted score was provided at the end of the test as a measure of working memory capacity.

6.3 Tests of Attentional Control

There are several tests of attentional control of from which one was chosen for the study. These included an Antisaccade task (Hallet, 1978), a Stroop task. (Stroop, 1935; MacCleaod, 1991), a Trail Making Test (Arbuthnott and Frank, 2002), and finally, an alternating runs paradigm (Rogers and Monsell, 1995). The test that was finally chosen as means to measure attentional control was the alternating runs paradigm. A description of the theoretical background behind the test which provided the rationale for using the test in the present research is described in detail below.

6.3.1 The Alternating runs paradigm

Rogers and Monsell (1995) created the alternating runs paradigm, an attention switching task by which subjects perform sets of alternating tasks such as identifying a letter as either a vowel or a consonant, or a number as either odd or even. The task proposed by Rogers and Monsell drew from work done by Dr. Arthur Thomas Jersild in 1927. Jersild developed a process which compared performance on two different kinds of blocks of tasks referred to as alternating trial blocks and pure blocks. To perform the alternating trial blocks, an individual switched between two different kinds of tasks (i.e. ABAB). For the pure blocks, the individual performed only one of the tasks (i.e. AAAA or BBBB). Jersild compared the time taken to complete alternating trial blocks with the time taken to complete pure blocks. The difference in compared times, referred to by Rogers and Monsell as the switch cost, is used as an index to measure the added cognitive difficulty imposed on the task performer by having to switch from one kind of task to a new task with a different set of requirements.

The Alternating Runs task switching paradigm developed by Rogers and Monsell (1995) differs from the task switching paradigm designed by Jersild in 1927 in that where Jersild compared the switch cost between entire blocks of tasks, Rogers and Monsell compared switch cost for switch and non-switch trials within a single block. In each block, a participant would perform alternating runs of trials of predictable length for each task. So, where the length of runs equals 2, the subject would perform a sequence of pairs of trials AABBAABB. To determine the switch cost, a comparison is then made between performance on the trials in which the participant had to switch between tasks (i.e. AB, BA) with trials where no switch was made (i.e. AA,BB). Measures of performance were the difference of reaction time, referred to as time cost, and the difference in error percentage which was referred to as error cost.

The alternating runs paradigm is considered an effective means of measure of attentional capacity. It is a measure of endogenous control; a capacity of control required for an individual to switch attention effectively from one task to another or, as in the case of the alternating runs paradigm, for an individual to adopt a new task paradigm before being presented with the actual stimuli simply by knowing which kind of task is going to be presented as the paradigm changes predictably, according to Rogers and Monsell (1995). Drawing from the executive system, regulating between endogenous and exogenous control of behavior during task switching is resource depleting, so measures provided by the test show difficulty exhibited by endogenous control mechanisms in overcoming exogenous activation. This is demonstrated not only by failure to behave according to task demands, but also by the slowing of performance to meet the task demands as inappropriate behavior is inhibited. The result is slower reaction times registered during switch trials compared with non-switch trials.

The alternating runs paradigm as developed by Rogers and Monsell has become a popular measure of cognitive control. Where the tasks they used required identification and classification of digits and letters, other researchers have used any number of tasks within an ever expanding range of variations of the paradigm. The current researcher has looked for a precedent after which he could model a paradigm that would require a participant to make a simple choice based on visual stimuli given that the stimuli used in the current investigation is likewise visual. In addition, the paradigm should allow that the subject’s choice be made without any need for linguistic production in order to avoid interference from as many other cognitive processes as possible apart from those involved in the control of attention during the performance of the task at hand. Such a precedent was found in a study which used an alternating runs paradigm requiring subjects to distinguish between forms of shapes and their colors.

In their work examining the origins of mixing cost in task switch paradigms, Rubin and Meiran (2005) had participants perform an alternating runs paradigm in which the two main tasks required them to discriminate between either the color of determinate shapes, or their form. In the current study, participants performed similar tasks also within an alternating runs paradigm. One task required participants to determine whether the forms of two shapes of equal size are the same or not, regardless of their color. The second task required them to determine whether the colors of two shapes of equal size are the same or not, regardless of their form.

During visualization of any given object, its dimensions may be perceived in two ways: holistically, where dimensions and their features are perceived unitarily as a whole, or independently where dimensions and their features are perceived separately from others. These concepts were adopted by Garner (1974; 1976) who describes the interaction of varying dimensions of visual stimuli during choice making processes. Garner defined a stimulus dimension as a variable attribute of a stimulus which may have two or more different values. A dimension of a specific stimulus might be the shape of an object, it being a circle, a square, a triangle, etc., or the color of an object. A stimulus can also vary on two or more dimensions, such as both shape and color, potentially resulting in being, as a matter of example, a red circle, or a blue square.

The dimensions which make up stimuli may interact in various ways during the visual perception process. Depending on which dimensions are being processed, they are analyzed either separately from other dimensions, or integrally, conjoined with other dimensions. Perception of the attributes of width, height, and the angles connecting the sides of a shape conjoin to form the features of the shape which distinguish it as a rectangle rather than, for example, a square or a triangle. Likewise, the features of tint and hue of a color are perceived integrally making it distinguishable from another color. Color itself, on the other hand, is a visual dimension the features of which are perceived separately from the features which determine the visual dimension of shape. So, the underlying cognitive processes required by a person toward the perception of an object's color are different from those required for the identification of its shape. In the case of a red square, an observer will initially perceive the color red and the form square, but will not have an integrated percept of a red square until these separate dimensions are joined at a later stage in the perception process (Garner, 1974, 1976; Treisman & Gelade, 1980; Cheng & Pachella, 1984).

According to the integration theory of visual attention (Treisman and Gelade, 1980), an observer first registers specific dimensions of an object such as color and shape during a pre-attentive stage, an automatic process done in parallel across the field of vision. Then, focal attention is required to serially process the various separately registered dimensions and combine them into a single, identifiable object. Once attention is focused, an overload of resources during processing could result in 'illusory conjunctions' or false images.

According to Garner (1976), were an experiment to use a stimuli formed by just the two dimensions of color and shape, that experiment could be called a perceptual classification problem and would lack what is referred to as irrelevant dimensions. Irrelevant dimensions are those dimensions which exist, but are not relevant for the definition of the object. So, where shape and color of an object are the separable dimensions by which it is identified for the intended ends of the experiment, size would be an irrelevant dimension. Nonetheless, dimensions apart from those used to define a stimulus, despite their irrelevance to the definition of the stimuli, may interfere with an observer's perception. This is true if they are integral together with a relevant dimension as happens between features of height and width when task demands require judgment of sameness of shape, or those of tint and hue while determining sameness of color (Dixon and Just, 1978). Dixon and Just (1978) demonstrated the role that irrelevant dimensions can play during the perception process by showing how variations in them affected reaction times in experiments in which subjects were asked to judge the sameness between stimuli according to specific relevant dimensions. When there was a variation in the irrelevant dimension where both the irrelevant and relevant dimensions comprised a single integral construct, reaction time was affected in determining sameness of the relevant dimension of the construct. One experiment used ellipses where subjects demonstrated mutual interference between shape altering integral dimensions of width and height while determining sameness. And another experiment used color, where the irrelevant dimension of tint interfered with the relevant dimension of hue.

According to Garner, concepts which are defined within sets of stimuli that have no irrelevant dimensions affecting perception are so simple that there are no meaningful differences between experimental conditions under which one or the other of these conditions have to be distinguished. So, for the current alternating runs paradigm, the experimental conditions of determining either sameness of color or sameness of shape are equal and without the interference that irrelevant dimensions would impose if the stimulus dimensions of shape and color were not separable.

6.3.2 Alternating runs task for the present investigation

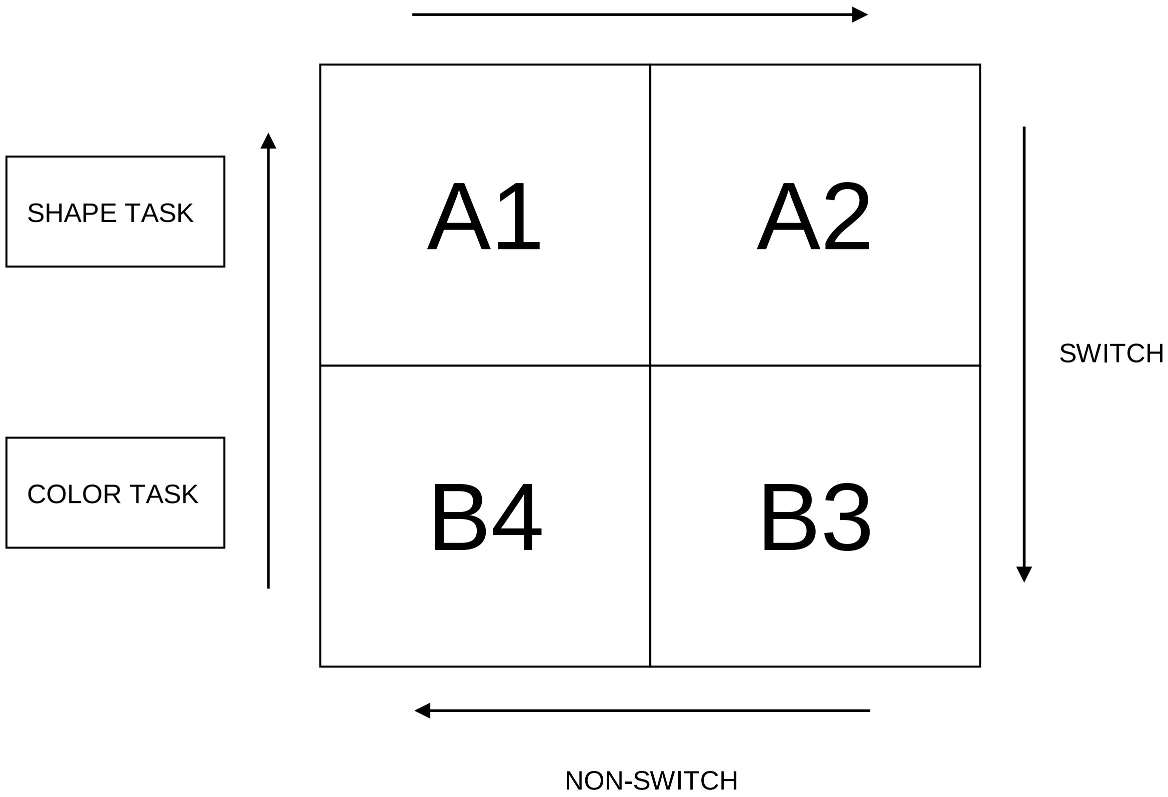
The aforementioned research led to the creation of the alternating runs paradigm used in the investigation. The test was created in E-prime by the investigator. In this paradigm, subjects are instructed first to choose whether shapes of objects are the same or different, and then to choose whether the colors of the stimuli are the same or different or which they must switch their attentional focus. The two tasks are drawn on different, separable cognitive dimensions implying no interference between them due to irrelevant dimensions that could alter perception. This provides for equal experimental conditions for both tasks; the processes for registering both color and shape being automatic and taking place pre-attentively. Additionally, according to the feature integration theory, attentional resources are later required by both tasks in order to focus attention on those features which either define the shapes of the objects, or their colors while requiring those resources to be properly economized and allocated in order to avoid cognitive overload leading to illusory conjunctions which would interfere with the accuracy of subjects' replies. Those subjects who demonstrate an ability to complete the tasks with fewer errors and with a smaller difference in reaction times between switch and non-switch trials can be expected to have demonstrated a greater level of cognitive control.

The alternating runs task is pertinent as a measure of non-linguistic task switching capacity as it pertains to language as it has been demonstrated that mechanisms involved in task switching activity is domain general and transferrable to task switching in bilingual contexts (Prior &MacWhinney, 2010; Weissberger *et al*, 2012).

6.3.3 Performance procedure for the alternating runs task

The procedure for performing the test created for the current study is as follows. Subjects are presented with a computer screen that is divided into four quadrants of equal size. Each consecutive trial takes place in one of the quadrants beginning with the quadrant in the upper left hand corner of the screen and continuing around the screen to the subsequent quadrant in a clockwise direction as each new trial begins (figure 2).

Figure 2 – Quadrants in the alternating runs paradigm



Each trial starts with a fixation point that appears in the form of a ‘+’ sign for 1.5 seconds in the center of the corresponding quadrant. Immediately after the fixation point disappears, two colored shapes of equal size appear side by side in the same quadrant. The forms of the shapes may be the same, or different, and likewise the colors of the shapes may be the same or different. The shapes may vary in form between a circle, triangle, or square, and they may appear in any of a range of six different colors. These are red, blue, green, yellow, brown, and pink. The colors were chosen based on work done on basic color terms by Berlin and Kay (1969) in which the investigators studied color terms and their comparative meanings in ninety-eight different languages. Apart from black and white, which Berlin and Kay found that every language in their study had terms for, the colors used in the present paradigm were those colors which were found to be more commonly accounted for in the languages studied. This is an indication that each of the colors is sufficiently different from the others in order for there to be any confusion between them as to sameness.

Each block consists of four separate trials. The first two trials take place in the two quadrants at the top of the screen (labeled A1 and A2 in figure 2), and these corresponded to the ‘shape’ task. Whenever the colored shapes appear in the quadrants on the top half of the screen, the subject is to decide whether the forms of the shapes are the same or different and to respond by pressing specific either one of two previously marked keys located on opposite extremes of the computer keyboard with either the right or the left hand. When the shapes appear in the quadrants on the bottom half of the screen the task changes and becomes the ‘color’ task. For the quadrants on the bottom half of the screen (labeled B3 and B4 in figure 2), the subject is to decide whether the colors of the shapes are the same or different and to respond accordingly by pressing the keys which have been marked on the keyboard. Switch conditions are considered to be undergone during performance of the trials after which the tasks have changed from ‘shape’ to ‘color’ and vice versa. These correspond to those trials in the first and third quadrants. Trials in the other quadrants are considered to be performed under non-switch conditions. Subjects have five seconds to respond after which time the following trial starts automatically if no response has been given for the previous trial.

Before performing the task itself, the subject is given a series of instructions in their mother tongue, explaining the operationalization of the task. The subject also performs a series of practice runs for both the ‘shape’ task and the ‘color’ task separately during both of which no switch condition is undergone. Finally, the subject performs another practice run which simulates the actual task in that both switch and non-switch conditions. During the practice runs, if an error is made, the subject is presented with a screen to point this out and to provide a reminder of what is to be done during the task at hand. Each practice run consists of eight separate trials. If the subject does not answer without an error for a minimum of ninety percent of the trials during any of the practice runs, then that run is repeated until ninety percent of the trials are responded to correctly. This is to ensure that the rules of operationalization are clear for the subjects before data collection takes place.

Once the subject finishes the practice stage of the paradigm, the actual task begins. During this stage, the subject is not informed of errors in order to avoid deliberate focus on accuracy during performance. In order to ensure that the subject is sufficiently accustomed to performing the task before data is collected, the first sixteen trials are done for practice without data collection. Once data collection begins, twenty-four blocks are performed during which reaction times of ninety-six trials are collected, forty-eight of which consist of switch trials and another forty-eight of which consist of non-switch trials.

6.3.4 Validation of the alternating runs task

A pilot test was undertaken in order to ensure that the alternating runs test that was created would be an effective means of measuring attentional capacity. Fifteen subjects performed the test twice not performing the second test sooner than two weeks prior to the first testing session. Descriptive statistics are provided in Table 3.

Table 3 - Descriptive statistics for validation of alternating runs paradigm. Mean values in milliseconds.

					Skewness		Kurtosis	
		N	Mean	Std. Deviation	Statistic	Std. Error	Statistic	Std. Error
Session 1	Quadrant A1	15	886.82	173.57	.081	.580	-1.230	1.12
	Quadrant A2	15	755.69	113.10	.906	.580	.175	1.12
	Quadrant B3	15	930.00	150.72	.370	.580	-.687	1.12
	Quadrant B4	15	809.28	129.34	1.024	.580	1.251	1.12
Session 2	Quadrant A1	15	725.27	100.07	.417	.580	-.753	1.12
	Quadrant A2	15	646.69	64.61	1.236	.580	2.072	1.12
	Quadrant B3	15	789.89	125.57	.338	.580	-.669	1.12
	Quadrant B4	15	699.50	60.92	.485	.580	.952	1.12

Paired samples t-tests were run in order to determine if differences between switch and non-switch trials were significant. A Bonferroni correction was applied, resulting in a significance level set at $p < 0.006$. Results are listed in table 4. The only instance of a non-significant difference occurred between quadrants B4 and A1 during the second round of testing sessions.

Table 4 Paired samples t-tests between switch and non-switch trials

				95% CI				
	Quadrant pairs	Mean	Std. Deviation	Lower	Upper	t	df	Sig. (2-tailed)
Time 1	A1 X A2	135.60	104.88	77.51	193.68	5.007	14	.000*
	A2 X B3	-178.77	122.06	-246.37	-111.18	-5.673	14	.000*
	B3 X B4	120.71	62.79	85.94	155.48	7.446	14	.000*
	B4 X A1	77.54	85.35	30.27	124.80	3.518	14	.003*
Time 2	A1 X A2	92.58	69.03	54.35	130.80	5.194	14	.001*
	A2 X B3	-149.83	91.27	-200.37	-99.28	-6.358	14	.000*
	B3 X B4	97.02	90.09	47.13	146.91	4.171	14	.001*
	B4 X A1	39.77	99.82	-15.51	95.05	1.543	14	.145

**α significant at p<.005 (Bonferroni correction)*

Switch costs for both testing sessions were calculated as described above. A Spearman correlation demonstrated a significant positive correlation between the results of each of the two testing sessions that were carried out with each participant, Spearman's rho (15) = .615, p =.015.

These results demonstrated that the alternating runs paradigm that was designed and created for purposes of the investigation could effectively be used to provide a measure attentional capacity.

6.4 Tests of proficiency

In the current study, proficiency was controlled by means of two measures. The first is the Oxford Quick Placement test, and the second is a test of vocabulary breadth test which consists of two parts, the X-Lex and Y-Lex vocabulary tests (Meara & Milton, 2003). Based on analysis of the results of these tests, subjects were categorized into two groups of higher and lower levels of proficiency. This distinction between levels of proficiency may offer insight into what degree proficiency level, apart from cognitive resource capacity, affects the observances of trade-off effects as the tasks increase in complexity.

6.4.1 Oxford Quick Placement test

The Oxford Quick Placement (OQP) test is a test of English language proficiency created by Oxford University Press and Cambridge ESOL and designed for quick and simple administration (Geranpayeh, 2003). The test which was used for the current investigation is a pen and paper test divided into 2 parts. Part I of the test consists of 40 multiple choice questions. Part II consists of an additional 20 questions. Instructions state that if a subject responds correctly to fewer than 36 of the first 40 questions, then part II of the test should not be considered for determining the proficiency level of the individual. A chart is provided in the instructions by which the examiner can determine the level of the subject based on the number of correct answers given. If the individual scores 36 or more questions correctly on part I of the test, then the final 20 questions are taken into consideration for determining the proficiency level. Again, the examiner refers to a chart which suggests the level of the test taker based on the number of correct responses out of a total of 60.

6.4.2 X_Lex and Y_Lex tests

Two different yes/no style tests of receptive vocabulary size were used as a complementary means to determine proficiency level of the subjects. The two tests included the X_Lex version 2.05 (Meara 2005) and Y_Lex version 2.05 (Meara & Miralpeix, 2006) both administered on a computer. Yes/no tests are meaning-recall tests for which subjects are shown a series of written words of which they indicate those that they think they know the meaning of. In order to compensate for the possibility of a subject's claim to know a word that they, in fact, do not really know, non-words are included in the list of words presented. A final score is then mathematically determined based on the accuracy of the subject's responses (Pignot-Shahov, 2012).

Use of the tests as a means of determining proficiency level is based on the premise that it is with difficulty that non-native speakers are able to express themselves fluently with words that are typically used by native speakers. It is therefore possible to distinguish non-native speakers from native speakers by comparing performance of the two groups. In the Lex tasks, native speakers tend not to vary in their results under the different conditions while non-native speakers do. How closely the results given by non-native speakers on the tests approach native like results is indicative of greater or lesser command of the target language (Meara, 2005).

Both the X_Lex and Y_Lex present the subject with a series of words free of context taken from several word frequency lists. The X_Lex test draws from a vocabulary of 5000 words from the JACET List of Basic Vocabulary. The Y_Lex extends the range up to 10,000 words drawing from the JACET 8000 list as well ask from Kilgarrif's listing of the British National Corpus (Miralpeix & Meara, 2014).

6.5 Picture arrangement tasks

The picture arrangement tasks are made up of a series of pictures which tell a short story once they are placed into the correct order. This form of test was originally intended to measure Performance Intelligence Quotient (PIQ), a measure that assesses capacity in dealing with nonverbal skills, but has been demonstrated to have a largely verbal component which has been correlated to other verbal subtests. It came to be interpreted as a measure 'social intelligence', or the capacity of one to interact within a social context. The test was used as part of a mental testing program used by the military during the First World War at which time David Wechsler, creator of a series of intelligence tests which bare his name, was purportedly exposed to it although the tests were limited in use (Tulsky, 2003). When Wechsler crested the first Wechsler Bellevue test in 1939, he used three picture arrangement tasks from the the Army Beta Prelimary Form and adapted four cartoons entitled 'King' from the New Yorker Magazine by the cartoonist Otto Soglow.

The picture arrangement task remained a part of the Wechsler series of tests over the various versions that have been created including the Wechsler-Bellevue II in 1946, and the Wechsler Adult Intelligence Scale (WAIS) in 1955. In 1976, the Spanish version of this test was first created. A revised version, the WAIS-R was published in 1981, and the WAIS-III appeared in 1997 with the Spanish version 1999. When the WAIS-III was created, the pictures were redrawn and content was modified to adapt to modern test takers. However, in the WAIS-III was made optional for test administrators in computing the general index score. This was due in part to investigation which demonstrated that other constructs were involved in the performance of the task other than PIQ (Tulsky, 2003). The picture arrangement task was finally eliminated from the latest version, the WAIS-IV, which appeared in 2008, and from the Spanish version in 2012 (Amador Campos, 2013).

The picture arrangement task used in the experiment was taken from the Spanish version of the WAIS-III and was designed to be carried out by Spanish speakers. The pictures used in the current study are already arranged in the WAIS-III test in order of difficulty from simple to more complex (Wechsler, 1958) determined from frequency tables of items passed or failed by the standardizing population during the design phase of the test. The increase in conceptual demands of the task is operationalized through the intentional reasoning requirements it imposes as this is the capacity which is interpreted from the score of the picture arrangement task of the WAIS-III.

6.5.1 Intentional Reasoning

Intentional reasoning is the ability to reason about cause, behavior, or intention, and to understand and predict the behavior of others and their mental states. It is referred to as theory of mind in the field of cognitive psychology. The term was first coined by Premack and Woodruff (1978) to name a phenomenon that they were investigating while attempting to determine whether chimpanzees inferred mental states such as purpose or intention. The concept of theory of mind has grown to become widely studied phenomena in various areas of cognitive psychology. It is understood to be carried out by an innately determined cognitive mechanism (Leslie, 1987) and independent of measures of intelligence (Nunez &

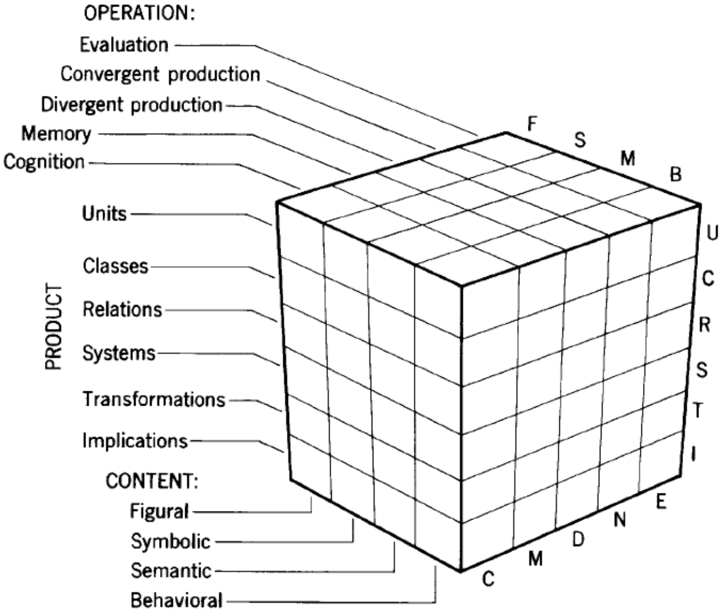
Riviere, 1990 in Fletcher, *et al*, 1995). Lack of theory of mind is characteristic of autistic individuals. This was demonstrated by Baron-Cohen, Leslie, and Firth (1986) in a study in which they used a picture arrangement task to investigate theory of mind in both autistic children and children with Down's syndrome. They determined that while both groups of children performed equally well on tasks of causal-mechanical as well as descriptive-behavioral criteria, the autistic children performed at a much inferior level on tasks of intentional reasoning.

As with that used by Baron-Cohen, Leslie and Firth (1986) study, the picture arrangement test within the WAIS-III has also been classified as a measure of intentional reasoning, or theory of mind. Kaufman and Lichtenberger (1999) provide a brief description of the categorization of the various tests of the WAIS-III according to the Guilford‘s structure-of-intellect model (Guilford 1967).

6.5.1.1 Guilford‘s structure-of-intellect model

Guilford (1967) proposed that intelligence can be evaluated based on three dimensions which he calls *product*, *content*, and *operation*. These are represented in a morphological model of intellect in which they are cross-classified, each dimension, or category, represented by one side of a cube (figure 3). Each category is further divided into factors of intelligence.

Figure 3 - The structure-of-intellect model (Guilford 1967)



The product category consists of factors which have to do with the way that information that we perceive or conceptualize is structured. It may be perceived as units such as words, shapes, or behavioral units like facial expression. We may perceive these units in classes or groups. Information may be perceived as relations linking units or concepts. Also, as systems which are patterns of interacting parts, transformations which are perceived changes, or to see things in a different way than we normally see them. And finally, implications which are things that are expected anticipated or predicted from available information.

The content category is concerned with the format of the information that we perceive. As a cognitive ability, he refers to people having capacity to pay more attention to input or who can reflect more efficiently on input. Factors making up this category include semantic (verbal-meaningful) information. This is information which carries meaning, but not necessarily in the form of word. Figural (visual-spatial) factors are perceived through the senses, and symbolic factors are such items as words or symbols carry meaning. In addition, behavioral content refers to information involved in cognitive operations which pertain to one’s ability to perceive what other people are thinking, feeling, attending to, and intending to do in order to draw inferences about their behavior.

Factors within the operational category are those cognitive processes which handle the information. Factors include cognition or the ability to perceive items. Memory involves storage and retrieval capacities. Two kinds of psychological production factors are referred to as divergent and convergent; divergent referring to the ability to access memory and retrieve various kinds of information that could solve a problem. Convergent factors describe the ability to access memory to find single answer to a problem. Finally, evaluation is the ability to pass judgment or to make critical comparisons of different kinds of information.

It was expected that people could excel in some factors while exhibiting deficit on others. So, a person may be an artist who has a great capacity to process visual (figural) information, but who struggles with symbolic content contained in words or numbers. Cross-classifying the different factors provide a description of cognitive capacities. Drawing from the content category and the product category one could refer to relationships between images or behavioral transformations such as changes in emotion.

According to Kaufman and Lichtenberger’s classification of the WAIS-III tests within Guilford’s model, the picture arrangement task is of semantic as well as figural-behavioral content. It is of semantic content in that the task consists of an inherent and meaningful narrative thread, an understanding of which leads to successful completion of the task. It is of figural-behavioral content in that the narrative is told through the figurative images of which it is made up, these being representations of people whose behavior and intentions are to be interpreted in order for the story to be completed.

Kaufman and Lichtenberger also claim the picture arrangement task draws on both the evaluation and convergent production operations. Evaluation is an ability to use logical reasoning to make differentiations between identities, similarities, and consistency, or criteria of aesthetic or ethical nature in order to meet task demands (Guilford, 1967). Psychological production factors are described as creative thinking abilities and draw heavily on memory. Convergent production is the capacity to use available information to converge upon one correct answer in accordance with task requirements. It is logical deduction, or the drawing of conclusions and the reasoning about and consequential understanding of implications. Both of these operations, evaluation and convergent production, are intrinsic to the successful completion of the picture arrangement task.

The logical foundation for the order of the factors within the model considers that as one reads from the front to the back of the model, there is an increasing dependency of each factor on those preceding it; if there is no cognition, there is no memory, and without memory there is no production, and so on (Guilford 1967). In accordance with Guilford’s model, we would expect successful completion of the picture arrangement tasks to be strongly dependent on cognitive capacity for memory, creativity, and intentional reasoning ability to understand and predict behavior from an interpretation of the figural representations of the pictures.

6.6 The Task

The main part of the present experiment consisted in subjects narrating the stories formed by a series of picture arrangement tasks. A description of the task, the protocol which was followed, and the experimental methodology are described below.

Eleven stories comprise the Picture Arrangement subtest of the Spanish version of the WAIS-III. As mentioned above, these are arranged in order of difficulty (Table 5). To elicit oral narratives from subjects in the current investigation, five stories were chosen so that task complexity could be operationalized at five different levels. Stories are arranged according the level of difficulty as operationalized by the WAIS-III.

Table 5 – Picture arrangement tasks in the WAIS-III

Level of Complexity	Task
1	CAP
2	BAKE
3	OPENS
4	CHASE
5	CLEAN
6	HUNT
7	SAMUEL
8	LUNCH
9	CHOIR
10	DREAM
11	SHARK

Within this succession, alternate storylines beginning with the third were used with the intention of keeping a sufficient distance in degree of complexity between any two. The storyline corresponding to the first two simplest levels were not chosen for fear of their not eliciting sufficient speech from which to collect data, due to their simplicity. The story lines that were chosen were *Opens, Clean, Samuel, Choir, Shark*. A brief description of each story follows.

For the investigation, *Opens* represents task complexity level one. This story depicts a woman approaching a door leading to a stairwell, apparently in an office building. She tries to open the door by pulling on the doorknob, but is unable to do so. She pulls harder as another woman approaches from behind. Finally, the first woman gives up and is seen walking away as the second woman continues to approach the door. In the last picture the second woman is seen successfully opening the door by pushing it as the first woman observes over her shoulder.

Clean represents the second complexity level. This story shows a young man who brings his laundry to a laundromat for cleaning. It depicts him first putting the clothes into the washing machine and then transferring them to the dryer. He then folds the clothes and puts them into a laundry basket and finally is seen leaving the laundromat with the clothes clean and folded inside his basket.

Samuel represents complexity level three. The story depicts a man walking along a street carrying a bust of a woman in his arms. He is seen calling a taxi. The remaining pictures are shown from a point of view behind the taxi. The man is seen from behind sitting in the back seat of the taxi with his arm around the bust. The impression is that he is sitting with a real woman. He is then seen looking toward the bust, and then looking back over his shoulder toward the viewer. He is blushing. The final picture depicts the man and the bust seated on opposite sides of the back seat of the taxi with a large space in the middle separating them.

Choir represents the fourth level of task complexity. This story depicts a choir of ten people standing on a riser in two rows. Each member of the choir is holding a sheet of music in their hands. Each sheet of music has a note depicted on the front which allows the viewer to see that one man is holding his music upside down. The choir can be seen singing as the conductor directs them from his podium. The conductor apparently hears that something is wrong and moves along the choir trying to hear who is singing badly. He arrives at the man who is holding the music upside down and brusquely turns it right side up as the man appears surprised. The final picture depicts the conductor returning to his podium irritated. The other members of the choir can be seen discretely smiling while the man who caused the trouble appears to be embarrassed.

Shark represents the fifth and highest level of task complexity. The pictures are all shown from a perspective that is slightly raised as if the viewer is looking down toward the events that take place. The story depicts a young man carrying a surfboard who arrives at a crowded beach. He can’t surf because of the amount of people in the water. He leaves and is then seen at a surf shop buying a shark mask with a dorsal fin. The next picture shows the man swimming in the water toward the beach while wearing the mask. The other bathers notice the shark fin protruding from the water, panic and leave the beach. The final picture depicts the young man paddling out into the water on his surfboard carrying the shark mask with him. However, in the water there is a real shark approaching him of which he is unaware.

Each of the stories were presented to each subject in one of three random sequences in which at no point is the sequence between any of the stories repeated nor were any of the stories presented in the same sequential position more than once. The chosen sequences were as follows:

A: OPENS, CHOIR, SHARK, CLEAN, SAMUEL

B: SHARK, OPENS, SAMUEL, CHOIR, CLEAN

C: SAMUEL, CLEAN, CHOIR, OPENS, SHARK

The tasks were performed monologically as they were designed to be done as a part of the WAIS-III test. A monologic task was also chose over a dialogic task for the experiment as this could affect the outcomes. Gilabert, Barón, and Levkina’s 2011 study of the effects of manipulating task complexity over task types and modes has shown that L2 learners’ oral performance is greatly influenced by their interlocutors while performing dialogic tasks. To study the way that individual differences affect performance, it is necessary to eliminate the influence of others who may either compliment or detract from the subjects’ own capacities. A monologic task allows the subject to perform with minimal outside influence.

6.6.1 Pilot Study

A pilot study was carried out the purpose of which was twofold. Firstly, the goal was to confirm the practicality of the tasks and testing procedures as means for collecting data, and additionally, to observe subjects’ perceptions of the tasks in terms of complexity.

The picture arrangement task was taken from the Spanish version of the WAIS-III but the current experiment was designed as such that subjects would use English as a non-native language to carry out the tasks. Therefore, the pilot study was also used to determine whether use of English as an L2 would have would have any effect on the perceptions of the conceptual complexity of the tasks that were intended to be carried out by native Spanish speakers and thereby ensuring that any increase in cognitive load is due to task characteristics and independent from the language in which the task is being performed.

Two experimental means were used to independently measure task complexity in this study. A method of subjective time estimation and a subjective rating scale of communicative difficulty performed on a 9 point Likert scale were used to collect data describing subjects' perception of difficulty while performing the tasks.

Subjective time estimation as a means of measuring cognitive load has grown out of a dual-task methodology paradigm from experimental psychology based on the idea that as the load imposed by a primary task increases, limited attentional resources are drawn away from a secondary task resulting in a decrease in the quality of performance of that second task (Block *et al.*, 2010; Dzaak *et al.*, 2007; Fink & Neubauer, 2001; Fraisse, 1984; Hicks *et al.*, 1976; Paas *et al.*, 2003). The subjective estimation of time spent performing a primary task has been described as a reliable means of determining the cognitive load of that task the cognitive demands of which afford subjects greater or fewer resources available to estimate passed time accurately (Fink & Neubauer, 2001; Fraisse, 1984). Subjective time estimation has been adopted by investigators in the area of second language acquisition as a means to measure task complexity (Baralt, 2010; Gilabert & Baralt 2013; Michel, Gilabert, & Révész, in press). In the current study, the subjects were asked to estimate how much time they needed to narrate each story upon completion. This was compared to real time on task to determine an index of task complexity.

The subjective rating scale of communicative difficulty was based on that described by Paas *et al*. (2003), who used one as a measurement of cognitive load and labeled as per Robinson (2001a). As participants completed each task, they used a nine point Likert scale to rate it in terms of how hard or easy they found the tasks. Samples of the Likert scales can be found in Appendix I. One was done immediately after performing each of both the arrangement part of the tasks and the narrative part.

The investigator also wished to determine whether performance of the tasks in an L2 would have any effect on results since the pictures were taken from the version of the Wechsler test that was intended for L1 speakers of Spanish. In order to do so, the task was performed in three different modalities; by ten L1 English speakers, by ten L1 Spanish or Catalan speakers, and by ten L2 English speakers. This number was chosen in order to have a sufficient amount of data to study for each one of the three modalities. All subjects were adults not younger than 18 years old.

6.6.1.1 Pilot test procedure

The procedure that was followed during the study began with instructions as to how the task was to be done. Before actual data collection, subjects performed a practice run using the story named *Chase* taken from the same picture arrangement part of the WAIS-III. This was done to eliminate any affect that task familiarity might have had on tasks performed later in the experiment. The picture frames for each task had been put onto individual cards. For each task, the cards were set out on the table in front of the subjects with the illustrations facing up. The sequence of the stories presented to the participants were determined by the sequence that that particular participant was pre-assigned and the order in which the pictures were set out on the table is the same as that established by the manual of the Wechsler exam. The cards were hidden behind a blind. When the participants indicated that they were ready, the blind was removed and a timer started. The subjects arranged the pictures in order to form a story. When they were finished the timer was stopped and the time it took them to finish the task was recorded on a data sheet.

The next step was to ask the learner to estimate how long it took them to complete the task. This estimate was also recorded on a data sheet and compared to the time recorded by the timer for the subjective time estimation analysis. Subjects were given a difficulty perception questionnaire with the Likert scales on it. After giving and estimation of time, they rated the arrangement part of the task in terms of difficulty on the Likert scale.

Next, they were instructed to narrate the story according to how they arranged the pictures. Previous to the narration, however, the researcher clarified any vocabulary that they might have needed to do this. The narration was digitally recorded.

After the narration, the researcher asked them to estimate how long it took them to narrate the story and the reply was recorded on the data sheet for the subjective time estimation analysis. Subjects then rated the narration for difficulty on the Likert scale. Finally, the order in which the pictures were placed by that participant on the data sheet was also recorded for the researcher's reference.

6.6.1.2 Results of the pilot study

Data collected during the pilot study was analyzed as described below in order to determine whether the scale of complexity suggested by the Wechsler test would be appropriate for the current study. As a first step, statistics were analyzed to determine whether the language group used in each of the modules of the pilot study had an effect on the results. After it was determined that the language paradigm had no statistically significant effect of the results, the data was merged and analyzed as a whole.

6.6.1.2.1 Subjective time estimation

Values of estimated of time were compared to real observed time on task by means of an index created by dividing estimated values by observed values as carried out in such studies as Brown (1985) and Block, Hancock, and Zackay (2010). Brown describes the use of expressing such measures as proportions of the amount of time judged as standard practice and a means to ensure that all measures are on the same relative scale. Index values greater than 1 will indicate an over estimation of time spent on task and suggest a perception of less complexity.

In the end, only the time estimation data corresponding to the narration task was analyzed for the current study. The picture arrangement task has already been studied extensively and its analysis at this point would be redundant. Descriptive statistics for the time estimation data for the narration task can be seen in Table 6.

Table 6 – Descriptive statistics for time estimation data for narration (estimated time on task/real time on task in seconds)

Tasks	Language paradigm	N	Mean	Median	Minimum	Maximum	SD	Standard Error	Skewness	Kurtosis
OPENS	L2 ENG	10	1.18	1.06	0.88	1.76	0.30	0.09	1.02	0.05
	L1 ENG	10	1.44	1.35	0.81	2.42	0.50	0.16	1.14	0.65
	L1 SP/CAT	10	1.53	1.43	0.97	3.00	0.56	0.18	2.23	5.95
CLEAN	L2 ENG	10	1.04	1.01	0.76	1.62	0.24	0.07	1.56	3.36
	L1 ENG	10	1.38	1.27	0.88	2.00	0.39	0.12	0.28	-1.50
	L1 SP/CAT	10	0.93	1.00	0.71	1.07	0.13	0.04	-1.01	-0.30
SAMUEL	L2 ENG	10	0.91	0.87	0.62	1.25	0.22	0.07	0.40	-1.27
	L1 ENG	10	1.25	1.04	0.65	2.86	0.65	0.20	1.98	4.32
	L1 SP/CAT	10	0.91	0.98	0.45	1.19	0.25	0.08	-0.81	-0.63
CHOIR	L2 ENG	10	0.98	0.99	0.66	1.17	0.16	0.05	-0.63	0.05
	L1 ENG	10	1.15	0.10	0.61	2.25	0.56	0.18	1.39	0.80
	L1 SP/CAT	10	1.05	0.92	0.65	1.69	0.36	0.11	0.77	-0.62
SHARK	L2 ENG	10	0.71	0.76	0.35	0.92	0.19	0.06	-0.81	-0.33
	L1 ENG	10	0.74	0.76	0.45	1.09	0.19	0.06	-0.08	0.89
	L1 SP/CAT	10	0.85	0.91	0.38	1.46	0.30	0.10	0.45	0.90

Means for each of the language paradigms for each task were compared to determine the effect of language on the perception of task complexity. A one-way ANOVA was carried out for the data corresponding to the task Shark to determine if there was an effect of language on perception of task complexity. There was not a statistically significant effect of language on perception of task difficulty at the $p<.05$ level for the three language conditions [F(2, 27) = 1.10, p = 0.35].

For data corresponding to the tasks Clean and Samuel, distribution was assessed as not normal and violated the assumption of homogeneity of variance. Data was transformed using a base-10 logarithmic function which corrected both distribution and variance. A one-way ANOVA was performed on both to determine whether an effect of language on perception of task complexity existed. For the data corresponding to Clean there was a statistically significant effect of language on perception of task difficulty at the $p<.05$ level for the three language conditions [F(2, 27) = 7.29, p = 0.003]. For data corresponding to the task Samuel, there was not a statistically significant effect of language on perception of task difficulty at the $p<0.05$ level for the three language conditions [F(2, 27) = 1.94, p = 0.16].

For the data corresponding to Opens, distribution of the data was assessed as not normal. A Kruskal-Wallis test was performed to determine the effect for language on the data for this task between the three different language paradigms. No statistically significant effect for language was demonstrated (H(2) = 3.89, p = 0.14).

The data corresponding to the task Choir violated the assumption of homogeneity of variance. To determine whether an effect of language on perception of task complexity existed, a test of Welch’s ANOVA was performed.

For the data corresponding to Choir there was not a statistically significant effect of language on perception of task difficulty at the $p<.05$ level for the three language conditions [F(2, 14.70) = 0.54, $p = 0.60$].

Results of the time estimation data demonstrated that language had no effect on performance of most of the tasks. The only exception was for the task Clean which represented the second level of task complexity.

6.6.1.2.2 Subjective rating scale of communicative difficulty

Means of the values reported by subjects on the nine point Likert scale as a subjective measure of communicative difficulty were compared to determine the effect of language on the perception of task complexity. Descriptive statistics are shown in table 7.

Table 7 – Descriptive statistics for subjective perception of narrative difficulty – Likert Scale (1-9)

Tasks	Language paradigm	N	Mean	Median	Minimum	Maximum	SD	Standard Error	Skewness	Kurtosis
OPENS	L2 ENG	10	3.4	3	1	6	1.51	0.48	0.47	-0.17
	L1 ENG	10	3.5	3.5	1	7	2.12	0.67	0.28	-0.63
	L1 SP/CAT	10	2.7	2	1	5	1.49	0.47	0.83	-0.64
CLEAN	L2 ENG	10	4.7	5	3	8	1.57	0.50	1.08	0.63
	L1 ENG	10	3.9	3.5	1	8	2.77	0.88	0.27	-1.23
	L1 SP/CAT	10	2.2	1.5	1	7	1.87	0.35	1.11	3.39
SAMUEL	L2 ENG	10	5	4	2	9	4.49	0.79	0.62	-0.98
	L1 ENG	10	4.1	3.5	1	8	2.81	0.89	0.42	-0.95
	L1 SP/CAT	10	3.5	3	2	8	2.17	0.69	2.10	-0.34
CHOIR	L2 ENG	10	5	5.5	3	7	1.41	0.45	-0.38	-0.90
	L1 ENG	10	4.2	4	1	8	2.57	0.81	0.40	-0.67
	L1 SP/CAT	10	2.8	3	1	7	1.81	0.57	1.74	1.69
SHARK	L2 ENG	10	4.9	5	2	8	1.79	0.57	0.24	-0.09
	L1 ENG	10	4.7	5	1	9	2.83	0.90	-0.08	-0.91
	L1 SP/CAT	10	4.1	3	2	8	2.08	0.66	1.22	-0.18

Separate one-way ANOVAs were carried out for the data corresponding to the tasks Opens, Choir, and Shark to determine if there was an effect of language on perception of task complexity. For data corresponding to the task Opens, there was not a statistically significant effect of language on perception of task difficulty at the $p<.05$ level for the three language conditions [F(2, 27) = 0.63, p = 0.54]. For data corresponding to the task Choir, there was also not a statistically significant effect of language on perception of task difficulty at the $p<0.05$ level for the three language conditions [F(2, 27) = 3.12, p = 0.06]. For data corresponding to the task Shark, there was also not a statistically significant effect of language on perception of task difficulty at the $p<0.05$ level for the three language conditions [F(2, 27) = 0.33, p = 0.72].

For the data corresponding to Samuel, distribution of the data was assessed as not normal. Therefore, a Kruskal-Wallis test was performed to determine the effect for language on the data for this task between the three different language paradigms. No statistically significant effect for language was demonstrated (H(2) = 2.49, p = 0.29).

The data corresponding to the task Clean violated the assumption of homogeneity of variance. To determine whether an effect of language on perception of task complexity existed, a test of Welch’s ANOVA was performed. For the data corresponding to Clean there was a statistically significant effect of language on perception of task difficulty at the $p<.05$ level for the three language conditions [F(2, 16.38) = 10.93, p = 0.001].

The subjective rating scale of communicative difficulty rated the task Clean as the only task which may be affected by the language which subjects use to perform it.

6.6.2 Discussion of the pilot study results

Of all the language paradigms, a statistically significant effect of language on the perception of difficulty was found only for the task Clean. This was true for both the time estimation measure as well as the Likert scale scores.

Table 8 – Multiple Comparisons – difference between means for the task Clean

Measure	Difference L2 English vs. L1 English	Difference L2 English vs. L1 Spanish/Catalan	Difference L1 English vs. L1 Spanish/Catalan
Time Estimation	-0.34*	0.11	0.45*
Likert scale	0.80	2.80*	2.00*

*$p < 0.05$

Examination of a multiple comparisons analysis of data from the task Clean provided in table 8, reveals that both measures correspond in showing a significant difference between L1 speakers of English and L1 speakers of Spanish/Catalan as they performed the task in their native language. However, a close look at the mean scores in tables 2 and 3suggest some inconsistencies in the data. According to the scores on the Likert scale, of the three language paradigms, L1 speakers of Spanish/Catalan are those subjects which least found the task complex to perform while the data from the time estimation measure indicates that these same subjects most found the task complex to perform. Other comparisons also indicate some inconsistencies between the two measures for the task Clean. While the time estimation data shows a significant difference between the L2 English and L1 English paradigms, the data from the Likert scale does not, and a comparison between L2 English and L1 Spanish/Catalan paradigms show significant difference between scores on the Likert scale where the time estimation data does not. Nonetheless, there is a tendency for the subjects within the L2 English paradigm to find the task more complex than the subjects within the L1 English paradigm, albeit significantly so only according to time estimation measures.

The data suggests that the primary difference in the case of the task Clean exists between native English speakers and native Spanish/Catalan speakers as they perform the task in their respective L1s, but evidence that any increase in cognitive load imposed on native Spanish/Catalan speakers due to their performance of the task Clean in L2 English is not definitive. The data shows that perceived task complexity may be attributed to task characteristics.

Based on the above, it may be concluded that use of English as an L2 to perform any of the experimental tasks as presented in the current study is not expected to have an effect on the perceptions of the conceptual complexity of the tasks while carried out by native Spanish/Catalan speakers. Any increase in cognitive load is due to task characteristics and is independent from the language in which the task is being performed.

6.7 Scale of complexity of oral narration task

As previously stated, the picture arrangement tasks chosen for the current study represent a continuum of conceptual complexity as administered by the Wechsler Adult Intelligence Scale III. The following section will demonstrate how this same scale of complexity transfers to the tasks as they are administered as oral production tasks. The scale is represented in table 9 below.

Table 9 – Scale of conceptual complexity

Complexity	Simple	OPENS
	↓	CLEAN
		SAMUEL
		CHOIR
	Complex	SHARK

As language was determined not to be a factor exerting influence on the perception of task complexity as discussed in the previous section, the data from all subjects across language paradigms was merged for further analysis to determine that the sequence of the tasks in order from greater to lesser level of conceptual complexity can be maintained as suggested by the Wechsler test. Descriptive statistics of the merged data corresponding to the time estimation index is shown in table 10 and that of the Likert scores is shown in table 11.

Table 10 – Descriptive statistics for merged data for the time estimation index

Tasks	N	Mean	Median	Minimum	Maximum	SD	Standard Error	Skewness	Kurtosis
OPENS	30	1.38	1.32	0.81	3.00	0.48	0.09	1.7841	3.8789
CLEAN	30	1.12	1.02	0.71	2.00	0.33	0.06	1.308	1.000
SAMUEL	30	1.02	0.96	0.45	2.86	0.44	0.08	2.7263	10.5228
CHOIR	30	1.06	0.97	0.61	2.25	0.34	0.07	1.6933	3.0243
SHARK	30	0.77	0.77	0.35	1.46	0.23	0.04	0.5089	1.4987

Table 11 – Descriptive statistics for merged data for the Likert scale

Tasks	N	Mean	Median	Minimum	Maximum	SD	Standard Error	Skewness	Kurtosis
OPENS	30	3.20	3.0	1.0	7.0	1.71	0.31	0.4203	-0.6777
CLEAN	30	3.50	3.0	1.0	8.0	2.22	0.41	0.5035	-0.8324
SAMUEL	30	4.20	3.0	1.0	9.0	2.50	0.46	0.5964	-1.0875
CHOIR	30	4.00	4.0	1.0	8.0	2.13	0.39	0.2511	-0.8804
SHARK	30	4.56	5.0	1.0	9.0	2.22	0.41	0.2078	-0.9463

In order to compare the data and the sequencing order suggested by the Wechsler test, each of the tasks was assigned a number from 1 to 5. Number 1 was assigned to the task considered to be of the lowest level of complexity and each successive number corresponding an increment of one level of complexity as illustrated in table 12.

Table 12 - Succession of tasks per complexity level as per Wechsler test

Complexity Level	Task
1	OPENS
2	CLEAN
3	SAMUEL
4	CHOIR
5	SHARK

First, the means of the data from the time-estimation index and the Likert scores corresponding to each of the levels of cognitive complexity represented by each of the tasks were compared to determine a correlation between the two kinds of measures. A strong negative correlation between the means from the time-estimation data and the Likert scores was shown, r(4)=-0.97, *p*=.007.

The merged data was analyzed using a one-way ANOVA to determine whether differences in the level of conceptual complexity expected to be imposed on subjects performing the tasks exists between the tasks. The merged time estimation data did not exhibit normal distribution, so it was transformed using a base-10 logarithmic function which corrected distribution. The test suggested a significant difference between the means of each of the tasks representing distinct complexity levels [F(4, 145) = 14.04, *p*<0.05]. Nonetheless, a multiple comparisons analysis showed that statistical significance did not exist between all of the tasks individually. These include tasks representing the second and third levels of complexity (Opens vs. Samuel), the second and fourth levels of complexity (Opens vs. Choir), nor the third and fourth levels of complexity (Samuel vs. Choir).

A one-way ANOVA was performed on the data corresponding to the Likert scale. There was not a statistically significant difference between the means at the $p<.05$ level for the five tasks [F(4, 145) = 0.90, $p = 0.11$]. A multiple comparisons analysis showed statistical significance only between the first and fifth levels of complexity (Opens vs. Shark).

Data was next examined in order to determine a correlation between increasing conceptual complexity levels of the tasks and the merged data from the time-estimation index as well as from the Likert scores. Data for the measures from the time estimation index was transformed using a base-10 logarithmic function which corrected distribution. A small yet significant negative correlation was found between complexity levels and the time estimation index, r(149) = -0.4902, p=0.0001. As expected, the result indicates that subjects tended to overestimate the time spent on the tasks as they increased in their conceptual complexity. This is supportive of the succession of task complexity as stated above. A negligible but significant positive correlation was found between complexity levels and the Likert scores, r(149) = 0.2087, p=0.01 and is confirmative of the succession of task complexity levels as afforded them by the Wechsler test.

6.8 Conclusions of the pilot study

The results of the pilot study confirmed that both the method chosen and the procedure followed were appropriate for collecting data which could be statistically analyzed for the main experimental phase.

Statistical analysis of the data collected during the pilot study suggested that any increase in cognitive load as operationalized by the various tasks is due to task characteristics and independent from the language in which the task is being performed. In addition, analysis of the data from the time-estimation index measure of cognitive complexity provides an indication that the levels of complexity may be for the most part significantly different with the exception of the most intermediate tasks.

The statistical analysis of the data collected during the pilot study was supportive, but not conclusive concerning whether the tasks do indeed represent a perfect continuum from simple to complex. At the same time, there is not sufficient evidence to suggest using any alternative sequencing of the tasks for the current investigation although a strong negative correlation between the means of the merged scores from both the time-estimation index and the Likert scores is indicative that both of the measures support the sequencing of the tasks as suggested by the Wechsler test.

The tasks have been demonstrated through extensive study over periods of decades to present distinct levels of difficulty for subjects as operalitionalized as a part of the Wechsler tests and the statistical analysis as described above is reflective of this. In addition, the tasks have previously been used in at least one study of task complexity (Robinson, 2007). Finally, the statistical analysis has demonstrated that they can be affectively used for collection of oral data without interference of the use of English as an L2 during task performance on perceived cognitive complexity.

6.9 Main study

Data collection for the main study was carried out within the installations at Blanquerna School of Communication Studies a Ramon Llull University in Barcelona and in the installations of the University of Barcelona. Details of how the experiment was carried out is described in detail below.

6.9.1 Participants

The subjects who volunteered to participate in the study were all university students from the Blanquerna School of Communication Studies at Ramon Llull University in Barcelona, and from the University of Barcelona. Eighteen subjects were male and twenty-nine were female with an average age of 19.87 years. Twenty students identified Spanish as their mother tongue, twenty identified Catalan as their mother tongue, and seven identified both Spanish and Catalan as being mother tongues. The average number of years that they claim to have studied English is 12.65 with a standard deviation of 3.3 years.

6.9.2 Working memory capacity

In order to determine how subjects were to be divided into groups of either high or low working memory capacity, scores drawn from the automatic reading span task were analyzed with a k-means cluster operation in SPSS. The data from each of the two groups that were formed was further analyzed with an independent samples t-test. It was determined that there was a significant difference between the data corresponding to both the high working memory capacity group (N=20, M=41.35, SD=7.44) and the low working memory capacity group (N=27, M=19.74, SD=6.22), t (45)=10.83, $p<0.001$, CI 95% [17.59, 25.63].

6.9.3 Attentional Capacity

Data collected from the alternating runs paradigm was used to calculate a value by which subjects were divided into one of two levels of attentional capacity. This value was determined by subtracting the recorded mean reaction times from the non-switch trials from those of the switch trials. The difference was then divided by the value recorded for the base, or non-switch trials and multiplied by one hundred. The result indicates how much more time was needed to perform the switch trials in terms of a percentage of the time needed to perform the non-switch trials.

Those whose performance demonstrated a lower switch cost were considered of high attentional capacity, and those who demonstrated greater switch cost were considered of low attentional capacity.

A k-means cluster operation was performed in SPSS to separate data into two groups. An independent samples t-test was on the data corresponding to each of the two groups. It was determined that there was a significant difference between the values for the high attentional capacity group (N=29, M=9.68, SD=50.86) and the low attentional capacity group (N=18, M=174.10, SD=73.40), t (45)=-9.10, $p<0.001$, CI 95% [-200.90, -127.93].

It must be stated that while all groups were formed based on the relation of their test scores to common means, subjects did, nevertheless, form a continuum of attentional and working memory capacity levels.

6.9.4 Proficiency level

Based primarily on the OQP test scores, subjects were divided into B1, B2, C1 and C2 levels as per the Council of Europe’s Common European Framework of Reference for Languages. Data corresponding to subjects who scored less than a B1 level on the test was discarded. Of the 56 subjects from whom data was collected, data from 9 of those were discarded based on these criteria. It was determined that subjects whose level was below that of B1 provided recorded performance during the narration task that was insufficient for proper analysis.

The results on the Lex tests were used in two different ways. The first was to confirm proper placement of subjects whose scores demonstrated to be close to the cut-off point between B1 and B2 levels. These cases included situations in which a level change would be determined by very few points difference on the Oxford Quick Placement test, but where the LEX scores suggested that a level change was appropriate. Subjects were considered outliers whose OQP test scores suggested their proficiency level to be low, but were nonetheless greater than the upper limit of the 95% confidence interval for the group and concurrently whose LEX scores were also greater than the upper limit of the 95% confidence interval for that test. These were subsequently placed into the ‘high’ level group. This affected three subjects. Likewise, students who would otherwise have been placed into the high level group but whose scores were less than the lower limit of the 95% confidence interval for the group for both the OQP test and LEX test were transferred into the ‘low’ proficiency level group. This affected one subject. Descriptive statistics can be viewed in table 13.

Table 13 – Descriptive statistics for LEX and QPT proficiency measures divided into categories as per high and low OQP scores.

Measure	Proficiency	N	Mean	Std. deviation	Minimum	Maximum	95% CI
LEX proficiency scores	Low	24	5175.00	712.01	3850	6250	4874.35, 5475.6
	High	23	6052.17	949.70	4350	7450	5641.49, 6462.85
OQP test scores	Low	24	26.79	2.65	20	30	25.67, 27.91
	High	23	42.17	8.80	31	58	38.37, 45.98

Secondly, in order to confirm that the two definitively formed groups did indeed represent populations of distinct proficiency level, an independent-samples t-test was conducted on the Lex scores for subjects of both groups. There was a significant difference in the scores between the high level group (N = 25, M = 6124.00, SD = 861.34) and low level group (N = 22, M = 5013.64, SD = 628.90); t (45) = -4.99, $p < 0.001$, CI 95% [-1558.86, -661.87]. These results suggest that subjects had been divided into two distinct groups of significantly different proficiency levels. It must be stated that while the two groups were formed based on the relation of their test scores to common means, subjects did, nevertheless, form a continuum of proficiency levels.

6.10 Testing Procedure

The procedure for administering the tests during the data collection sessions followed the following protocol.

Once a subject agreed to participate, a time and place was arranged to meet in order for the tests to be administered. Sessions took an average of approximately one and a half hours to complete. At the beginning of the session, subjects received a consent to participate form which they read. If they agreed to the conditions proposed in the form, the researcher asked them to sign two copies. One was given to them for their own records, and the second was kept by the researcher. Copies of the Spanish and Catalan forms are available for review in Appendix II. Next, an affective variables questionnaire was given to be completed. The purpose of the questionnaire was twofold. It allowed for the collection of useful information for the analysis of the data. Secondly, it functioned to as a vehicle to begin small talk before actual testing took place in order to develop rapport with the subject. A copy of this form is available for review in Appendix III. Subjects received a copy of the Oxford Quick Placement test which they were to complete on their own time outside of the testing session. They were given instructions to spend no more than thirty minutes to complete the test. They were also asked to sign a statement at the front of the form by which they committed to completing the test without use of any reference material or outside aid. A copy of this page of the test is available for review in Appendix IV. Once the test was completed they were to return it to the researcher for evaluation.

The next step was to administer the computer based tests. As these tests were sometimes relatively long, a short break was made between each in order to mitigate an effect of fatigue for the subject. Small talk was made and a short explanation of the test which they just completed and its function was given before the next test was administered. The first cognitive test to be administered was the automated reading span task. This test needed approximately thirty minutes to complete. For this reason, it was completed first. Following the reading span task, both the X-Lex and Y-Lex tests were completed, and finally the alternating runs paradigm test. This first stage of the session took approximately forty-five minutes to one hour to complete.

The next part of the session consisted in the picture arrangement task and story narration. Subjects were first given instructions on what to do. Pictures would be placed in front of them for them to arrange in such a way that they felt a story could be narrated. The stories were presented to the subject in one of three possible random sequences as explained in the description of the pilot study. Pictures corresponding to each story were presented in the same order as that recommended in the instructions for their administration as part of the WAIS-III tests. This order did not change between subjects. Subjects were not given a time limit which allowed for planning. Once the subjects arranged the pictures in the order that they preferred, the researcher asked them if they needed any specific vocabulary which they felt would be necessary in order to complete the story.

The first story was intended as a practice run. The subjects completed the procedure exactly as if they were performing for data collection, but the researcher did not record them. Every story after the first was digitally recorded for subsequent analysis. The session ended after the five stories were narrated and recorded.

6.11 Measures

The current study intends to determine how dimensions of linguistic complexity, accuracy, and fluency interact as task demands are increased along the resource-directing variable of +/- intentional reasoning demands. Eighteen different measures were calculated in total to try to cover all CAF dimensions as suggested by Housen & Bulté (2012).

6.11.1 Fluency measures: Speed

For measures of speed fluency, Rate A was calculated by counting the number of total syllables used during each task following Segalowitz (2010). This number was divided by the total amount of performance time spent on the task in seconds, and then multiplied by sixty to calculate the number of syllables of unpruned in speech per minute.

6.11.2 Fluency measures: Breakdown fluency

A second measure of breakdown fluency was calculated by means of a ratio of filled pauses occurring between analysis of speech unit (ASU, or AS-unit) boundaries. AS-units were defined following the guidelines set by Foster, Tonkyn and Wigglesworth (2000). This measure was decided upon according to Skehan

(2009) who explained that the AS-unit boundary is a natural place for native speakers to pause in order to undertake online planning. Non-native speakers, on the other hand, pause more frequently where it is less natural to do so resulting in more haltered speech. Tavakoli (2011) explains that pausing is an indication that a subject is performing a lexical search, or is concerned about accuracy of structure, or pronunciation, or to plan for speech. Such processes involve the Formulator and Conceptualizer stages of Levelt's model and require mental effort as subjects review and possibly need to repair speech in order to meet task demands. The greater mental effort required for these processes under increasingly complex tasks as they are performed by non-native speakers is expected to be reflected in the measure manifested by more pausing as task complexity increases.

6.11.3 Accuracy measures: Errors

It was predicted that speaker-external forces would direct attention to distinct pragmatic requirements of the task which could be met through linguistic aspects of lexical and syntactic choice, and that accuracy measures along these lines would be appropriate to study. In addition to the number of total errors per AS-unit, errors were broken down into three distinct categories: lexical, morpho-syntactical, and 'other', for remaining errors. The category of other errors included cases of pragmatic errors, wrong expressions, or superfluous words. In other words, they were any errors which were categorized as neither lexical nor morpho-syntactical errors. Ratios per AS-units were calculated for each following Gilabert (2007). This was done in an attempt to determine how subjects' focus on form may have been affected through lexical choice and in the grammaticalization of the conveyed message both processes corresponding to the Formulization stage of Levelt's model.

6.11.4 Accuracy measures: Repairs

Self-repairs are problem solving techniques related to the perception that one has concerning deficiencies in one's own production (Gilabert, 2007). Monitoring one's own speech is accounted for in Levelt's model of speech production but is limited by attentional capacity. An increase in task demands may direct resources from monitoring with a negative affect for accuracy. So, investigating error repairs was expected to offer insight into how subjects focused attentional resources during their narrations. Repair calculation focused on both error repairs and non-error repairs. Error repairs were concerned with instances of self-correction of lexical and morpho-syntactical errors. Non-error repairs were of two types. A *different information repair* is concerned with the monitoring of an error resulting in the realization that the content of the pre-verbal plan associated with the Conceptualizer stage of Level's model needs to be modified resulting in a reconceptualization of the message that the speaker wants to transmit. An *appropriacy repair* is also involved in changing the pre-verbal plan but while the message remains the same, it must be re-worded. For example, an appropriacy repair occurs when the message is expressed in pragmatically inappropriate language, or in language whose message is incoherent (Kormos, 2006).A ratio of repaired to unrepaired errors was measured by dividing the number of repaired errors by the number of unrepaired errors. As per Gilabert (2007), a corrected ratio of repair compensates for differences in text length which may account for large differences in the number of errors. The corrected ratio is calculated by dividing the number of error repairs by the square root of twice the number of errors. This measure was also taken into account.

6.11.5 Complexity: Structural

Structural complexity measures included measures of subordinate as well as coordinate clauses per AS-unit, words per AS-unit as a sentence level measure, and words per clause as a clausal level measure. These are standard measures in the literature and have been recommended by Norris & Ortega (2009).

6.11.6 Complexity: Lexical

Two measures of lexical complexity were used in order to cover both lexical diversity and sophisitication. As a measure of lexical richness, D was employed. This value was calculated using the software program D-Tools developed by Meara & Miralpeix (2007).

Lambda as a measure of lexical sophistication was calculated using a web based program call P-Lex (Meara & Bell, 2001) available on lognostics.com. According to Skehan (2009), the comparison between structural complexity indexed by subordination and values for lambda as a measure of lexical richness provides a means to explore the relationship between lexis and syntax. For native speakers, making more demanding and less obvious lexical choices does not hinder syntax, but rather drives it as they may find the need to use less common vocabulary which is more likely to require more complex syntactic accompaniment. Non-native speakers are likely to find that making more complex lexical choices results in more effort spent during lexical retrieval resulting in a syntax which is less complex as well as less accurate.

Skehan (2009) observed that native speakers demonstrated a positive correlation between lambda and structural complexity indexed by subordination while for non-native speakers this correlation is negative. This will be performed in order to investigate what influence lexis may have on syntax as subjects perform the tasks.

6.11.7 Complexity: Propositional

Propositional complexity was measured in idea units per task. Identification of idea units followed the guidelines described in Vasylets, Gilabert, and Manchón (forthcoming). Both semantic and intonational criteria are considered in identifying idea units. Semantically, as defined in Vasylets, Gilabert, and Manchón, an idea unit is a “meaningful, semantically integral chunk of discourse”. Intonationally, idea units form a single intonation contour, bounded by some kind of hesitation and ending in a clause final intonation. The idea unit may be a clause, or a part of a clause (Chafe, 1985). As per its definition, an idea unit will either be smaller that, but not longer than an AS-unit. The guidelines in Vasylets, Gilabert, and Manchón (forthcoming) were interpreted in the following way for the present investigation. Clauses with a coordinate relationship were considered separate idea units. In the examples below, idea units are separated by a double slash (//).

[1] *at that moment he's folding a towel // and putting them again to that basket.*

Subordinate clauses were considered separate idea units if there was a week relationship between them. Indicators include non-restrictive relative pronouns, clauses which began with *while*, *because*, *although*, *as for,* or *since*. Also, adverbial clauses placed at the beginning of a phrase would be considered a separate idea unit.

[2] *they are playing volleyball into the water // because maybe they are on his free days.*

Where there was a stronger conceptual connection between a main and subordinate clause, only one idea unit was counted. Vasylets, Gilabert, and Manchón refer to these as *extended idea units*. Indicators included restrictive relative pronouns, clauses beginning with *that*, complement clauses beginning with *to*, and adverbial clauses in a final position. The examples below illustrate phrases counted as single idea units.

[3] *we can appreciate that he's well dressed*

[4] *and jumps into the surfboard to start practicing*

As per Chafe (1985), an idea unit roughly corresponds to the amount of information that one may hold at any given moment in short term memory; a phrase of approximately seven words, or an idea which may be held for about two seconds. This, of course, is subject to variation. According to Ellis and Barkhuizen (2005), calculating the number of idea units provides a measure of the extent that a speaker engages in conceptualization as the content of narration is expressed. More idea units would be an expression of reversion to the Conceptualizer stage of Levelt's model for creating ideas to be encoded, an attentional resource depleting function as mentioned earlier.

6.12 Statistical Instruments

Statistical analyses were carried out using SPSS Version 15.0. Descriptive statistics for all measures of linguistic complexity, accuracy and fluency at each of the five levels of task complexity are provided (tables 39-44). Medians and the measure of interquartile range are given due to the non-parametric nature of most of the data. Means and standard deviations are provided where appropriate. Normality of distribution was determined by means of Shapiro-Wilk tests and homogeneity of variance by means of Levene's Test for Equality of Variances. Data which violated the assumption of homogeneity of variance or normality of distribution was rank order transformed and non-parametric statistical tests were performed for analysis.

Tests were all performed for data that was not split into groupings of high and low levels of proficiency, and attentional and working memory capacity. This is referred to in most cases as the 'non-split' data. Tests were also run on the data that was split into these groupings. This is referred to in the dissertation as 'split' data. The differentiation was made for the purpose of comparison.

Correlations were run for all measures against ordinal values ascribed to each level of task complexity in order from simple to complex. This was done for both the split and non-split data to identify differences in performance between the different groupings of subjects between measures of CAF as task complexity increased.

Next, a series of repeated measures procedures were performed in order to identify significant differences in measures between the various task complexity levels. Friedman's tests were run for non-parametric data and a repeated measures ANOVA was used where data was parametric. In addition, paired samples tests were performed between measures at each level of task complexity. For non-parametric data, these were performed using a Wilcoxon signed-rank test and for the parametric data paired samples t-tests were run.

To help find possible trade-off effects between CAF measures, correlation analyses were performed between each of the measures at each of the task complexity levels in order to determine at which point along the complexity continuum measures appeared to be influenced in similar ways as a result of increased task complexity. Where patterns in correlations were found, comparisons were made with measures which demonstrated significant differences between levels as well as with paired samples tests to find differences in how each group performed at each task complexity level.

Outliers were not removed in considering them an integral part of the data. Data for a small number of participants was found to be missing for some measures in which cases the null value was substituted by an average value calculated from available data. Significance levels were set at $\alpha = .05$ except where otherwise indicated. The CA mode of CHILDES (MacWhinney, 2000) was used for the calculation of items (e.g. words or tags) in the transcripts. Reliability was determined by means of an interrater method. This was carried out with the help of experienced researchers which were provided with written instructions as to the protocols to be carried out. An intrarater method was used for the calculations of repairs in which the researcher counted data a second time for comparison. Rate A. Measures for D, and Lambda are computer generated values. Mean percentage rate of interrater agreement out of a randomly selected sample of 10% can be seen for each measure in table 14.

Table 14 - Interrater reliability: Mean percentage of rater agreement

	Dependent variables	N	Mean
Fluency	▪ Filled pauses per AS-unit	25	90.29
	▪ Rate A – syllables per minute	25	100
Accuracy	▪ Morpho-syntactical errors per AS-Unit	25	91.79
	▪ Lexical errors per AS-Unit	25	82.34
	▪ Other errors per AS-Unit	25	97.56
	▪ Total Errors per AS-Unit	25	89.39
	▪ Error repairs per AS-Unit	25	94.19
	▪ Appropriacy repairs per AS-Unit	25	83.33
	▪ Different repairs per AS-Unit	25	85.00
	▪ Repaired to unrepaired errors per AS-unit	25	99.95
	▪ Repaired to unrepaired errors per AS-unit (corrected)	25	100
Lexical Complexity	▪ D-Value	25	100
	▪ Lambda	25	100
Structural Complexity	▪ Words per AS-Units	25	80.19
	▪ Subordinate clauses per AS-unit	25	92.40
	▪ Coordinate clauses per AS-unit	25	90.82
	▪ Words per clause	25	92.18
Propositional Complexity	▪ Idea units per task	25	86.19

6.13 Summary

This chapter reviewed the methodology and protocols which were followed during the experimental phase of the present research. Tests which were administered were described in detail as well as the statistical methods which were employed during the analysis phase of the research. Details of the results of the analysis will be described in the following section.

CHAPTER 7: RESULTS

7.1 Introduction

This chapter presents results of the statistical analysis performed on quantitative data collected from subjects. In each case, data was analyzed in two ways; first it was analyzed in a ‘non-split’ format where no distinctions were made between high and low levels of proficiency, and attentional and working memory capacity. Then it was analyzed in a ‘split’ format where the distinction was made. This way, trade-offs effects that exist in the data can be identified and compared between groups of subjects so that differences between them can be analyzed.

First, the results of correlations between dimensions of CAF and task complexity are presented. This will be presented first for non-split data. Then results of the split data will be presented. Secondly, the results of the repeated measures tests will be given. This will also be presented for non-split data first, and then for the split data. Finally, the results of the analysis of the correlations between CAF measures will be presented. The results of these correlations for both split and non-split data will be compared jointly. Results will be compared and discussed in the following chapter.

7.2 Correlations between CAF measures and task complexity

In order to carry out the correlation process, each level of task complexity was assigned a value between 1 and 5. The simplest task was assigned the value of 1 and the most complex, the value of 5. Because of the ordinal nature of the variable for task complexity, measurement data corresponding to the dependent variables was ranked, and Spearman’s Rank-Order correlation was used.

7.2.1 Non-split data

Forty-seven subjects performed each of the five separate tasks one time which resulted in n = 235 total observations. Table 15 shows data for measures for which significance was demonstrated.

Table15 – Spearman's Rank-Order correlation test between measures and values of task complexity

		Lexical complexity		Structrual complexity		Fluency
		Lambda	D	Words per clause	Words per ASU	Rate A
Task Complexity	Correlation Coefficient	.558(**)	.284(**)	.203(**)	.181(**)	-.379(**)
	Sig.	.000	.000	.002	.005	.000
	n	235	235	235	235	235

*correlation significant at *p*<.05
**correlation significant at *p*<.01
n = total observations

7.2.1.1 Fluency

Of the two fluency measures, mid-ASU pauses and rate A, only rate A demonstrated a significant negative correlation with task complexity. As tasks became more cognitively complex, subjects produced fewer syllables per minute.

7.2.1.2 Accuracy

None of the measures of accuracy demonstrated any correlation whatsoever with increased task complexity.

7.2.1.3 Linguistic complexity

Both measures of lexical complexity showed positive correlations with increased task complexity (table 15). This shows that increased task complexity promoted a greater variety of vocabulary and the use of less common words.

Two measures of structural complexity; words per clause and words per AS-unit also demonstrated a positive correlation (table 15). This demonstrates that increased task complexity pushed subjects to speak in structurally more complex ways.

7.2.2 Split data

Up to now, the data shows that as task complexity increased, fluency decreased, accuracy showed no change, and linguistic complexity increased in some aspects. This is supportive of Skehan’s Trade-off Hypothesis which predicts that as task complexity increases, we can expect fluency to decrease while attentional resources are allocated either to accuracy or linguistic complexity, but not to both simultaneously.

The question that remains, however, is whether these results can be generalized, or if subjects who differ in their levels of proficiency, and attentional and working memory capacity will show that they perform the same tasks in a different way.

The same correlations as described above were run again. This time, however, the data was split according to levels of proficiency, and attentional and working memory capacity. Table 16 only shows the information for measures which demonstrated significance. The data shows six coefficients per measure; one coefficient for each level (high and low) of the three groupings (proficiency, attentional capacity, working memory capacity). Data shows the coefficient, significance, and n=number of total observations from each group.

Table 16 – Spearman's Rank-Order correlation test between measures and values of task complexity

			Lexical complexity		Structrual complexity		Fluency
			Lambda	D	Words per clause	Words per ASU	Rate A
Proficiency	Low	Correlation Coefficient	.620(**)	.214(*)	.135	.149	-.412(**)
		Sig.	.000	.025	.160	.121	.000
		n	110	110	110	110	110
	High	Correlation Coefficient	.514(**)	.347(**)	.274(**)	.215(*)	-.390(**)
		Sig.	.000	.000	.002	.016	.000
		n	125	125	125	125	125
Attentional Capacity	Low	Correlation Coefficient	.56(**)2	.285(**)	.30(**)3	.131	-.393(**)
		Sig.	.000	.000	.006	.248	.000
		n	80	80	80	80	80
	High	Correlation Coefficient	.599(**)	.285(**)	.201(*)	.163(*)	-.370(**)
		Sig.	.000	.000	.012	.042	.000
		n	155	155	155	155	155
Working Memory Capacity	Low	Correlation Coefficient	.614(**)	.267(**)	.189(*)	.147	-.335(**)
		Sig.	.000	.002	.028	.088	.000
		n	135	135	135	135	135
	High	Correlation Coefficient	.485(**)	.316(**)	.232(*)	.224(*)	-.520(**)
		Sig.	.000	.001	.020	.025	.000
		n	100	100	100	100	100

*correlation significant at *p*<.05
**correlation significant at <.01
n = total observations

7.2.2.1 Fluency

The breakdown fluency measure of mid-AS-unit pause position showed no correlation with increased task complexity. On the other hand, rate A fluency measurements showed significant negative correlations with task complexity for all groups (table 16).This is supportive of both the Trade-off Hypothesis and the Cognition Hypothesis which both predict that as task complexity increases, fluency will decrease.

7.2.2.2 Accuracy

Only one accuracy measure, the corrected repaired to unrepaired errors index, showed significant positive correlations. This was true for three of the groupings: The low attentional capacity group showed a significant correlation, Spearman's rho (80) = .229, *p* <.05, as well as the high working memory capacity group Spearman's rho (100) = .212, *p* <.05, and the high proficiency level group, Spearman's rho (125) = .183, *p*<.05.

This is in contrast with the non-split data described above which showed no correlation whatsoever with any of the accuracy measures. Results for the split data lend partial support to Robinson's Cognition Hypothesis which predicts that as task complexity increases, fluency will decrease as attentional resources are allocated to both accuracy and linguistic complexity. To this end, some measures of linguistic capacity also demonstrated significant correlations as explained below.

7.2.2.3 Linguistic complexity

As seen in table 16 both measures of lexical complexity, D and lambda, showed positive correlations in all cases. In addition, two measures of structural complexity showed correlations. Words per clause also demonstrated weak but significant correlations in all cases except for the low proficiency group. The measure of words per AS-unit demonstrated significant correlations with task complexity in all cases for the high level proficiency, working memory and attentional capacity groups while the low level groups did not. Other measures of structural and propositional complexity remained unaffected.

A Fisher’s r-to-z transformation was performed for each of the correlation coefficients where significance was found. This was done in order determine whether high and low proficiency, and working memory and attentional capacity groups performed with greater or less intensity from one another. None of the groups exhibited significant differences from their counterparts.

7.2.2.4 Summary of correlations between CAF measures and task complexity

The results of the correlations between task complexity and dimensions of CAF are supportive of the second hypothesis of this thesis which states that differences in working memory and attentional capacity may determine the clarity with which subjects manifest differences in how they perform tasks. When no distinctions were made between individual differences of proficiency, and attention and working memory capacity, there was evidence that attentional resources were directed to structural complexity to the detriment of fluency, but that accuracy remained unaffected. This follows prediction of the Trade-off Hypothesis. When individual differences were taken into account, evidence that measures of both accuracy and linguistic complexity increased to the detriment of fluency became apparent or some of the groups, but not for all. This offers some support for the Cognition Hypothesis. Nevertheless, since only one accuracy measure experienced growth along dimensions of task complexity, more structurally complex forms may have been elicited by task characteristics particular to some of the more complex tasks rather than drawn out by an actual increase in cognitive complexity of the tasks. These results recall the dual increase of measures of accuracy and syntactic complexity found in experiments performed by Tavakoli and Skehan (2005), and Tavakoli and Foster (2008; 2011), and Foster and Tavakoli (2009) which Skehan (2009) attributed to cumulative effects of task demands.

7.3 Comparisons between task complexity levels

Repeated measures tests were used to investigate where CAF measures showed significant differences between task complexity levels. Tests were first run on non-split data. Then the tests were run on split data. Tables 39-44 provide descriptive statistics for all measures per each of the five levels of task complexity. Data was analyzed using the Shapiro-Wilk test of normality (tables 45-48).

7.3.1 Fluency: Non-split data

A repeated measures ANOVA showed that increased task complexity had a strong significant effect on the rate A fluency measure for non-split data, $F(4,46) = 17.50$, $p = .000$, $\eta_p^2 = .400$.

Table 17 shows results of the test of within-subjects contrasts. There were significant differences between all levels except for between levels four and five.

Table 17 - Tests of within-subjects contrasts for rate A fluency measure (non-split data)

	Sum of Squares	df	F	Sig.	η_p^2
Level 1 vs. Level 2	17756.11	1	17.50	.000**	.276
Level 2 vs. Level 3	5078.74	1	7.12	.010*	.134
Level 3 vs. Level 4	13653.47	1	21.11	.000**	.315
Level 4 vs. Level 5	1272.65	1	1.22	.275	.026

*α significant at $p<.05$
**α significant at $p<.001$

7.3.2 Fluency: Split data

A repeated measure ANOVA showed that increased task complexity had a significant effect for both high and low level proficiency, and attention and working memory capacity groups on measures of rate A fluency between different levels of task complexity (table 18).

Table 18 – Repeated measures ANOVA for Rate A (split data)

	Level	N	Sum of the squares	df	*F*	*p*-value	η_p^2
Proficiency	Low	22	30247.03	4	15.25	.000**	.412
	High	25	36330.77	2.99	15.46	.000** [a]	.392
Attentional Capacity	Low	16	24644.13	4	10.21	.000**	.403
	High	31	41865.60	4	12.95	.000**	.404
Working Memory Capacity	Low	20	38957.56	4	19.06	.000**	.501
	High	27	28865.20	3.09	12.94	.000** [a]	.332

*α significant at p<.05
**α significant at p<.001
a: Greenhouse-Geisser

The test of within-subjects contrasts brings to light patterns of significantly different means between levels of task complexity for the different levels of proficiency, and working memory and attentional capacity (table 19). None of the groups showed significance in the difference between means of task complexity levels four and five while all showed significance between levels three and four. There is a distinctive pattern in significantly different means between levels one and two, and two and three. The low working memory capacity group showed significant differences in all cases apart from between levels four and five. However, the low proficiency and low attentional capacity group both showed significance between task complexity levels two and three, as well as between levels three and four. In contrast, all three of the higher level groups followed the same pattern in showing significant differences between levels one and two, but not between levels two and three.

Table 19 - Tests of within-subjects contrasts for Rate A fluency measures (split data)

		Task Complexity	Sum of Squares	df	F	Sig.	η_p^2
Proficiency	Low	Level 1 vs. Level 2	4642.92	1	2.842	.107	.119
		Level 2 vs. Level 3	3784.33	1	4.445	.047*	.175
		Level 3 vs. Level 4	6059.33	1	13.32	.001*	.388
		Level 4 vs. Level 5	194.71	1	.389	.537	.018
	High	Level 1 vs. Level 2	14110.11	1	29.81	.000**	.554
		Level 2 vs. Level 3	1600.48	1	2.629	.118	.099
		Level 3 vs. Level 4	7602.45	1	9.037	.006*	.274
		Level 4 vs. Level 5	1283.36	1	.827	.372	.033
Attentional capacity	Low	Level 1 vs. Level 2	3823.57	1	3.353	.087	.182
		Level 2 vs. Level 3	3423.13	1	4.813	.044*	.244
		Level 3 vs. Level 4	4103.68	1	6.340	.024*	.297
		Level 4 vs. Level 5	44.46	1	.056	.817	.004
	High	Level 1 vs. Level 2	14316.42	1	14.72	.001**	.330
		Level 2 vs. Level 3	2090.02	1	2.890	.099	.088
		Level 3 vs. Level 4	9575.48	1	14.35	.001**	.324
		Level 4 vs. Level 5	1531.63	1	1.290	.265	.041
Working memory capacity	High	Level 1 vs. Level 2	9309.18	1	7.572	.013*	.284
		Level 2 vs. Level 3	4706.31	1	5.396	.031*	.222
		Level 3 vs. Level 4	52869.13	1	6.631	.018*	.259
		Level 4 vs. Level 5	138.50	1	.284	.599	.015
	Low	Level 1 vs. Level 2	8606.02	1	9.667	.004*	.272
		Level 2 vs. Level 3	1223.72	1	2.069	.162	.074
		Level 3 vs. Level 4	8388.88	1	14.961	.001*	.365
		Level 4 vs. Level 5	1364.48	1	.923	.345	.034

*α significant at *p*<.05
**α significant at *p*<.001

A multivariate ANOVA showed that neither proficiency, nor working memory or attentional capacity in and of themselves, exerted a significant influence on differences between the means of the fluency data at each of the levels of task complexity. Significance was only found when effects for proficiency were factored out, $F(4, 36) = 2.76$, $p = .043$, $\eta_p^2 = .234$.

Two-way repeated measures ANOVAs were run between the high and low level groups of each of the between-subjects independent variables in order to investigate how the groups may have differed in fluency measures as they performed the tasks. Tests of between-subjects effects showed that differences in the means between the high and low proficiency level groups was significant, $F(1,45) = 7.321$, $p = .01$, $\eta_p^2 = .140$, as well as between the high and low working memory capacity groups, $F(1,45) = 7.194$, $p = .01$, $\eta_p^2 = .138$, but not between the high and low attentional capacity groups, $F(1,45) = .875 p = .355$, $\eta_p^2 = .019$.

The results suggest that that differences between subjects in proficiency level, and level of working memory capacity, as well as differences in levels of attention and working memory capacity as a merged influence, may be factors which account for significantly different effects on fluency at lower levels of task complexity during performance on an oral task.

A series of Friedman's tests were performed to investigate whether there were any differences in the mid-AS-unit pause measurement of fluency between the task complexity levels. This was done separately for both high and low levels of proficiency, and working memory and attentional capacity in order to determine differences in the performances of the groups. In no case was any significant difference for mid AS-unit pause measurement between task complexity levels detected. This indicates that proficiency and working memory affected speed fluency, but not their pausing behavior as a result of variations in task complexity.

7.3.3 Accuracy measures: Non-Split data

Friedman's tests were performed on data for accuracy measures between all task complexity levels. For the data which was not split according between-subjects differences in proficiency, and attentional and working memory capacity, no significant differences were detected in any case (table 20).

Table 20 – Friedman's test for accuracy measures across all task complexity levels (non-split data)

Measure	N	χ^2	df	*p*-value
Morpho-syntactical errors	47	4.39	4	.356
Lexical errors	47	7.12	4	.130
Other errors	47	1.55	4	.818
Total errors	47	8.90	4	.062
Appropriacy repairs	47	5.14	4	.273
Different repairs	47	1.66	4	.798
Error repairs	47	1.54	4	.820
Repaired/Unrepaired errors	47	.215	4	.995
Repaired/Unrepaired errors (Corrected)	47	5.83	4	.212

*α significant at *p*< .05
**α significant at *p*< .001

7.3.4 Accuracy measures: Split data

Friedman's tests were performed to investigate whether there were any differences in the accuracy measurements between the task complexity levels. This was done separately for both high and low levels of proficiency, working memory capacity, and attentional capacity in order to determine differences in the performances of the three groups. Only in the case of total errors per AS-unit for the high level proficiency group was any significant difference found (table 21). Post-hoc analysis with Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at *p*<.005. The results did not indicate significance between any of the task complexity levels for errors per AS-unit for this proficiency level group. Table 22 shows data for consecutive task complexity levels.

The between measures tests support the observations made from the correlational analysis between values of task complexity and accuracy measures that variations in task complexity had little influence on dimensions of accuracy regardless of cognitive complexity capacity.

Table 21 – Friedman's test for accuracy measures across all task complexity levels (split data)

	Measure	Level	N	χ^2	df	*p*-value
Proficiency	Morpho-syntactical errors	Low	22	7.04	4	.134
		High	25	4.22	4	.377
	Lexical errors	Low	22	7.55	4	.110
		High	25	2.50	4	.644
	Other errors	Low	22	1.01	4	.908
		High	25	4.03	4	.402
	Total errors	Low	22	2.08	4	.721
		High	25	9.51	4	.049*
	Appropriacy repairs	Low	22	5.17	4	.270
		High	25	3.87	4	.423
	Different repairs	Low	22	6.97	4	.137
		High	25	5.27	4	.261
	Error repairs	Low	22	4.82	4	.307
		High	25	.603	4	.963
	Repaired/Unrepaired errors	Low	22	1.07	4	.898
		High	25	1.90	4	.754
	Repaired/Unrepaired errors (Corrected)	Low	22	.685	4	.953
		High	25	8.82	4	.066
Attentional Capacity	Morpho-syntactical errors	Low	16	2.19	4	.700
		High	31	4.27	4	.370
	Lexical errors	Low	16	3.13	4	.536
		High	31	12.82	4	.012
	Other errors	Low	16	3.52	4	.475
		High	31	.531	4	.970
	Total errors	Low	16	4.56	4	.336
		High	31	6.86	4	.143
	Appropriacy repairs	Low	16	2.26	4	.688
		High	31	3.55	4	.470
	Different repairs	Low	16	1.42	4	.841
		High	31	3.18	4	.528
	Error repairs	Low	16	6.64	4	.156
		High	31	.853	4	.931
	Repaired/Unrepaired errors	Low	16	2.26	4	.688
		High	31	1.75	4	.782
	Repaired/Unrepaired errors (Corrected)	Low	16	7.97	4	.093
		High	31	2.70	4	.609

*α significant at *p*<.05
**α significant at *p*<.001

Table 21 (continued) – Friedman's test for accuracy measures across all task complexity levels (split data)

	Measure	Level	N	χ^2	df	*p*-value
Working Memory Capacity	Morpho-syntactical errors	Low	20	2.89	4	.576
		High	27	1.89	4	.757
	Lexical errors	Low	20	2.34	4	.673
		High	27	6.29	4	.179
	Other errors	Low	20	2.68	4	.613
		High	27	2.96	4	.990
	Total errors	Low	20	2.27	4	.685
		High	27	7.18	4	.127
	Appropriacy repairs	Low	20	3.53	4	.473
		High	27	4.29	4	.368
	Different repairs	Low	20	2.02	4	.732
		High	27	4.85	4	.303
	Error repairs	Low	20	2.59	4	.629
		High	27	.473	4	.976
	Repaired/Unrepaired errors	Low	20	1.17	4	.883
		High	27	2.27	4	.686
	Repaired/Unrepaired errors (Corrected)	Low	20	4.74	4	.315
		High	27	4.60	4	.331

*α significant at *p*<.05
**α significant at *p*<.001

Table 22 - Wilcoxon signed-rank test for the high level proficiency group

		Task Complexity Levels			
		Levels1-2	Levels 2-3	Levels 3-4	Levels 4-5
Total errors	Z	-2.126	-1.238	-.400	-1.547
	Sig.	.034	.216	.689	.122
	r	-.45	-.26	-.09	-.33

**α significant at p<.005 (Bonferroni correction)*

7.3.5 Linguistic complexity measures: Non-Split data

Friedman's tests were performed on data for linguistic complexity measures between all task complexity levels to investigate whether there were any differences in the measurements of linguistic complexity between the task complexity levels. This was first done for the non-split data.

Table 23 shows the results of Friedman's test. Significance was detected in all cases except for the coordination index by total AS-units. For all measures where significance was found, a post-hoc analysis with Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at *p*<.005. Table 24 shows data only between consecutive levels both for brevity and clarity.

Table 23 – Friedman's test for linguistic complexity measures across all task complexity levels (non-split data)

Measure	N	χ^2	df	*p*-value
Lambda	47	99.38	4	.000**
D	47	36.90	4	.000**
Subordination X ASU	47	19.72	4	.001**
Coordination X ASU	47	5.39	4	.249
Words X Clause	47	61.88	4	.000**
Words X ASU	47	13.79	4	.008*
Idea Units	47	57.74	4	.000**

*α significant at *p*< .05
**α significant at *p*< .001

Table 24 - Wilcoxon signed-rank test for linguistic complexity measures

		Task Complexity Levels			
		Levels1-2	Levels 2-3	Levels 3-4	Levels 4-5
Lambda	Z	-1.597	-5.260	-4.836	-5.534
	Sig.	.110	.000*	.000*	.000*
	r	-.16	-.54	-.50	-.57
D	Z	-.751	-3.492	-2.836	-2.730
	Sig.	.452	.000*	.005*	.006
	r	-.08	-.36	-.29	-.28
Subordination X ASU	Z	-3.456	-1.262	-.972	-1.699
	Sig.	.001*	.207	.331	.089
	r	-.36	-.13	-.10	-.18
Words X Clause	Z	-5.265	-.497	-1.233	-1.683
	Sig.	.000*	.619	.218	.092
	r	-.54	-.05	-.13	-.17
Words X ASU	Z	-2.307	-.884	-.429	-.534
	Sig.	.021	.377	.668	.593
	r	-.24	-.09	-.04	-.06
Idea Units	Z	-2.975	-3.591	-3.784	-5.425
	Sig.	.003*	.000*	.000*	.000*
	r	-.31	-.37	-.39	-.56

**α significant at p<.005 (Bonferroni correction)*

7.3.5.1 Lambda: Lexical sophistication

Lambda showed significant differences in most cases between levels except for between levels one and two and levels two and four. In two of three cases of significance the higher task complexity level produced a higher value than the lower complexity level (table 43). These were between levels one and two, and levels three and four suggesting that as task complexity increased, subjects used more sophisticated vocabulary.

7.3.5.2 D: Lexical richness

Higher task complexity produced greater variety of vocabulary in most cases. The only exception was between levels one and two, and levels four and five although these differences were not significant (table 43).

7.3.5.3 Subordination per AS-unit

Significance was only found between levels one and two where level two produced more subordinate clauses than level one. The only other significant difference occurred where level five produced more clauses than level two (table 43). In general, increase task demands did not result in increases in the dimension of subordination.

7.3.5.4 Words per clause

Significance was also only found for level one which produced fewer words than all other levels (table 44).

7.3.5.5 Words per AS-unit

Significance was only found for level one which produced fewer words per unit than levels three, four, and five (table 44). According to this measure, increased task demands did not especially produce increases in structural complexity.

7.3.5.6 Idea Units (propositional complexity)

Greater task complexity produced more idea units in most cases except for level four which produced significantly fewer idea units than levels one, and three (table 43).

7.3.5.7 Summary: Linguistic complexity measures for non-split data

This section was a review of the linguistic complexity measures of the non-split data. The data indicates that propositional complexity and lexical complexity were most affected by increases in task complexity. There were cases in which a task of higher conceptual complexity actually produced a smaller measure than less complex tasks as operationalized in the study. This indicates that tasks considered less complex may, nonetheless, contain characteristics that require greater variety of vocabulary, more sophisticated vocabulary, or may require a greater number of concepts to be explained sufficiently well.

7.3.6 Linguistic complexity measures: Split data

Table 25 shows the results of a series of Friedman's tests which was performed to investigate whether there were any differences in the measurements of linguistic complexity between the task complexity levels. This was done separately for both high and low levels of proficiency, and working memory and attentional capacity in order to determine differences between the performances of these groups.

Table 25 – Friedman's test for linguistic complexity measures across all task complexity levels

	Measure	Level	N	χ^2	df	*p*-value
Proficiency	Lambda	Low	22	52.34	4	.000**
		High	25	48.77	4	.000**
	D	Low	22	12.80	4	.012*
		High	25	25.95	4	.000**
	Subordination X ASU	Low	22	9.48	4	.050*
		High	25	13.57	4	.009*
	Coordination X ASU	Low	22	3.29	4	.510
		High	25	5.20	4	.267
	Words X Clause	Low	22	32.22	4	.000**
		High	25	36.14	4	.000**
	Words X ASU	Low	22	6.84	4	.145
		High	25	9.28	4	.054
	Idea Units	Low	22	32.04	4	.000**
		High	25	56.46	4	.000**
Attentional Capacity	Lambda	Low	16	38.21	4	.000**
		High	31	61.95	4	.000**
	D	Low	16	12.45	4	.014*
		High	31	24.64	4	.000**
	Subordination X ASU	Low	16	9.62	4	.047*
		High	31	12.43	4	.014*
	Coordination X ASU	Low	16	6.57	4	.161
		High	31	6.42	4	.170
	Words X Clause	Low	16	21.75	4	000**
		High	31	43.28	4	.000**
	Words X ASU	Low	16	6.35	4	.174
		High	31	8.697	4	.069
	Idea Units	Low	16	12.73	4	.013*
		High	31	47.68	4	.000**

*α significant at *p*<.05
**α significant at *p*<.001

Table 25 (continued) – Friedman's test for linguistic complexity measures across all task complexity levels

	Measure	Level	N	χ^2	df	*p*-value
Working Memory Capacity	Lambda	Low	20	43.40	4	.000**
		High	27	59.16	4	.000**
	D	Low	20	22.92	4	.000**
		High	27	17.83	4	.001**
	Subordination X ASU	Low	20	17.70	4	.001**
		High	27	7.45	4	.114
	Coordination X ASU	Low	20	7.813	4	.099
		High	27	.542	4	.969
	Words X Clause	Low	20	27.80	4	.000**
		High	27	35.35	4	.000**
	Words X ASU	Low	20	8.76	4	.067
		High	27	7.62	4	.107
	Idea Units	Low	20	25.83	4	.000**
		High	27	34.28	4	.000**

*α significant at $p<.05$
**α significant at $p<.001$

Significance was determined for all measures except for measures of coordination per AS-unit, words per AS-unit, and in only one instance for subordination per AS-unit was significance not determined. This was the case for the high working memory capacity group. For instances in which significance was determined, a post-hoc analysis with Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at $p<$ .005. Table 26 shows data only between consecutive levels both for brevity and clarity. This was done to investigate at which points there was significant difference between measures at consecutive complexity levels along the continuum from simple to complex. Here, a pattern begins to emerge that indicates how individual differences seem to influence the way that some subjects approached the tasks.

Table 26 - Wilcoxon signed-rank test where significance was determined

				Task Complexity Levels			
				Levels 1-2	Levels 2-3	Levels 3-4	Levels 4-5
Lambda	Proficiency	Low	Z	-1.460	-3.799	-3.328	-3.911
			Sig.	.144	.000*	.001*	.000*
			r	-.22	-.57	-.50	-.59
		High	Z	-.800	-3.646	-3.445	-4.023
			Sig.	.423	.000*	.001*	.000*
			r	-.11	-.52	-.49	-.57
	Attentional capacity	Low	Z	-.931	-3.155	-3.361	-3.258
			Sig.	.352	.002*	.001*	.001*
			r	-.16	-.56	-.59	-.58
		High	Z	-1.276	-4.155	-3.469	-4.597
			Sig.	.202	.000*	.001*	.000*
			r	-.16	-.53	-.44	-.58
	Working memory capacity	Low	Z	-.483	-3.398	-3.846	-3.659
			Sig.	.629	.001*	.000*	.000*
			r	-.08	-.54	-.61	-.58
		High	Z	-1.715	-3.977	-3.015	-4.178
			Sig.	.086	.000*	.003*	.000*
			r	-.23	-.54	-.41	-.57
D	Proficiency	Low	Z	-.860	-2.127	-1.834	-.893
			Sig.	.390	.033	.067	.372
			r	-.13	-.32	-.28	-.13
		High	Z	-.309	-2.650	-2.381	-2.812
			Sig.	.757	.008*	.017	.005*
			r	-.04	-.37	-.34	-.40
	Attentional capacity	Low	Z	-.052	-1.913	-1.396	-1.603
			Sig.	.959	.056	.163	.109
			r	-.01	-.34	-.25	-.28
		High	Z	-.941	-2.939	-2.567	-2.234
			Sig.	.347	.003*	.010	.025
			r	-.12	-.37	-.33	-.28
	Working memory capacity	Low	Z	-1.904	-3.061	-1.643	-2.277
			Sig.	.057	.002*	.100	.023
			r	-.30	-.48	-.26	-.36
		High	Z	-.865	-1.634	-2.427	-1.490
			Sig.	.387	.102	.015	.136
			r	-.12	-.22	-.33	-.20

**α significant at $p<.005$ (Bonferroni correction)*

Table 26 (continued) - Wilcoxon signed-rank test where significance was determined

				Task Complexity Levels			
				Levels 1-2	Levels 2-3	Levels 3-4	Levels 4-5
Subordination X ASU	Proficiency	Low	Z	-2.062	-.097	-.765	-1.477
			Sig.	.039	.922	.444	.140
			r	-.31	-.01	-.12	-.22
		High	Z	-2.584	-1.529	-.686	-.815
			Sig.	.010	.126	.493	.415
			r	-.37	-.22	-.10	-.12
	Attentional capacity	Low	Z	-2.795	-1.500	-.621	-1.526
			Sig.	.005*	.134	.535	.127
			r	-.49	-.27	-.11	-.27
		High	Z	-2.234	-.309	-1.563	-.998
			Sig.	.025	.758	.118	.318
			r	-.28	-.04	-.20	-.13
	Working memory capacity	Low	Z	-2.913	-1.569	-.382	-2.204
			Sig.	.004*	.117	.702	.028
			r	-.46	-.25	-.06	-.35
		High	Z	-1.718	-.120	-1.490	-.279
			Sig.	.086	.904	.136	.780
			r	-.23	-.02	-.20	-.04
Words X clause	Proficiency	Low	Z	-3.782	-.179	-1.981	-.552
			Sig.	.000*	.858	.048	.581
			r	-.57	-.03	-.30	-.08
		High	Z	-3.861	-.486	-.431	-1.816
			Sig.	.000*	.627	.667	.069
			r	-.55	-.07	-.06	-.26
	Attentional capacity	Low	Z	-2.999	-.957	-.465	-.362
			Sig.	.003*	.339	.642	.717
			r	-.53	-.17	-.08	-.06
		High	Z	-4.370	-.267	-1.695	-1.793
			Sig.	.000*	.789	.090	.073
			r	-.55	-.03	-.22	-.23
	Working memory capacity	Low	Z	-3.808	-.112	-.653	-1.568
			Sig.	.000*	.911	.513	.117
			r	-.60	-.02	-.10	-.25
		High	Z	-3.808	-.571	-1.141	-.961
			Sig.	.000*	.568	.254	.337
			r	-.52	-.08	-.16	-.13

**α significant at p<.005 (Bonferroni correction)*

Table 26 (continued) - Wilcoxon signed-rank test where significance was determined

				Task Complexity Levels			
				Levels 1-2	Levels 2-3	Levels 3-4	Levels 4-5
Idea Units	Proficiency	Low	Z	-1.805	-2.908	-2.943	-3.953
			Sig.	.071	.004*	.003*	.000*
			r	-.27	-.44	-.44	-.60
		High	Z	-2.400	-2.423	-2.483	-3.745
			Sig.	.016	.015	.013	.000*
			r	-.34	-.34	-.35	-.53
	Attentional capacity	Low	Z	-1.894	-2.277	-2.236	-2.856
			Sig.	.058	.023	.025	.004*
			r	-.33	-.40	-.40	-.50
		High	Z	-2.391	-2.809	-3.024	-4.533
			Sig.	.017	.005*	.002*	.000*
			r	-.30	-.36	-.38	-.58
	Working memory capacity	Low	Z	-1.566	-1.291	-2.385	-3.928
			Sig.	.117	.197	.017	.000*
			r	-.25	-.20	-.38	-.62
		High	Z	-2.701	-3.668	-2.913	-3.895
			Sig.	.007	.000*	.004*	.000*
			r	-.37	-.50	-.40	-.53

**α significant at $p<.005$ (Bonferroni correction)*

7.3.6.1 Lambda: Lexical sophistication

Significance was shown between all consecutive levels of task complexity for both high and low levels of proficiency, and working memory and attentional capacity groups. The descriptive statistics (table 27) indicate that level three produced greater measures than level four, but indiscriminately for high or low levels of proficiency, and working memory and attentional capacity. This was also true for the measure of idea units described below. The nature of the story at level three seems to have required less frequent vocabulary and, as per the data corresponding to idea units, more ideas than the story for level four in spite of its supposedly lesser degree of cognitive complexity. This was also reflected in the non-split data.

Table 27 – Descriptive statistics for lambda per high and low levels of proficiency, and attentional and working memory capacity

			Proficiency					Attentional capacity					Working memory capacity				
Measure	level	Task Complexity	N	Median	Interquartile range	Min.	Max.	N	Median	Interquartile range	Min.	Max.	N	Median	Interquartile range	Min.	Max.
Lambda	Low	Level 1	22	.57	.26	.20	1.50	16	.55	.29	.20	1.60	20	.70	.47	.30	1.60
		Level 2	22	.75	.28	.25	1.11	16	.78	.36	.25	2.00	20	.82	.39	.29	2.00
		Level 3	22	1.26	.70	.67	2.11	16	1.43	.57	.92	3.97	20	1.32	.53	.83	3.97
		Level 4	22	.92	.44	.33	1.63	16	.84	.64	.38	2.00	20	.80	.73	.33	2.00
		Level 5	22	1.51	.35	.92	2.11	16	1.63	.77	.95	2.50	20	1.62	.85	.92	2.50
	High	Level 1	25	.73	.52	.17	1.60	31	.63	.50	.17	1.50	27	.58	.36	.17	1.50
		Level 2	25	.80	.39	.25	2.00	31	.75	.35	.25	1.43	27	.75	.26	.25	1.43
		Level 3	25	1.31	.45	.61	3.97	31	1.18	.55	.61	2.11	27	1.27	.55	.61	2.11
		Level 4	25	.80	.79	.41	2.00	31	.88	.67	.33	1.63	27	.96	.64	.40	1.63
		Level 5	25	1.63	.60	.95	2.50	31	1.48	.35	.92	2.50	27	1.53	.33	.92	2.50

7.3.6.2 D: Lexical richness

Results again showed no meaningful patterns in the data. This contrasts with the non-split data which showed that task complexity produced greater variety of vocabulary in most cases.

7.3.6.3 Subordination per AS-units

No pattern was detected for the measure of subordination per AS-units. This is similar to results for non-split data. There is no indication that increased task complexity affected the subordination regardless of proficiency level or cognitive capacity.

7.3.6.4 Words per clause

As with the measure of subordination per AS-units, no pattern was detected. This was also true of the non-split data.

7.3.6.5 Idea Units (propositional complexity)

Although significance was shown more often between successive complexity levels for low proficiency subjects (three instances) compared with high proficiency subjects (one instance), one of those demonstrated fewer idea units in the lower of the two complexity levels; this was the case of levels three and four. This also occurred for the non-split data. For the high proficiency group, the only successive levels of complexity that produced significant results between them were levels four and five where level five produced more idea units. Results indicate that the lower proficiency level group seemed a bit more susceptible to immediate variations in task complexity.

As with the high proficiency group, the low attentional capacity group showed significance between only levels four and five with level five producing more idea units than level four.

The high attentional capacity group demonstrated significance in seven out of the ten pairwise comparisons. In most cases, the higher task complexity level produced more idea units except for level four which produced fewer idea units than level three. Beginning between levels two and three, and except for levels three and four, pairwise comparison of ranked data demonstrated that each consecutive task complexity level was associated with more idea units produced. The effect of task complexity on idea unit production for high attentional capacity subjects is in stark contrast to the low attentional capacity subjects who showed only a single case of significance.

7.3.6.6 Summary: Linguistic complexity measures for split data

In general terms, the results for the split data tell a similar story to the non-split data. The primary effect of increased task complexity seems to have been a greater need for more sophisticated vocabulary at the expense of fluency.

Splitting the data into high and low levels of proficiency, and attentional and working memory capacity, however, put into evidence how results might be different when individual differences between subjects are taken into consideration. Results suggest that differences in task complexity affect the production of idea units, but that level of attentional capacity is determinant as to whether differences will be significant. This is supportive of my second hypothesis which states that differences in attentional capacity may be responsible for the clarity with which measures demonstrate trade-offs between other measures of CAF. In addition, for the working memory capacity groups, results point in a similar direction. High working memory capacity subjects showed more variation than low capacity subjects in terms of significance between task complexity levels for the measure of idea units. This suggests that differences in task complexity may have a greater effect on high working memory capacity subjects than on low capacity subjects. This is also supportive of my second hypothesis with respect to differences in working memory capacity.

7.4 Correlations between CAF dimensions

Spearman correlations were performed between all measures at all complexity levels in order to look for patterns which could indicate trends in how dimensions varied in similar ways as task complexity increased. Special emphasis was placed on searching for correlational trends between measures of CAF in order to identify possible points where trade-off effects occurred along the continuum of complexity. A trend was considered where two values correlated over more than one consecutive complexity level. It was considered that where a trend either began or ended was an indication that attentional resources were reallocated to a different aspect of production provoking a change in one or both of the correlating measures. Where a trend was detected, data for repeated measures and pairwise comparisons were consulted in order to search for information indicating that attentional resources may have been allocated in such a way that trade-off effects could have occurred. For the sake of clarity, cells that are shaded in the tables showing correlation coefficients are those cells whose data shows statistical significance.

7.4.1 Influence of lexis on syntax

The results of the correlations between values of task complexity and measures of CAF suggest that lexis may be driving syntax as task complexity increases. Measures of subordination per AS-unit were correlated with lambda values. This was done at each level of task complexity. First, correlations were run on non-split data and then on the split-data.

To investigate the relationship between lexis and syntax comparisons were made between lambda and the subordination index. According to Skehan (2009) native speakers are expected to manifest a positive correlation between lambda and structural complexity indexed by subordination while for non-native speakers this correlation is negative. A positive correlation would indicate greater structural complexity.

7.4.1.1 Correlation between lexis and syntax: Non-split data

The correlations run on non-split data showed no instances of significance in any case (table 28).

Table 28 - Spearman correlations between subordination index and lambda (non-split data)

	Task complexity level				
	1	2	3	4	5
Correlation Coefficient	-.116	-.136	.025	-.214	.012
Sig. (2-tailed)	.439	.363	.869	.148	.938
N	47	47	47	47	47

*α significant at *p*<.05
**α significant at *p*<.001

7.4.1.2 Correlation between lexis and syntax: Split data

Correlations were run on data divided into high and low levels of proficiency, and attentional and working memory capacity. Only in one instance did a correlation show significance for the high level proficiency group at task complexity level four (table 29).

7.4.1.3 Summary of the correlation between lexis and syntax

Although correlations were mostly negative as expected according to Skehan’s (2009) findings for non-native speakers, results suggest that while lambda correlated positively with increased task complexity, lexis did not play a significant role in driving syntax.

Table 29 - Spearman correlations between subordination index and lambda (split data)

		Task complexity level														
		Proficiency					Attentional capacity					Working memory capacity				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Low	Correlation Coefficient	-.266	.043	.111	-.059	.326	-.330	-.326	.145	-.466	-.171	-.050	-.041	-.035	-.114	.116
	Sig. (2-tailed)	.232	.850	.623	.796	139	.212	.217	.592	.069	.526	.789	.827	.852	.542	.533
	N	22	22	22	22	22	16	16	16	16	16	31	31	31	31	31
High	Correlation Coefficient	.003	-.188	-.070	-.150(**)	-.231	-.050	-.041	-.035	-.114	.116	.112	.165	-.055	-.152	.334
	Sig. (2-tailed)	.988	.368	.739	.024	.267	.789	.827	.852	.542	.533	.579	.412	.786	.448	.089
	N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27

*α significant at *p*<.05
**α significant at *p*<.001

7.4.2 Rate A and errors

It had been shown earlier in this section that rate A measure of fluency showed that as task complexity increased, subjects tended to speak more slowly. However, correlations between error measures and rate A are also negative across the board (table 30). This is reflected for non-split data as well as for data which was split. The indication is that greater fluency was associated with fewer errors and greater accuracy at all levels of task complexity independently of individual differences. This is contrary to predictions by both the Cognition Hypothesis and the Trade-off Hypothesis.

Table 30 - Spearman correlations between rate A and error measures (non-split data)

		Task complexity level				
		1	2	3	4	5
Morpho-syntactical errors	Correlation Coefficient	-.366(*)	-.475(**)	-.324(*)	-.155	-.532(**)
	Sig. (2-tailed)	.011	.001	.026	.297	.000
	N	47	47	47	47	47
Lexical errors	Correlation Coefficient	-.180	-.413(**)	-.137	-.159	-.456(**)
	Sig. (2-tailed)	.226	.004	.359	.285	.001
	N	47	47	47	47	47
Other errors	Correlation Coefficient	.105	-.140	-.299(*)	.183	-.232
	Sig. (2-tailed)	.481	.348	.041	.219	.116
	N	47	47	47	47	47
Total errors	Correlation Coefficient	-.331(*)	-.537(**)	-.278	-.129	-.487(**)
	Sig. (2-tailed)	.023	.000	.059	.389	.001
	N	47	47	47	47	47

*α significant at *p*<.05
**α significant at *p*<.001

7.4.3 Pausing and errors

In the non-split data, pauses accompany the measure of total errors for the first two levels of task complexity. In addition, pauses correlate with both measures of error repairs and the corrected measure of repaired to unrepaired errors for the first two levels (table 31). However, when looking at the split data (table 32), it becomes clear that individual differences between subjects influence how the task is performed. Data for total errors shows several cases of significance but no trend or pattern is evident. For the measure for error repairs, only the low attentional capacity group and the high working memory group maintained the same pattern as shown in the non-split data. For the measure of the corrected measure of repaired to unrepaired errors, the high working memory group, and the low proficiency and low attentional capacity groups all show the same positive correlations as the non-split data showed for the first two task complexity levels where their counterparts did not. The bottom line is that high and low proficiency, and attentional and working memory capacity groups approach accuracy differently in order to meet task demands.

Table 31- Spearman correlations between error and repair measures, and mid-ASU pausing (non-split data)

		Task complexity level				
		1	2	3	4	5
Total errors	Correlation Coefficient	,430(**)	,368(*)	0,267	0,277	,329(*)
	Sig. (2-tailed)	0,003	0,011	0,070	0,059	0,024
	N	47	47	47	47	47
Error repairs	Correlation Coefficient	,499(**)	,426(**)	0,187	0,186	,340(*)
	Sig. (2-tailed)	0,000	0,003	0,209	0,212	0,019
	N	47	47	47	47	47
Corrected repaired to unrepaired errors	Correlation Coefficient	,406(**)	,320(*)	0,103	0,075	0,132
	Sig. (2-tailed)	0,005	0,028	0,489	0,615	0,377
	N	47	47	47	47	47

*α significant at *p*<.05
**α significant at *p*<.001

Table 32 - Spearman correlations between error and repair measures, and mid-ASU pausing (split data)

			Task complexity level														
			Proficiency					Attentional capacity					Working memory capacity				
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Total errors	Low	Correlation Coefficient	.337	.429(*)	.251	.486(*)	.255	.651(**)	.369	.326	-.044	.341	.408	.283	-.215	.348	-.112
		Sig. (2-tailed)	.125	.047	.259	.022	.252	.006	.159	.218	.871	.196	.074	.226	.362	.133	.638
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.610(**)	.273	.270	.216	.221	.297	.343	.111	.439(*)	.313	.421(*)	.334	.478(*)	.286	.425(*)
		Sig. (2-tailed)	.001	.186	.192	.301	.288	.105	.059	.551	.013	.086	.029	.089	.012	.148	.027
		N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27
Error repairs	Low	Correlation Coefficient	.685(**)	.417	.207	.332	.441(*)	.679(**)	.733(**)	.298	.452	.430	.342	.005	-.290	-.101	.035
		Sig. (2-tailed)	.000	.053	.355	.131	.040	.004	.001	.262	.079	.097	.140	.982	.215	.673	.883
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.307	.426(*)	.200	.059	.263	.377(*)	.283	.006	.091	.376(*)	.532(**)	.612(**)	.362	.369	.364
		Sig. (2-tailed)	.135	.034	.338	.781	.203	.037	.124	.974	.625	.037	.004	.001	.063	.058	.062
		N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27
Corrected repaired to unrepaired errors	Low	Correlation Coefficient	.614(**)	.468(*)	.240	.261	.330	.573(*)	.616(*)	.318	.432	.310	.319	.132	.119	-.099	.107
		Sig. (2-tailed)	.002	.028	.283	.241	.134	.020	.011	.231	.095	.243	.170	.580	.618	.677	.652
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.325	.169	.008	-.114	.113	.340	.249	-.105	-.063	.065	.423(*)	.464(*)	.079	.204	.228
		Sig. (2-tailed)	.113	.420	.968	.588	.590	.061	.177	.575	.737	.728	.028	.015	.694	.308	.252
		N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27

*α significant at $p<.05$

**α significant at $p<.001$

Tavakoli (2011) remarks that pausing is an indication of a lexical search, or expresses a concern about structure, pronunciation, or to plan for speech. In the data we are observing the same metalinguistic phenomenon as subjects seem to pause as a demonstration of concern for accuracy. However, as mentioned above, there is an indication that different groups are approaching this in different ways.

In order to further investigate data concerning errors, a series of Mann-Whitney tests was performed to investigate performance at the different levels of task complexity by each of the two groups representing high and low levels of proficiency, and attentional and working memory capacity. The results showed that the low proficiency subjects committed a greater number of errors at all task complexity levels for every error measure. However, results for the working memory and attentional capacity groups pointed in the opposite direction. For both working memory and attentional capacity groups, along all task complexity levels, the higher level group committed more errors in most cases. Table 33 on the following pages shows descriptive statistics for morpho-syntactical errors, lexical errors, and total errors for both high and low levels of proficiency, and attentional and working memory capacity. Table 34 shows results of the Mann-Whitney tests.

Table 33 – Descriptive statistics for error measures per high and low levels of proficiency, and attentional and working memory capacity

			Proficiency					Attentional capacity					Working memory capacity				
Measure	level	Task Complexity	N	Median	Interquartile range	Min.	Max.	N	Median	Interquartile range	Min.	Max.	N	Median	Interquartile range	Min.	Max.
Morpho-syntactical errors	Low	Level 1	22	.440	.51	.00	1.13	16	.125	.29	.00	.56	20	.15	.29	.00	.75
		Level 2	22	.420	.40	.00	1.71	16	.190	.34	.00	.75	20	.22	.29	.00	1.00
		Level 3	22	.415	.36	.00	.82	16	.195	.37	.00	.73	20	.22	.31	.00	.60
		Level 4	22	.330	.49	.00	1.40	16	.245	.31	.00	.50	20	.225	.39	.00	.60
		Level 5	22	.445	.51	.10	1.00	16	.145	.19	.00	.45	20	.22	.34	.00	.83
	High	Level 1	25	.140	.26	.00	.71	31	.360	.54	.00	1.13	27	.33	.57	.00	1.13
		Level 2	25	.200	.20	.00	1.33	31	.350	.50	.00	1.71	27	.38	.57	.00	1.71
		Level 3	25	.220	.28	.00	1.17	31	.290	.38	.00	1.17	27	.36	.38	.00	1.17
		Level 4	25	.170	.37	.00	.78	31	.250	.54	.00	1.40	27	.29	.52	.00	1.40
		Level 5	25	.210	.34	.00	.94	31	.420	.42	.00	1.00	27	.39	.35	.00	1.00
Lexical errors	Low	Level 1	22	.530	.41	.10	1.00	16	.345	.35	.00	1.00	20	.39	.39	.00	1.00
		Level 2	22	.815	.91	.22	2.33	16	.440	.52	.00	1.43	20	.42	.47	.00	1.43
		Level 3	22	.695	.44	.13	1.30	16	.220	.40	.00	1.00	20	.31	.10	.00	1.25
		Level 4	22	.780	.74	.00	1.86	16	.210	.35	.00	1.13	20	.33	.42	.11	1.13
		Level 5	22	.630	.65	.13	1.63	16	.570	.60	.00	1.50	20	.30	.51	.00	1.45
	High	Level 1	25	.330	.25	.00	.80	31	.430	.31	.00	1.00	27	.44	.27	.00	1.00
		Level 2	25	.400	.39	.00	1.14	31	.500	.67	.11	2.33	27	.50	.81	.17	2.33
		Level 3	25	.350	.29	.00	1.00	31	.600	.58	.12	1.30	27	.60	.56	.00	1.30
		Level 4	25	.390	.35	.11	1.13	31	.580	.68	.11	1.86	27	.50	.47	.00	1.86
		Level 5	25	.350	.44	.00	.80	31	.440	.38	.00	1.63	27	.50	.36	.09	1.63

Table 34 – Mann-Whitney U for error measures per high and low levels of proficiency, and attentional and working memory capacity

		Morpho-syntactical errors			Lexical errors			Total errors		
Task complexity level		Proficiency	Attention	Working memory	Proficiency	Attention	Working memory	Proficiency	Attention	Working memory
1	Mann-Whitney U	141	128	193.5	136	213.5	244.5	71.50	147.5	191
	Z	-2.86	-2.70	-1.65	-2.97	-.78	-.55	-4.34	-2.26	-1.70
	Sig.	.000**	.010*	.10	.000**	.44	.58	.000**	.02*	.09
	r	.42	.39	.24	.43	.11	.08	.63	.33	.25
2	Mann-Whitney U	142	155.5	199.50	136	199.5	18.00	115	173	165
	Z	-2.84	-2.01	-1.52	-2.97	-1.09	-1.94	-3.41	-1.69	-2.26
	Sig.	.000**	.04*	.13	.000**	.28	.05*	.000**	.09	.02*
	r	.41	.29	.22	.43	.16	.28	.50	.25	.33
3	Mann-Whitney U	171	155.5	167.5	101.5	121	166.5	97.5	142.5	212.5
	Z	-2.23	-2.09	-2.22	-3.70	-2.85	-2.23	-3.79	-2.37	-1.47
	Sig.	.03*	.04*	.03*	.000**	.000**	.03*	.000**	.02*	.14
	r	.33	.30	.32	.54	.42	.32	.55	.35	.21
4	Mann-Whitney U	16.5	191	193	135.5	106.5	188	136.5	114	178.5
	Z	-2.45	-1.29	-1.66	-2.98	-3.18	-1.77	-2.95	-3.01	-1.97
	Sig.	.01*	.20	.10	.000**	.000**	.08	.000**	.000**	.05*
	r	.36	.19	.24	.43	.46	.26	.43	.44	.29
5	Mann-Whitney U	126	125.5	191	136	211.5	18.50	94	167.5	174.5
	Z	-3.18	-2.75	-1.70	-2.96	-.82	-1.93	-3.86	-1.81	-2.06
	Sig.	.000**	.01*	.09	.00**	.41	.05*	.000**	.07	.04*
	r	.46	.40	.25	.43	.12	.28	.56	.26	.30

r = effect size
*α significant at *p*<.05
**α significant at *p*<.001

7.4.4 Pausing and error repair

While pausing accompanied by errors may be an indication of awareness of interlanguage accuracy, those accompanied by error repairs suggest conscientiousness toward performing accurate interlanguage. Pauses accompany errors across most task complexity levels, but at higher levels, the correlations between pauses and the measure for error repair as well as the corrected value for repaired to unrepaired errors disappear. This is an indication that increased task complexity led the subjects away not from awareness of erroneous speech, but away from concern about accuracy of speech.

To further compare data with the findings described in Kormos (2006) which showed that lower proficiency learners correct a smaller proportion of mistakes than learners of higher proficiency, Mann-Whitney tests were performed on both measures of repaired to unrepaired errors per AS-unit, and the corrected measure for repaired to unrepaired errors. This was to investigate possible differences in the performance of subjects of high and low proficiency, and attentional and working memory capacity. The results show that all subjects corrected at the same proportion regardless of level of proficiency, and attentional and working memory capacity with only two exceptions where the high proficiency group corrected a greater proportion of errors; at task complexity level five for the measure of repaired to unrepaired errors, $U = -183$, $p = .05$, $r = -.29$, and the corrected measure of repaired to unrepaired errors, $U = 179$, $p = .04$, $r = -.3$. These results reflect findings in Gilabert (2007) who also found no significant difference on measures of self-repair between subjects of proficiency levels similar to those who participated in the current task.

7.4.5 Error and error repairs

For the non-split data, correlations between error measures and the corrected measure of repaired to unrepaired errors indicate that at the highest level of task complexity, subjects showed significant negative correlations with most error measures except for morpho-syntactical measures. A greater number of errors of all kinds resulted in a smaller percentage of them being repaired. But when correlations for the split data are consulted, it shows that only the high proficiency group and the low working memory group showed significant negative correlations. This was true for all error measures (table 35). None of the other groups showed such a pattern. It seems that these two groups followed a similar strategy to deal with task demands at the highest level of complexity.

Table 35 - Spearman correlations between error measures and the corrected measure of repaired to unrepaired errors

			Task complexity level														
			Proficiency					Attentional capacity					Working memory capacity				
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Morpho-syntactical errors	Low	Correlation Coefficient	.359	.267	.061	.101	.138	.429	.171	-.146	.025	-.640(**)	.359	-.057	-.311	.198	-.596(**)
		Sig. (2-tailed)	.101	.230	.788	.655	.540	.097	.526	.589	.926	.008	.120	.810	.183	.402	.006
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.232	.111	-.126	.080	-.425(*)	.025	.193	-.126	-.012	-.038	.000	.387(*)	-.003	-.154	.132
		Sig. (2-tailed)	.265	.598	.550	.705	.034	.893	.298	.499	.949	.838	.999	.046	.989	.444	.511
		N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27
Lexical errors	Low	Correlation Coefficient	-.041	.145	.026	-.199	-.053	.163	.140	-.174	-.205	-.176	-.088	-.026	-.240	.093	-.675(**)
		Sig. (2-tailed)	.858	.519	.909	.376	.814	.546	.604	.520	.446	.513	.712	.912	.309	.697	.001
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.229	-.065	-.189	.020	-.451(*)	-.102	.072	-.167	-.103	-.423(*)	.066	.215	-.043	-.369	-.008
		Sig. (2-tailed)	.272	.757	.365	.923	.024	.586	.698	.370	.580	.018	.745	.280	.829	.058	.969
		N	25	25	25	25	25	31	31	31	31		27	27	27	27	27
Other errors	Low	Correlation Coefficient	.105	-.051	-.168	.216	.003	-.009	.040	.019	-.009	-.493	.196	.019	-.238	.046	-.674(**)
		Sig. (2-tailed)	.641	.822	.456	.335	.991	.973	.883	.945	.973	.212	.409	.936	.312	.847	.001
		N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
	High	Correlation Coefficient	.109	-.005	-.005	-.138	-.436(*)	-.047	-.055	-.220	.077	-.199	-.094	.003	.051	.039	-.017
		Sig. (2-tailed)	.604	.983	.981	.512	.029	.802	.769	.235	.679	.283	.642	.988	.801	.845	.933
		N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27

*α significant at *p*<.05
**α significant at *p*<.001

7.4.6 Lexical richness and pausing

High and low working memory groups demonstrated opposite effects on their correlations between D and pausing. The low working memory group showed a significant positive correlation for the first three levels while the high level working memory group produced significant negative correlations for the first two levels of task complexity (table 36). Where the low working memory group paused more while producing a greater variety of lexis, the high working memory capacity group paused less as they used a richer vocabulary.

Table 36 - Spearman correlations between D and mid-ASU pauses

Working memory capacity		Task complexity level 1	2	3	4	5
Low	Correlation Coefficient	.612(**)	.518(*)	.446(*)	-.116	.276
	Sig. (2-tailed)	.004	.019	.049	.626	.239
	N	20	20	20	20	20
High	Correlation Coefficient	-.478(*)	-.412(*)	-.380	-.351	-.199
	Sig. (2-tailed)	.012	.033	.051	.072	.320
	N	27	27	27	27	27

*α significant at *p*<.05
**α significant at *p*<.001

7.4.7 Lexical sophistication and propositional complexity

For the non-split data, there was a strong negative correlation between lambda and the measure for idea units at the highest task complexity level (table 37). The indication is that more idea units were formed at the expense of lexical sophistication; subjects used simpler vocabulary. However, a review of correlations performed on split data (table 38) indicated that while this was true for both high and low proficiency groups, only the low working memory and attentional capacity groups sacrificed more complex vocabulary while their high level counterparts were not significantly affected in this way by task complexity.

Table 37 - Spearman correlations between lambda and idea units (non-split data)

	Task complexity level				
	1	2	3	4	5
Correlation Coefficient	.128	.176	-.144	.198	-.446(**)
Sig. (2-tailed)	.391	.237	.333	.181	.002
N	47	47	47	47	47

*α significant at *p*<.05
**α significant at *p*<.001

Table 38 - Spearman correlations between lambda and idea units (split data)

Levels of proficiency, attention, working memory		Proficiency					Attentional capacity					Working memory capacity				
		Task complexity					Task complexity					Task complexity				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Low	Correlation Coefficient	-.407	.255	.014	.117	-.468(*)	.165	.214	-.163	.237	-.572(*)	.088	.200	.048	.214	-.654(**)
	Sig. (2-tailed)	.060	.252	.952	.604	.028	.542	.425	.546	.377	.020	.711	.398	.842	.364	.002
	N	22	22	22	22	22	16	16	16	16	16	20	20	20	20	20
High	Correlation Coefficient	.371	.039	-.293	.257	-.509(**)	.093	.129	-.130	.178	-.271	.227	.177	-.344	.170	-.226
	Sig. (2-tailed)	.068	.853	.156	.215	.009	.620	.491	.486	.338	.140	.255	.376	.079	.395	.257
	N	25	25	25	25	25	31	31	31	31	31	27	27	27	27	27

*α significant at *p*<.05
**α significant at *p*<.001

7.5 Descriptive Statistics

Table 39– Descriptive statistics for accuracy measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Morpho-syntactical errors	Level 1	47	.33	.27	.41	.31	.00	1.13
	Level 2	47	.40	.29	.47	.39	.00	1.71
	Level 3	47	.31	.27	.42	.27	.00	1.17
	Level 4	47	.33	.25	.40	.31	.00	1.40
	Level 5	47	.35	.29	.36	.28	.00	1.00
Lexical errors	Level 1	47	.44	.38	.30	.26	.00	1.00
	Level 2	47	.63	.50	.59	.47	.00	2.33
	Level 3	47	.52	.45	.55	.35	.00	1.30
	Level 4	47	.59	.50	.60	.43	.00	1.86
	Level 5	47	.52	.47	.43	.39	.00	1.63
Other errors	Level 1	47	.18	.13	.29	.18	.00	.70
	Level 2	47	.16	.14	.20	.14	.00	.57
	Level 3	47	.20	.17	.28	.19	.00	.70
	Level 4	47	.20	.15	.29	.24	.00	1.17
	Level 5	47	.18	.15	.18	.16	.00	.70
Total errors	Level 1	47	.94	.73	.76	.53	.13	2.25
	Level 2	47	1.19	.94	1.19	.78	.14	3.29
	Level 3	47	1.04	.92	.93	.60	.00	2.50
	Level 4	47	1.13	.92	.87	.76	.14	3.43
	Level 5	47	1.06	1.00	.93	.63	.00	2.38

*α significant at *p*<.05
**α significant at *p*<.001

Table 40 – Descriptive statistics for accuracy measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Appropriacy errors	Level 1	47	.03	.00	.00	.05	.00	.20
	Level 2	47	.03	.00	.00	.06	.00	.20
	Level 3	47	.04	.00	.08	.06	.00	.20
	Level 4	47	.03	.00	.00	.06	.00	.22
	Level 5	47	.05	.00	.10	.07	.00	.30
Different repairs	Level 1	47	.06	.00	.13	.08	.00	.27
	Level 2	47	.05	.00	.07	.08	.00	.29
	Level 3	47	.05	.00	.09	.07	.00	.22
	Level 4	47	.06	.00	.11	.08	.00	.33
	Level 5	47	.06	.00	.10	.10	.00	.43
Error repairs	Level 1	47	.10	.08	.17	.10	.00	.30
	Level 2	47	.14	.11	.19	.22	.00	1.29
	Level 3	47	.10	.09	.17	.10	.00	.35
	Level 4	47	.13	.11	.17	.16	.00	.67
	Level 5	47	.12	.10	.09	.11	.00	.60
Repaired/Unrepaired errors	Level 1	47	.39	.25	.67	.46	.00	2.00
	Level 2	47	.33	.17	.47	.48	.00	2.40
	Level 3	47	.28	.20	.35	.53	-2.00	2.33
	Level 4	47	.40	.29	.46	.48	.00	2.00
	Level 5	47	.41	.19	.42	1.14	-4.00	5.00

*α significant at $p<.05$
**α significant at $p<.001$

Table 41 – Descriptive statistics for accuracy measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Repaired/Unrepaired errors (corrected)	Level 1	47	.48	.50	.82	.39	.00	1.42
	Level 2	47	.42	.29	.59	.41	.00	2.06
	Level 3	47	.49	.41	.51	.38	.00	1.57
	Level 4	47	.48	.45	.64	.38	.00	1.41
	Level 5	47	.67	.49	.68	.64	.00	3.67

*α significant at $p<.05$
**α significant at $p<.001$

Table 42 – Descriptive statistics for fluency measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Rate A	Level 1	47	164.69	168.89	60.92	41.91	77.54	256.22
	Level 2	47	145.25	138.95	54.55	39.70	79.50	240.00
	Level 3	47	134.86	135.00	51.43	33.30	68.28	191.54
	Level 4	47	117.81	116.33	43.64	30.03	67.58	192.75
	Level 5	47	123.02	123.91	44.90	33.09	59.24	208.00
Pauses Mid-ASU	Level 1	47	.40	.33	.43	.38	.00	1.75
	Level 2	47	.53	.33	.49	.58	.00	2.50
	Level 3	47	.56	.38	.71	.64	.00	3.00
	Level 4	47	.60	.40	.72	.59	.00	2.40
	Level 5	47	.54	.36	.48	.52	.00	2.11

*α significant at $p<.05$
**α significant at $p<.001$

Table 43 – Descriptive statistics for linguistic complexity measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Lambda	Level 1	47	.69	.60	.38	.33	.17	1.60
	Level 2	47	.78	.75	.34	.31	.25	2.00
	Level 3	47	1.34	1.31	.50	.53	.61	3.97
	Level 4	47	.93	.88	.67	.39	.33	2.00
	Level 5	47	1.57	1.54	.41	.39	.92	2.50
D	Level 1	47	33.24	32.97	11.91	9.27	15.47	57.80
	Level 2	47	32.15	32.09	13.58	1.07	14.81	64.17
	Level 3	47	37.81	35.58	14.10	1.44	18.70	58.69
	Level 4	47	44.02	42.88	12.78	12.36	15.22	91.58
	Level 5	47	38.52	35.77	11.75	1.30	2.26	64.01
Subordination X ASU	Level 1	47	.99	1.00	.51	.44	.20	2.29
	Level 2	47	.75	.67	.45	.50	.11	2.83
	Level 3	47	.84	.77	.50	.48	.00	2.50
	Level 4	47	.89	.83	.53	.41	.20	2.00
	Level 5	47	1.02	.88	.61	.50	.15	2.30
Coordination X ASU	Level 1	47	.20	.14	.17	.27	.00	1.69
	Level 2	47	.15	.03	1.01	.00	.71	.71
	Level 3	47	.18	.14	.14	.18	.00	.78
	Level 4	47	.17	.14	.17	.14	.00	.50
	Level 5	47	.24	.20	.23	.20	.00	1.00

*α significant at *p*<.05
**α significant at *p*<.001

Table 44 – Descriptive statistics for linguistic complexity measures per level of task complexity

Measure	Complexity	N	Mean	Median	Interquartile range	Std. deviation	Min.	Max.
Words X Clause	Level 1	47	5.24	5.15	0.81	.749	3.77	7.62
	Level 2	47	6.54	6.17	1.80	1.23	3.27	9.52
	Level 3	47	6.49	6.16	1.07	1.36	4.43	12.48
	Level 4	47	6.17	6.10	1.29	.809	4.50	8.59
	Level 5	47	5.99	5.81	1.03	.971	4.38	9.19
Words X ASU	Level 1	47	11.41	11.44	3.34	2.42	7.20	18.76
	Level 2	47	12.61	11.67	3.20	3.25	7.20	22.86
	Level 3	47	12.99	12.20	4.64	3.44	7.45	23.50
	Level 4	47	12.74	12.60	4.17	2.82	7.67	21.83
	Level 5	47	13.21	12.78	3.77	3.28	7.38	22.80
Idea Units	Level 1	47	16.02	14	7	8.17	7	47
	Level 2	47	13.72	12	7	7.14	5	40
	Level 3	47	16.17	14	6	7.44	6	38
	Level 4	47	13.32	11	6	7.09	6	40
	Level 5	47	19.06	17	10	9.30	8	57

*α significant at $p<.05$
**α significant at $p<.001$

Table 45 - Shapiro-Wilk test: Accuracy measures

Measure	Complexity	N	Shapiro-Wilk Statistic	df	Sig.
Morpho-syntactical errors	Level 1	47	.891	47	.000**
	Level 2	47	.857	47	.000**
	Level 3	47	.916	47	.002*
	Level 4	47	.882	47	.000**
	Level 5	47	.926	47	.006*
Lexical errors	Level 1	47	.963	47	.147
	Level 2	47	.877	47	.000**
	Level 3	47	.951	47	.049*
	Level 4	47	.895	47	.000**
	Level 5	47	.907	47	.001**
Other errors	Level 1	47	.871	47	.000**
	Level 2	47	.886	47	.000**
	Level 3	47	.902	47	.001**
	Level 4	47	.802	47	.000**
	Level 5	47	.882	47	.000**
Total errors	Level 1	47	.933	47	.010*
	Level 2	47	.863	47	.000**
	Level 3	47	.968	47	.229
	Level 4	47	.852	47	.000**
	Level 5	47	.968	47	.217
Appropriacy errors	Level 1	47	.576	47	.000**
	Level 2	47	.531	47	.000**
	Level 3	47	.681	47	.000**
	Level 4	47	.537	47	.000**
	Level 5	47	.742	47	.000**
Different repairs	Level 1	47	.745	47	.000**
	Level 2	47	.602	47	.000**
	Level 3	47	.724	47	.000**
	Level 4	47	.732	47	.000**
	Level 5	47	.674	47	.000**
Error repairs	Level 1	47	.873	47	.000**
	Level 2	47	.593	47	.000**
	Level 3	47	.860	47	.000**
	Level 4	47	.792	47	.000**
	Level 5	47	.823	47	.000**

*α significant at *p*<.05
**α significant at *p*<.001

Table 46 - Shapiro Wilk test: Accuracy measures

			Shapiro-Wilk		
Measure	Complexity	N	Statistic	df	Sig.
Repaired/Unrepaired errors	Level 1	47	.785	47	.000**
	Level 2	47	.666	47	.000**
	Level 3	47	.716	47	.000**
	Level 4	47	.769	47	.000**
	Level 5	47	.689	47	.000**
Repaired/Unrepaired errors (corrected)	Level 1	47	.918	47	.003*
	Level 2	47	.852	47	.000**
	Level 3	47	.923	47	.004*
	Level 4	47	.939	47	.017*
	Level 5	47	.781	47	.000**

*α significant at *p*<.05
**α significant at *p*<.001

Table 47- Shapiro-Wilk test: Fluency measures

			Shapiro-Wilk		
Measure	Complexity	N	Statistic	df	Sig.
Rate A	Level 1	47	.978	47	.511
	Level 2	47	.970	47	.262
	Level 3	47	.970	47	.277
	Level 4	47	.971	47	.284
	Level 5	47	968	47	.216
Mid-ASU pause	Level 1	47	.840	47	.000**
	Level 2	47	.799	47	.000**
	Level 3	47	.771	47	.000**
	Level 4	47	.855	47	.000**
	Level 5	47	.813	47	.000**

*α significant at *p*<.05
**α significant at *p*<.001

Table 48 – Shapiro-Wilk test: Linguistic complexity measures

Measure	Complexity	N	Shapiro-Wilk Statistic	df	Sig.
Lambda	Level 1	47	.941	47	.020*
	Level 2	47	.907	47	.001**
	Level 3	47	.783	47	.000**
	Level 4	47	.961	47	.116
	Level 5	47	.949	47	.041*
D	Level 1	47	.973	47	.340
	Level 2	47	.968	47	.214
	Level 3	47	.962	47	.132
	Level 4	47	.934	47	.011*
	Level 5	47	.943	47	.023*
Subordination X ASU	Level 1	47	.967	47	.199
	Level 2	47	.831	47	.000**
	Level 3	47	.932	47	.009*
	Level 4	47	.962	47	.131
	Level 5	47	.939	47	.016*
Coordination X ASU	Level 1	47	.614	47	.000**
	Level 2	47	.898	47	.001**
	Level 3	47	.846	47	.000**
	Level 4	47	.926	47	.005*
	Level 5	47	.888	47	.000**
Words X Clause	Level 1	47	.877	47	.000**
	Level 2	47	.967	47	.204
	Level 3	47	.787	47	.000**
	Level 4	47	.973	47	.347
	Level 5	47	.908	47	.001**
Words X ASU	Level 1	47	.972	47	.326
	Level 2	47	.867	47	.000**
	Level 3	47	.956	47	.072
	Level 4	47	.959	47	.096
	Level 5	47	.955	47	.070
Idea Units	Level 1	47	.765	47	.000**
	Level 2	47	.848	47	.000**
	Level 3	47	.851	47	.000**
	Level 4	47	.782	47	.000**
	Level 5	47	.830	47	.000**

*α significant at $p<.05$
**α significant at $p<.001$

CHAPTER 8: DISCUSSION

8.1 Introduction

To reestablish the context of the current investigation, it shall be recalled that the introduction of TBLT has brought on a need for establishing criteria by which pedagogic tasks may be sequenced from simple to complex. In this way, task-based syllabi may recreate the ontological learning process which is expected to aid in the efficiency and efficacy of the language learning process. In order to create such criteria, researchers in the field of SLA have worked to understand how the manipulation of task characteristics affects speech production. Emphasis has been placed on attempting to predict dimensions of language production based on changing the cognitive demands placed on learners as they perform tasks. Skehan’s Trade-off hypothesis (Skehan, 1998; Skehan & Foster, 2001) and Robinson’s Cognition hypothesis (Robinson 2001a, 2001b, 2003a, 2005) are the two principle theories which serve as the basis by which research is commonly compared.

The Trade-off hypothesis states that once a threshold is reached, limited attentional resources enter into competition for allocation to aspects of language production in order for a task to be successfully completed. Improved performance in one area will result in decreased performance in other areas. Fluency will tend to decrease while resources are allocated to meaning or form. Meaning is prioritized, but within form, there is competition between accuracy and linguistic complexity. Generally, it would be expected that attentional resources may be directed to one or the other, but not to both simultaneously.

The Cognition Hypothesis also states that once a threshold is reached, limited attentional resources must be allocated in such a way to ensure successful completion of a task. Fluency generally decreases, but attentional resources may be allocated to both accuracy and linguistic complexity simultaneously. So, it may be expected to observe increases in both accuracy and linguistic complexity as cognitive demands imposed on a learner are increased.

To date, research has not provided undeniable evidence in support of one or the other hypothesis. The current investigation has intended to provide support for the point of view that individual differences, particularly in attentional and working memory capacity, play a role in results of research having provided less than conclusive outcomes. To do so, collected data was analyzed in both as a whole, and split according to groups of participants who demonstrated similar levels of attentional and working memory capacity. Proficiency was also taken into account as it is expected to account for variance in dimensions of oral production. In this way, it could be observed whether individual differences in attentional and working memory capacity may also account for variation in dimensions of linguistic production which could be responsible for inconclusive observations found in task complexity research.

In the previous chapter, results of the statistical analysis of the collected data were described in detail. The rest of the present chapter will review and interpret the results in light of the research questions and hypotheses that were proposed earlier in the dissertation. Selected observations taken from the correlations made between CAF measures and task complexity values, repeated measures tests, and the correlational analyses between CAF measures will be discussed and conclusions drawn.

8.2 Research question 1:

The first research question assumes that where trade-off effects appear in the data, both high and low level attentional and working memory capacity subjects will be affected by them. The issue is, however, whether individual differences will influence how quickly the groups reach a threshold of attentional resource capacity resulting in trade-offs between CAF measures.

> ***Research question 1:****As learners perform a series of oral tasks that vary in level of cognitive complexity along a continuum from simple to more complex, what influence does working memory and attentional capacity have in determining the point along the continuum at which they begin to demonstrate competition for attentional resources between dimensions of fluency, accuracy and linguistic complexity?*

8.2.1 Correlations between task complexity and dimensions of CAF

Correlations between task complexity and CAF measures showed that as task complexity increased along the continuum of simple to complex, some measures of lexical and structural complexity showed improvement to the detriment of fluency. This was true for the non-split data. The result is supportive of the Trade-off hypothesis in that attentional resources appear to have been allocated only to linguistic complexity but not to accuracy. On the other hand, there was no observed decrease in dimensions of accuracy as might be expected if there was competition for resources as predicted by the hypothesis. Possibly task complexity pushed learners to prioritize complexity, but did not impose a sufficient degree of complexity to where accuracy was observably affected.

The split data indicates a few principle differences in task performance dependent on subjects' levels of proficiency, and working memory and attentional capacity. Firstly, structural complexity, particularly in the form of words per AS-unit, showed that there were positive, although weak, correlations with task complexity for the high level proficiency, and attentional and working memory capacity groups. At the same time, their lower level counterparts showed no correlations at all. Secondly, the split data showed significant correlations for one measure of accuracy, the corrected value for repaired to unrepaired errors. The non-split data showed no such correlation. While these correlations were not strong, they are supportive of the hypothesis that individual differences may be responsible for differences in the way subjects perform tasks. Indications along these lines have been demonstrated in previous research. Results of studies by Niwa (2000), Trebits & Kormos (2008), and Gilabert & Muñoz (2010) suggest that working memory capacity may play a role in fluency. Bergsleithner (2007) showed that differences in working memory were reflected in measures of accuracy. Finally, Borges Mota (2003) found results suggesting that differences in working memory capacity were reflected in measures of fluency, accuracy and linguistic complexity.

In the case of the current study, it was the low attentional capacity group, the high working memory capacity group, and the high proficiency level group which demonstrated that increased task complexity may have pushed these learners to greater accuracy along the dimensions of repaired to unrepaired errors. From this perspective, results would be in partial support of the Cognition Hypothesis, as measures of both complexity and accuracy show an increase. On the other hand, as only one measure of accuracy was affected, it might be suspected that an effect of task characteristics, rather than increased cognitive complexity, was involved in influencing these subjects to attend to accuracy along this particular dimension where their counterparts did not. Or it may be a case where these subjects selectively chose to focus on this aspect of accuracy. Skehan (2009b) suggested that learners may choose to direct attention to aspects of production as an economizing strategy to deal with task demands. Interestingly, splitting the data leads to different outcomes and different interpretations of the how task complexity may affect the results.

The importance of the observations drawn from the data lies primarily within the domain of investigation. Research carried out on a population which is heavily skewed in terms of levels of proficiency, or attentional or working memory capacity may show very different effects of task complexity on speech from a population which is not skewed, or skewed in the opposite direction. Borges Mota Fortkamp (1998), for example, found no significant correlations between measure of working memory capacity and measures of fluency which contrasts with several of the studies mentioned above. There is potential that this discrepancy could be due in part to differences in the makeup of the study populations.

The correlations between task complexity and CAF measures are not intricate enough to show where trade-offs exist along the continuum, but they do provide an indication that trade-offs are present, especially between fluency and linguistic complexity as task complexity increases. Further analysis provides more in-depth answers.

8.2.2 Repeated measures tests

For the non-split data, propositional complexity and lexical complexity were the measures which were most affected by increases in task complexity while accuracy was not affected. Greater task complexity produced more idea units in most cases, as well as greater lexical sophistication and richness. At the same time, fluency decreased significantly between most levels as task complexity increased. These results are supportive of the Trade-off Hypothesis. However, the most interesting results of the repeated measures tests in terms of determining a point along the continuum of simple to complex where attentional resources shifted focus may be found in the measure of idea units. Between consecutive levels of task complexity, there was a strong pattern of in the performance of certain groups as demonstrated by the split data.

The low proficiency group, the high attentional capacity group, and the high working memory group all performed identically in their use of idea units, showing significant differences between all consecutive levels from three to five. In all cases, level three elicited more idea units than level four, likely due to aspects of task three which required more concepts to be expressed in order for it to be narrated successfully as suggested by Skehan (2009) who says that task characteristics may be responsible for such an effect. The nature of task three may have required learners to reflect on background information that was not inherent in the story. Tavakoli & Skehan (2005) and Tavakoli & Foster (2008, 2011) found that the need to carry over background information into the narration of the task increased complexity.

The counterparts of these three groups also performed identically to one another, however, none of these showed significant differences between consecutive levels of task complexity until level five where they used significantly more idea units than level four. So, beginning at level three it is clear that high attentional and working memory capacity groups in addition to the low proficiency group were all affected by some aspect of the tasks while their counterparts were not affected until task complexity level five. In addition, high working memory capacity subjects showed more variation than their lower capacity counterparts in terms of significance between task complexity levels for the measure of propositional complexity. This may imply that differences in task complexity may affect high working memory capacity subjects to a greater degree than lower capacity subjects. This is supportive of the suggestion by the Cognition Hypothesis that differences in working memory and attentional capacity are most important in more cognitively demanding tasks.

As stated earlier in the dissertation, task complexity was operationalized in terms of increased intentional reasoning demands incorporating processes of evaluation, and psychological and convergent production which include creative thinking abilities while drawing heavily on working memory (Guilford, 1967; Kaufman & Lichtenberger, 1999). So, as intentional reasoning demands increased, subjects with greater working memory capacity as well as those with greater attentional capacity, attention being a construct of working memory, may be expected to produce a greater number of idea units if this involves processes requiring creative thinking abilities. This seems to have been demonstrated in current data. Groups of lower cognitive capacity may have found the creative thinking processes more difficult to handle resulting in resources allocated to another aspect of the narration which seems to have been accuracy as will be identified below. It is unclear whether this observation is the result of a trade-off between the measure of propositional complexity and accuracy. This, as well as a possible reason why the low proficiency group was also affected as it was, will be dealt with further on in the chapter.

8.2.3 Correlations between dimensions of CAF

There were no instances in which high or low levels of either proficiency, or attentional and working memory capacity indicated that they were affected by similar trade-off effects between measures of CAF at distinctly different points along the continuum of task complexity. So, the first research question remains for the most part unanswered. Although correlations between CAF measures and task complexity suggested that there was a trade-off observed between measures fluency and lexical and structural complexity as task complexity increased, this trade-off was not manifested universally for groups of high and low levels of proficiency, and working memory and attentional capacity in the correlational analysis between CAF measures. The first hypothesis which states that high level capacity subjects would manifest trade-off effects at a later point along the continuum of simple to complex has not been demonstrated.

What has become apparent however, are the different strategies which were employed by each of the groups in order to meet task demands. Attentional resources seem to have been allocated to propositional complexity, as measured by idea units, for some but not for all groups. This is an issue which is which will be discussed in light of the second research question.

8.3 Research question 2

The second research question assumes that where trade-off effects are manifested in the data, all subjects may be affected regardless of level of proficiency, and working memory and attentional capacity. The issue is whether individual differences in levels of attentional and working memory capacity are going to influence the intensity with which trade-off effects happen.

Research question 2: *What relationship does working memory and attentional capacity have with the degree to which trade-off effects appear in speech samples produced under increasingly complex conditions?*

8.3.1 Correlations between task complexity and dimensions of CAF

As described above, the correlations between task complexity and CAF established that as task complexity increased, measures of linguistic complexity increased, as well as for certain cases of accuracy to the detriment of fluency. These results were conditioned, however, by differences between levels of proficiency, and attentional and working memory capacity. The question as to whether there were significant differences in the correlations was answered with the Fisher's r-to-z transformation which indicated that coefficients between the split groups were not significantly different. This is contrary to what was predicted by the second hypothesis which stated that subjects of higher attentional and working memory capacity would manifest a distinction between measures where levels of individual differences were responsible for variation in the results of the tests. In this respect, the answer to the second research question is negative; it seems that differences in attentional and memory capacity do not affect the degree to which trade-off effects are manifested as demonstrated by differences between strengths of correlations between CAF measures and values of task complexity. As stated earlier, however, individual differences in these aspects do seem to affect strategies which subjects take in order to perform the tasks. This was manifest through the repeated measures tests and the correlational analysis between CAF measures.

8.3.2 Repeated measures tests

As task complexity increased from simple to complex, the repeated measures tests showed that some but not all measures demonstrated differences between levels of task complexity. Differences appeared for rate A fluency measures which generally decreased, and measures of linguistic complexity which generally increased with task complexity. This is a reflection of the correlational analysis between dimensions of CAF and task complexity. But once data was split, it was shown especially for measures of propositional complexity that proficiency level, and attentional and working memory capacity determined whether differences were significant. This supports the second hypothesis that states that individual differences may determine the degree to which results are manifested.

The multivariate ANOVA which was run on the rate A data suggested that the combined effects of working memory and attentional capacity accounted for approximately 23% of the variance in the results. This reflects findings in Niwa (2000), Trebits & Kormos (2008), and Gilabert & Muñoz (2010) who also attributed effects of fluency to working memory, but contrasts with Borges Mota Fortkamp (1998) who found no correlation between these two aspects.

In summary, as subjects performed tasks which increased in their level of conceptual complexity, fluency decreased and there was a tendency to use a greater variety of vocabulary and more sophisticated vocabulary. However, levels of cognitive capacity played a strong role in the variance of fluency between task complexity levels as well as determining whether a significantly different number of idea units were produced between different task complexity levels. This provides confirmation to the hypothesis that individual differences may play a role in manifesting research outcomes that may differ if they are not taken into consideration.

8.3.3 Correlations between dimensions of CAF

Examination of the correlations between various CAF measures provides insight into how attentional resources shifted as subjects performed the various tasks. Focus was primarily set on identifying trends where significant correlations appeared over two or more consecutive task complexity levels. It can be assumed that where correlations either begin or end, is an indicator that the focus of attentional resources has switched. At these points, comparisons could be made with data from the other tests in an attempt to find out where attentional resources may have been reallocated.

However, the correlational analysis between CAF measures did not put into evidence instances of clear trade-off effects between measures, at least not in the expected way. Where trends were identified in the non-split data, the split data showed that high and low proficiency, and attentional and working memory capacity subjects seemed, in fact, to perform differently from one another. Where a trade-off might be detected at a particular task complexity level in the non-split data, rarely was it universal among all subject categories. The few patterns that were identified seemed to imply that individual differences determined strategies that subjects used to complete the tasks, but they did not offer any reasonably clear indication of predictability as to how attentional resources might be allocated as a result of increasing task complexity. Some exemplary observations are discussed below.

8.3.3.1 Errors and error repair

The most outstanding connection was made in observing correlations between mid-ASU pauses, and errors and error repairs. As described in the previous chapter for the non-split data, the corrected measure of errors to error repairs correlated with pauses for the first two levels of task complexity, but then stopped. Nevertheless, pausing continued to correlate with other measures of error. Pausing is associated with lexical search, pronunciation, planning, or message conceptualization and formulation (Tavakoli, 2011). So, it could be assumed that while subjects were aware of error, less attentional resources were being dedicated to concern for their proper repair. A review of descriptive statistics for error measures provided a clue to what was happening.

Low proficiency subjects committed a greater number of errors than their high level counterparts at all task complexity levels for every error measure. But the working memory and attentional capacity groups did just the opposite. It was the high working memory and attentional capacity groups that were committing more errors than their low capacity counterparts. Reflections on why this is the case will be dealt with below.

Although significance was not always reached, a pattern appears in the trends where the subjects in the low proficiency group, the high attentional capacity group and the high working memory group all followed similar strategies in how they attended to accuracy. It seems that high working memory and attentional capacity subjects, as well as low proficiency subjects all allocated resources to another aspect of speech production to the detriment of accuracy while their counterparts did not. This phenomenon concerning dimensions of accuracy did not seem to be subject to higher or lower task complexity levels, rather general tactics that these groups of subjects appeared to rely on in order to meet task demands. It will be recalled from earlier in this section that these same groups are those that seemed to focus attention to the production of idea units.

8.3.3.2 Lexical richness

For the high working memory capacity group, pausing correlates negatively with D for the first two levels of task complexity. On the other hand, for the low working memory capacity group, pausing correlates positively with D for the first three levels. For the low working memory group, pausing correlates with little else than with D. So, as per Tavakoli (2011), it may be assumed that for the subjects in the low working memory capacity group, pausing is associated with lexical search; subjects paused in order to find the best words to use in order to convey the message. Contrarily, the high working memory group seems to be allocating resources differently.

For the high capacity working memory group, while pausing correlates negatively with the measure for lexical richness, it corresponds positively for measures of error repair and the corrected value of repaired to unrepaired errors. So, for the high working memory capacity group, there seems to be an awareness of accuracy issues as mentioned earlier, while the concern of low capacity group is directed toward lexis. It might be speculated that by drawing concern away from accuracy and toward lexis allowed the low capacity working memory group to lose inhibitions or to minimize uncertainties about accuracy issues. Accuracy could be maintained through processes of automatization while attentional resources are economized allowing for a focus on lexical search as both accuracy and lexical search concern processes which form a part of the Formulator stage of Levelt’s speaking model (1989). Economization of resources at this level would allow for more efficient processing of the message. In contrast, the high capacity group continued to commit more errors than the low capacity group as seen in the last section, despite concern afforded to accuracy. It is to be recalled that while the high working memory capacity group demonstrated less concern for lexical variety, but concern for accuracy, attentional focus appeared to have been directed toward the creation of idea units as described above. A greater number of idea units indicates attentional focus directed toward the Conceptualizer stage of Levelt's model as attention is directed toward generation of the pre-verbal plan. Though greater cognitive capacity enjoyed by this group could afford some attention allocated to awareness of accuracy issues, the real concern seems to have been place on the meaning that was to be conveyed through the message, rather than on the form. This would confirm Skehan's proposal that meaning is prioritized above form (Skehan, 1998, Skehan & Foster, 2001).

The phenomenon just described does seem to be limited to the lower levels of task complexity. As of task complexity level three, the correlations between pausing and D, as well as between pausing and error repairs and the corrected measure of repaired to unrepaired errors end for the high working memory capacity group, as do correlations between pausing and D for the low capacity group as of level four. At these points, attentional resources seem to have been allocated to discourse aspects of the narration. This will be discussed further below.

8.3.4 Summary

Kormos (2006; 132) explains from a review of literature on development of competence and metalinguistic awareness in L2 that language learners of a lower proficiency tend to make more mistakes and correct a smaller proportion of those mistakes than learners of higher proficiency. The current data corroborated with the findings about number of errors for the low proficiency group. It also corresponds to the high attentional and working memory capacity groups. Further, Kormos explains that findings indicate that as learners develop in their interlanguage, greater automatization allows them to shift attentional resources away from metalinguistic concerns about accuracy and toward issues arising at the discourse level. Observations described above suggest that attentional resources were diverted from several different aspects of task performance to conveyance of a greater number of messages while narrating the stories for the low proficiency group, the high attentional capacity group, and the high working memory capacity group.

There is an indication that attentional resources began to be reallocated to other aspects of the discourse where pausing stopped correlating with the measure for corrected repaired to unrepaired errors at complexity level two for the non-split data as described above. At level three, significant differences between task complexity levels for measures of lambda began to appear universally. This occurred as well at level three for the measure of idea units although, in this case, the occurrence is not universal. In the case of idea units, these significant differences occur for high attentional and working memory capacity groups, as well as for low proficiency subjects, but not for their counterparts. Again, these correspond to the same groups that committed greater amounts of errors than their counterparts. It may be conjectured that subjects of higher working memory and attentional capacity, as well as subjects of low proficiency allocated more resources to discourse and the conceptualization of the narrative, neglecting accuracy.

Differences between performance of high and low proficiency and attentional and working memory capacity groups may to be linked to a strategy which could be explained in part with Skehan’s (2009) reference to Levelt’s model used as a framework for understanding speech production under varying task conditions. The speaking model separates speech production into conceptual areas which relate to the Conceptualizer stage, and linguistic areas which relate to the Formulator stage. Certain influences imposed by the task connect with the pre-verbal message developed during conceptualization. Other influences connect with how the message is formulated for expression. Limitations in cognitive resource capacity are linked to how these two stages function together smoothly and effortlessly, but subjects, regardless of their cognitive capacities, may target accuracy or linguistic complexity as they look for the easiest way to perform the task. According to Skehan (2009b: 210), studies have suggested that subjects can prioritize specific performance areas, according to difficulty of the demands that the task imposes. So, subjects who prefer to focus on developing a narrative by conveying more messages will revert more frequently to the Conceptualizer stage where the communicative intent of each message is created. This may be reflected by use of a greater number of idea units.

Studies within the field of cognitive psychology have demonstrated a link between working memory and attentional capacity with measures of creativity (e.g. DeDreu *et al*, 2012; Chiappe & Chiappe, 2007). According to Ellis and Barkhuizen's (2005) suggestion that a calculating a number of idea units provides a measure of the extent to which a speaker engages in conceptualization as concepts are created for encoding, then idea units can be considered an expression of creative thinking within a narrative. It would follow that learners who exhibit a higher level of working memory capacity might be expected to tend toward creative thinking processes and prioritized meaning over form while reverting frequently to the Conceptualizer stage of Levelt's model as communicative intentions are transformed into pre-verbal plans.

Elaboration of communicative intention is attentional resource depleting as per Levelt (1989), so high working memory and attentional capacity subjects may be more adapt at conceptualizing of a larger number of idea units than their lower capacity counterparts. Each concept expressed by a speaker depends on what has been said before. For a narration to be coherent, the speaker must rely on bookkeeping, or storage, to make relevant connections between what they want to say and what has already been expressed (Levelt, 1989). Frequent reversion to the process of message conceptualization may deplete resources to the extent where the quality of message formulization is diminished. This may account for poor accuracy on the part of subjects who produced more idea units. On the other hand, low working memory and attentional capacity subjects may have found that avoiding the creation of new messages allowed for economization of cognitive resources which they have less of. This would free resources for greater attention to message formulization resulting in more accurate expression of a smaller number of concepts.

The same strategy may have been used by the low proficiency subjects, prioritizing the content of the story over accuracy. Subjects with high proficiency did not demonstrate this, possibly because automaticity of processes at the levels of formulization allows for greater accuracy and for cognitive resources to be allocated toward fluency; these subjects demonstrated significantly faster speech than low proficiency learners. This follows Finardi and Weissheimer (2008) who found that as proficiency level increases, more L2 production processes become automatized freeing up working memory resources that are required for controlled executive processes.

Different kinds of strategies may result from individual differences in cognitive capacities. Where the strategies employed may be consciously chosen by a learner, that learner might remain unaware of the underlying reason why they prefer to do so. From the learner's perspective, it may simply be easier to perform the task in the chosen way. Nevertheless, if by conscious choice subjects of low proficiency, and high working memory and attentional capacity focus attention on propositional or lexical complexity while their counterparts prefer to focus on accuracy, the results of research would be affected if that research involved comparisons between two skewed populations.

8.4 Conclusions

The research questions as well as their corresponding hypotheses proposed in the current dissertation assumed that differences in task complexity would affect subjects indiscriminately of their levels of proficiency, and working memory and attentional capacities. So, where non-split data showed effects of differences in task complexity, these effects were assumed to be manifested in all the data despite the division into high and low level groups. The proposals of the hypotheses, however, were that the effects would be manifested differently for each of the groups in terms of their point of salience along the continuum of task complexity, and in terms of the intensity with which effects were manifested.

The data analysis revealed a very different situation. The non-split data was not an accurate indicator of how each of the groups performed individually. In addition, on rare occasions did groups perform in such a way that practical comparisons could be made between how task complexity affected them differently as proposed by the research questions. Consequently, the first research question remains primarily unanswered.

This notwithstanding, the investigation did provide insight into how individual differences may affect dimensions of performance. The between measures analysis demonstrated that statistical significance in differences between task complexity levels were not universal between subjects of different proficiency levels, and attentional and working memory capacities. Also, level of cognitive capacity played a role in the difference in the number of syllables produced per minute between task complexity levels as well as whether a significantly different number of idea units were produced between different task complexity levels. In addition, the correlations between measures of CAF and levels of task complexity demonstrated that only when the data was split, did evidence appear that could be argued in favor of the Cognition Hypothesis. An apparent trade-off between rate A fluency and the corrected measure of repaired to unrepaired errors appeared where it didn't for the non-split data.

The second research question, therefore, has been partially answered. Working memory and attentional capacity do appear to play a role in the way that subjects perform tasks which is reflected in dimensions of CAF. Repeated measures analyses reveal that individual differences may determine the degree to which linguistic aspects which result from manipulation of task characteristics become salient in the data. The data as it was handled in the current investigation, however, did not afford the opportunity to demonstrate the degree to which individual differences may have influenced trade-offs differently. Nevertheless, the observations do suggest that in data where trade-offs between measures of

CAF appear, there is potential for them to appear differently depending on differing levels of cognitive complexity.

An interesting conclusion drawn from the observations made during data analysis points toward the potential influence of individual differences on communicative strategies that subjects adhere to in accordance with their cognitive capacities. This conclusion is based on similarities between how three groups performed in similar ways as they prioritized propositional complexity to the detriment of accuracy while their counterparts maintained a focus on accuracy. It is unclear as to whether this was the result of a trade-off between accuracy and complexity where task complexity influenced the allocation of attentional resource in contradiction predictions made by the Cognition Hypothesis. It appears to be the result of a chosen strategy chosen which would likely result in an economization of cognitive resources to the benefit of narrative efficiency, but a strategy which did not necessarily come about because of changes in task complexity.

Some questions may be raised: How much is task performance affected by the employment of communicative strategies rather than by cognitive resources being subconsciously allocated to meet task demands as a result of increased task complexity? Can strategies be predicted based on external measures of cognitive complexity? If strategies do present an important influence in task performance, and it is possible to predict what kind of strategies are used by learners who share levels of cognitive capacities, then how can strategies be directed by instructors to benefit the acquisition process? It may be interesting to investigate whether individual differences in proficiency, and attentional and working memory capacity do indeed affect the predictability of the employment of communicative strategies and to what degree these can influence task performance.

8.5 Implications for research

The inspiration behind the current research lay in inconsistent findings of previous research about how differences in task complexity affect dimensions of performance. The underlying rationale was that task performance might not be consistent between individuals because of differences in their working memory and attentional capacities. If the differences in performance dimensions were great enough between individuals, then they could feasibly cancel each other out to some degree, resulting in data which would provide little or erroneous information. On the other hand, if subjects are divided into groups of high and low working memory and attentional capacity, and the subjects within these groups performed tasks in a more homogeneous manner, then data would shed a clearer light on the influence that increased task complexity might have on performance.

Results of the current research indeed show that there may be variations in the way that subjects approach task performance based on differences in their levels of proficiency, and attentional and working memory capacity. In cases, there is potential that increased task complexity influences the approach taken by some subjects. This is demonstrated by the way that subjects seemed to adopt different strategies to cope with the way they perform a task. The strategies may not necessarily be a result of demands imposed by increased task complexity, but rather because these strategies enable subjects either to perform tasks successfully within their capacities, or to perform the task with greater efficiency so that cognitive resources can be allocated to other aspects of the task. This is best demonstrated in the way that the low proficiency and high attentional and working memory capacity groups performed differently than their counterparts in terms of accuracy. While subjects within each of these groups produced less accurate speech than their counterparts, the effect did not seem to be linked to changes in task complexity.

The implications are that there is an apparent variety in how subjects approach task performance resulting from individual differences and that this may affect the comparability of results between experiments. If a population is skewed heavily enough in favor of one or another group which inherently performs tasks differently from the population of another research project, then one can certainly expect to find conflicting results. This could account for some of the inconsistencies present in previous task complexity research.

8.6 Pedagogical implications

Sequencing of tasks within a TBLT syllabus is intended to stimulate the acquisition process as students consolidate what they learned from attempts at tasks performed previously into their performance of later tasks. Moving from simple to more complex tasks can motivate learners to attempt more ambitious language as they try to meet the demands that the more challenging task imposes on them (Robinson, 2011). Task and syllabus designers would benefit from the potential to anticipate the results of task manipulation and sequencing to carry out the intentions of the TBLT methodologies in an effective and efficient manner. Of course, the variety of learner characteristics makes this a formidable task, and especially so when implications of research must be generalized so that they can be applied to classroom settings.

The study of individual differences and how these influence language production during task performance could allow task and syllabus designers to create tasks and syllabi that cater to learners who share similar characteristics and present them with effective learning opportunities. However, links must be made between individual character factors which can anticipate capacity of attainment and linguistic performance on tasks. Indications from the current research suggest that such links may be accessible to researchers and future investigation may be able to identify more clearly how learners which share specific individual differences approach tasks in foreseeable ways.

8.7 Future research

The conclusions which were drawn from the observations of the statistical analyses described above might impel future research to continue to investigate the influence of individual differences on results in task complexity research. Results of the current investigation suggest that differences in cognitive complexity may very well play a determining role in distinguishing whether data favors one or another position as concerns the effects of increased task complexity on linguistic performance.

Secondly, observations of the potential role that individual differences may play in learners' choices of communicative strategies as they search for the most efficient way to approach a task raised interesting questions in this respect. Questions involved the degree to which communicative strategies influenced analysis of performance dimensions, whether kinds of strategies could be predicted base on measures of cognitive capacities, and finally, whether strategies could be managed in a way that they could aid in the language acquisition process. Future research could attempt to identify whether learners prioritize between either meaning or form to meet task demands in the most efficient way possible within the allowances of their cognitive capacities, but based on a conscious awareness of ease of communication. If this is the case, researchers may wish to understand the degree to which this may affect results of research which investigates the cognitive processes responsible for directing attention to aspects of the language which best allows learners to meet task demands.

8.8 Limitations

The greatest limitation of the research described in this dissertation was due to the volume of data which was handled. This limited the researcher’s capacity to efficiently investigate all possible affects that may have been relevant to the investigation. This issue notwithstanding, an attempt was made to identify those elements which were most indicative of the effects of task complexity on dimensions of CAF for each of the groups that were studied. Future research on the same data should focus on reducing the amount of data which is to be treated.

A second limitation can be found in the operationalization of the continuum of task complexity as represented by the picture stories used in the investigation. Although it was determined that task complexity was operationalized properly, the indicators which resulted from the analysis of the data drawn from the pilot study were not entirely conclusive. The task was designed as a picture arrangement task for use in the WAIS-III, but not as a narrative task. As a picture arrangement task, the stories which form the different tasks are generally, although not universally (Costello & Connolly, 2005), considered to represent a continuum from simple to complex. As a narrative task, while the stories may form a continuum of complexity along some dimensions, it is quite feasible that the same series of stories may form a different continuum of complexity as they concern other dimensions. So, as a matter of example, where the sequence of stories may form a proper continuum from simple to complex for dimensions of fluency, it is quite possible that it is not a proper continuum for dimensions of linguistic complexity or accuracy due to particularities of characteristics inherent in the tasks. Révész (2014) stresses the importance that task versions which are designed to be more cognitively complex are indeed so. It would be interesting for future research to have tasks available which have been demonstrated to represent a clear and unquestionable continuum for all dimensions.

A third limitation may be found in the kinds of measures which were used in the study. It is possible that other measures of CAF may have shed clearer light on trade-off affects that were present in the narrative performances, but which did not become salient through the dimensions that were employed.

A fourth limitation of the study concerns the number of subjects that were studied. The entire range of subjects numbered forty-seven, but when these were split into two smaller groups the corresponding analytical results would become less reliable. This could explain why observed phenomenon was often unexpectedly not shared between non-split and split data.

References:

Ahmadian, M. J., Tavakoli, M., and VahidDastjerdi, H. (2015). The combined effects of online planning and task structure on complexity, accuracy and fluency of L2 speech. *The Language Learning Journal*, *43*(1), 41-56.

Amador Campos, J.A. (2013). *La escala de inteligencia de Wechsler para adultos, cuarta edición (WAIS-IV).* Documento de trabajo. Facultat de Psicologia. Universitat de Barcelona.

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

Arbuthnott, K. and Frank, J. (2002). Trail Making Test, Part B as a Measure of Executive Control: Validation Using a Set-Switching Paradigm. *Journal of Clinical and Experimental Neuropsychology, 22(4),* 518-528.

Baddely, W. (1986). *Working memory*. Oxford, England: Oxford University Press.

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

Baddeley, A.R. (2003).Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208

Baddeley, A.R. and Hitch, G. (2006). Is the binding of visual features in working memory resource demanding? *Journal of Experimental Psychology: General* 2006, Vol. 135, No. 2, 298–313

Baralt, M. (2010). *Task complexity, the Cognition Hypothesis and interaction in CMC and FTF environments*. Unpublished PhD dissertation, Georgetown University.

Baron-Cohen S., Leslie, L., and Firth, U. (1986). Mechanical, Behavioural, and Intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113-125.

Bergsleithner, J.M. (2007). Working Memory Capacity, Noticing, and L2 Speech Production. Unpublished Doctoral Dissertation, Universidade Federal de Santa Catarina.

Berlin, B., and Kay, P. (1969). *Basic color terms: Their university and evolution*. California UP.

Bialystok, E. and Martin, M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science 7:3* 325–339.

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Bliss, T. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, *232* (2), 331–356.

Block, R. A., Hancock, P. A., and Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *ActaPsychologica, 134,* 330-43.

Borges Mota Fortkamp, M. (2007). Relationship among individual differences in working memory capacity, noticing, and L2 speech production. *Signo*. *32*(52), 40-53.

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

Borges Mota Fortkamp, M. (1998). Working memory capacity and L2 oral fluency. *Ilha do Desterro*, 35, 201-238.

Broadbent, D.E. (1958). *Perception and Communication*. Pergamon Press.

Brown, S.(1985). Time perception and attention:The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception & Psychophysics, 38 (2),* 115-124.

Bygate, M., P. Skehan, M. Swain (Eds.). (2001). *Researching pedagogical tasks: second language learning, teaching, and assessment*. London: Pearson.

Carpenter, P.A. and Just, M.A (1989). The Role of Working Memory in Language Comprehension. In Kahr, D. and Kotowsky, k. (Ed.) *Complex Information Processing. The impact of Herber A. Simon*, (p.p. 31-68). Lawrence Erlbaum Associates Publishers.

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. *Literacy, language, and learning: The nature and consequences of reading and writing*, *105*-123.

Cheng, P.W. and Pachella, R.G. (1984). A psychophysical approach to dimensional separability. *Cognitivepsychology16*, 279-304.

Chiappe, D. L., and Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, *56*(2), 172-188.

Collins, A.M., Loftus, E.F. (1975) A Spreading-Activation Theory of Semantic Processing. *Psychological Review, 82*(6), 407-428.

Conway, A.R.A. and Engle, R.W. (1994). Working Memory and Retrieval: A Resource-Dependent Inhibition Model. *Journal of Experimental Psychology: General*, *123*(4), 354-373.

Costello, R. M., and Connolly, S. G. (2005). Item difficulty scaling for WAIS-III picture arrangement. *Journal of clinical psychology*, *61*(6), 781-786.

Cowan, N. (1993). Activation, attention and short-term memory. *Memory and Cognition, 21*, 162-167.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research, 169*, 323-338.

Cromer, R. (1974). The development of language and cognition: the cognition hypothesis. In *New Perspectives in Child Development*, Brian Foss (ed.), 204-252. Cambridge: Harmondsworth: Penguin.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*(4). 367-383.

Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422-433.

De Bot, K. (1992), *A bilingual production model: Levelt's 'Speaking' model adapted*, Applied Linguistics, 13, 1-24.

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, *1*(1), 17-32.

Declerck, M., and Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, *15*(04), 782-796.

De Dreu, C. K., Nijstad, B. A., Baas, M., Wolsink, I., and Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, *38*(5), 656-669.

Dehn, M.J. (2008). *Working Memory and Academic Learning: Assessment and Intervention*. Wiley. Hoaboken, NJ.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283-321.

Dixon, P. and Just, M.A. (1978). Normalization of Irrelevant Dimensions in Stimulus Comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 36-46.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA review*, *19*(1), 42-68.

Doughty, C.(2001). Cognitive underpinning of focus on form. In Robinson, P. (ed.), *Cognition and Second Language Instruction*, (p.p 206-257). Cambridge: Cambridge University Press.

Doughty, C. J., Campbell, S. G., Mislevy, M. A., Bunting, M. F., Bowles, A. R., and Koeth, J. T. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. In M. T. Prior, Y. Watanabe, and S-K. Lee (Eds.), *Selected proceedings of the 2008 Second Language Research Forum*, (p.p. 10-31). Somerville, MA: Cascadilla Proceedings Project. Retrieved fromhttp://www.lingref.com/cpp/slrf/2008/paper2382.pdf

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., and Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*(2), 105-132.

Dzaack, J., Trösterer, S., Pape, N., and Urbas, L. (2007). A computational model of retrospective time estimation. *Cognitive Systems Research, 8,* 208-215.

Ellis, N. (2001). Memory for language. In *Cognition and Second Language Instruction*. Robinson, P. (Ed.), (p.p 33-68). Cambridge University Press. Cambridge.

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and 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.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19-23.

Eviatar, Z. (1998). Attention as a psychological entity and its effects on language and communication. In B. Stemmer and H. A. Whitaker (Eds.), *The handbook of neurolinguistics, (*p.p. 275- 287). London: Academic Press.

Fink, A., and Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence, and time estimation as an index of cognitive load. *Personality and Individual Differences, 30,* 1009-1021.

Finardi, K. (2008). Working memory capacity and L2 speech production in a picture description task with repetition. *Revista de Estudios da Linguagem,16*(2), 129-144.

Finardi, K. and Weissheimer, J. (2008). On the relationship between working memory capacity and L2 speech development. *Signotica, 20*(2), 335-389.

Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak R.S., Frith, C.D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*(2), 109-28.

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

Foster, P. and Tavakoli, P. (2009). Lexical diversity and lexical selection: A comparison of native and non-native speaker performance. *Language Learning*, *59*(4), 866–896.

Fortkamp, M. (1999). Working Memory Capacity and L2 Speech Production. Retrieved from ERIC database.

Fraisse, P. (1984). Perception And Estimation Of Time. *Annual Review of Psychology 35*, 1-36.

Fukuta, J.and Yamashita, J. (2015). Effects of cognitive demands on attention orientation in L2 oral production. *System*, *53,* 1-12.

Garner, W. R. (1974) *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum,

Garner, W. R. (1976). Interaction of Stimulus dimensions in Concept and Choice Processes. *Cognitive Psychology 8*, 98-123.

Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *Research notes*, *12,* 8-10.

Gilabert, R. (2005). Task complexity and L2 narrative oral production. Unpublished PhD. dissertation. University of Barcelona, Spain.

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

Gilabert, R., Baralt, J. (2013). The impact of increasing task complexity on L2 pragmatic moves.In *Second Language Interaction in Diverse Educational Contexts, Language learning and language teaching*, Eds. Kim McDonough, Alison Mackey, (p.p. 45-70). John Benjamins Publishing.

Gilabert, R., Barón, J., and Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance*, (p.p. 105-140). Amsterdam: John Benjamins.

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

Guará-Tavares, M. G. (2009). The relationship among pre-task planning, working memory capacity, and L2 speech performance: a pilot study. *Linguagem & Ensino*, *12*(1), 165-194.

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill, New York.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279-1296.

Hardan, A. A. (2013). Language Learning Strategies: A General Overview. *Procedia-Social and Behavioral Sciences, 106*, 1712-1726.

Hicks, R., Miller, G.W., and Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology, 89*, 719-30.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In Housen, A. Kuiken, F., and Vedder, I.(Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, (p.p. 21-46). John Benjamins Publishing.

Ishikawa, T. (2007). The effect of manipulating task complexity along the +/- here and now dimension on L2 written narrative discourse. In García-Mayo, M.P. (Ed.). *Investigating Tasks in Formal Language Settings* (p.p. 136-156). Clevedon: Multilingual Matters.

Iwashita, N., McNamara, T., and Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language learning*, *51*(3), 401-436.

Jersild, A.T. (1927) Mental set and shift. *Archives of Psychology*, Whole No. 89.

Kahneman, D. (1973). *Attention to Effort*. Prentice-Hall. New Jersey.

Kane, M. J., Bleckley, M. K., Conway, A. R., and Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169.

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

Kaufman, A. S., and Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. John Wiley & Sons Inc. New York.

King, J., and Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of memory and language*, *30*(5), 580-602.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning, 50*, 343–384.

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

Kormos, J., and Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*(02), 261-271

Kuiken, F., and Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 261-284.

Kuiken, F., and 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.

Kuiken, F., Mos, M., and Vedder, I. (2005). Cognitive task complexity and second language writing performance. *Eurosla yearbook*, *5*(1), 195-222.

Leslie, A.M. (1987). Pretense and representation: the origins of "Theory of Mind". *Psychological Review, 94,* 412-426.

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

Levelt, W. J. (1999). Producing spoken language: A blueprint of the speaker. In *The neurocognition of language* (p.p. 83-122). Oxford University Press.

Lømo, T. (2003). The Discovery of Long Term Potentiation. In *Philosophical Transactions: Biological Sciences. Long-Term Potentiation: Enhancing Neuroscience for 30 Years.* Vol 358, No. 1432, (p.p. 617-620). The Royal Society.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In Hyltenstam, K. and Pienemann, M. (Eds.), *Modelling and assessing second language acquisition,* (p.p. 77-79). Clevedon, England: Multilingual Matters.

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

Long, M. H., and Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL quarterly*, *26*(1), 27-56.

Long, M. H. and Robinson, P. (1998). Focus on form Theory, research, and practice. In Doughty, C. and Williams, J. (Eds). *Focus on form in classroom second language acquisition,* (p.p.15-41). Cambridge University Press.

Luo, L., Luk, G., and Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29-41.

Lynch M.A. (2004). Long-Term Potentiation and Memory. *Physiological Review,* 84, 87–136.

MacCleaod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109,* 163-203.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Mahwah, New Jersey. Lawrence Erlbaum Associates.

Meara, P. M. (2005). X_Lex: the Swansea vocabulary levels test. v2. 05. *Swansea: Lognostics*.

Meara, P. and Bell, H. (2001) P-Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect 16(3),* 323-337.

Meara, P.M. and Miralpeix, I. (2006). Y_Lex: the Swansea Advanced Vocabulary Levels Test. v2.05. Swansea: Lognostics.

Meara, P. M., and Miralpeix, I. (2007). D_Tools. Version 2.0. *Computer software]. Swansea: Lognostics.(Source: http://www. lognostics.co uk/)*.

Meara, P. M., and Milton, J. (2003). X_Lex: the Swansea vocabulary levels test. *Swansea: Lognostics*.

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

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (p.p. 141–173). Philadelphia/Amsterdam: John Benjamins.

Michel, M. C. (2013). The use of conjunctions in cognitively simple versus complex oral L2 tasks. *The Modern Language Journal*, *97*(1), 178-195.

Michel, M., Gilabert, R. and Révész, A. (under review). Eye-tracking and stimulated recall as means of investigating cognitive load and processes during L2 tasks. Language Learning

Michel, M. C., Kuiken, F., and Vedder, I. (2012). Task complexity and interaction: (combined) effects on task-based performance in Dutch as a second language. *Eurosla Yearbook*, *12*(1), 164-190.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miralpeix, I. and Meara P.M. (2014). Knowledge of the written Word. In J. Milton and T. Fitzpatrick (eds.) *Dimensions of vocabulary knowledge*, (p.p. 30-44). Basingstoke: Palgrave Macmillan.

Mislevy, M., Annis, R., Koeth, J., Campbell, S., Linck, J., Bowles, A., Doughty, C. (2009). *Final Hi-LAB Assessment Utilization Argument*. University of Maryland Center for Advanced Study of Language.

Miyake, A. and Shah, P. (1999). *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press, N.Y.

Muñoz, C. (2010). On how age affects foreign language learning. *Advances in research on language acquisition and teaching.* 39-49.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics,* 463-482.

Nee, D. E., and Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Sciences, 105*(37), 14228-14233.

Niwa Y. (2000) Reasoning demands of L2 tasks and L2 narrative production: Effects of individual differences in working memory, intelligence, and aptitude. Unpublished M.A. dissertation, Aoyama Gakuin University, Tokyo.

Norris, J. M., and Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

Norman, D.A, and Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R.J., Schartz, G.E., and Shapiro, D. (Eds.), *Consciousness and self-regulation*. Vol. 4, (p.p. 1-18). New York: Plenum.

Nuevo, A. (2006). Task complexity and interaction. Unpublished PhD. dissertation, Georgetown University, USA.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition,* 21(1), 109-148.

Paas, F., Tuovinen, J. E., Tabbers, H. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38,* 63-71.

Perdue, C. (1993). Adult language acquisition: Crosslinguistic perspectives Vols. 2: The results. Cambridge: Cambridge University Press.

Pica, T., Kang, H. S., and Sauro, S. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition, 28*(02), 301-338.

Pignot-Shahov, V. (2012). Measuring L2 Receptive and Productive Vocabulary Knowledge. In: Ciarlo and, C. and Giannoni, D.S.(eds.) *Language Studies Working Papers*, Vol. 4, (p.p. 37-45). University of Reading.

Posner, M.I. and Peterson, E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25-42

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515-526.

Prior, A., and MacWhinney, B. (2010). A bilingual advantage in task-switching. *Bilingualism: Language and cognition*, *13*(02), 253-262.

Rahimpour, M. (1997). Task condition, task complexity and variation in L2 discourse. *Unpublished Ph. D. dissertation, University of Queensland, Australia*.

Revesz, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*(s1), 162-181.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied linguistics*, *35*(1), 87-92.

Robinson, P. (1995a). Attention, Memory and the "Noticing Hypothesis". *Language Learning, 45*(2), 283-331.

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

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In P. Robinson (Ed.) *Cognition and second language instruction.* (p.p. 287-318). Cambridge: Cambridge University Press.

Robinson, P. (2001b). Task Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework. *Applied Linguistics* 22, 1. 27-57.

Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA. *Individual differences and instructed language learning*, 113-133.

Robinson, P. (2003a). The Cognition Hypothesis, task design, and adult task-based language learning. Second Language Studies, 251, 2. 45-105.

Robinson, P. (2003b). Attention and memory during SLA. In *Handbook of Second Language Acquisition*, Catherine Doughty and Michael Long (eds.), (p.p. 631-678). Oxford: Blackwell.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a triadic componential framework for second language task design. *IRAL. International Review of Applied Linguistics in Language Teaching*, 43, 1, 1-32

Robinson, P. (2007). 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 (IRAL), 45,* 193-213.

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, *61*(s1), 1-36.

Robinson, P. and Gilabert, R. (2007). Task complexity, the Cognition Hyothesis and second language learning and performance. *IRAL* 45; 161-176.

Robinson, P., and Urwin, J. J. (1995a). Investigating second language task complexity. *RELC Journal*, *26*(2), 62-79.

Robinson, P. (1995a). Attention, Memory and the "Noticing Hypothesis". *Language Learning*, *45*(2), 283-331.

Rogers, R.D., and Monsell S. (1995). Costs of a Predictable Switch between Simple Cognitive Tasks. *Journal of Experimental Psychology: General*, Vol. 124, No. 2, 207-231.

Rubin, O. and Meiran, N. (2005). On the Origins of the Task Mixing Cost in the Cuing Task-Switching Paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1477–1491.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11,* 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics 13,* 206-226.

Schmidt, R. (2001). Attention. In *Cognition and Second Language Instruction*. Robinson, P. (Ed.) Cambridge University Press. Cambridge. Pp 3-32.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, and I. Walker, *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N., and Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory& Cognition*, *33*(4), 644-653.

Sheen, R. (1994). A critical analysis of the advocacy of the task-based syllabus. *Tesol Quarterly*, *28*(1), 127-151.Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17,* 38-62

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

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

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

Skehan, P. (2009b). Models of speaking and the assessment of second language proficiency. In Benati, A. (ed.), *Issues in Second Language Proficiency*, (pp. 202-218). Continuum.

Skehan, P. and Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, v. 1,* n. 3, pp. 185-211.

Skehan, P., and Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language learning*, *49*(1), 93-120.

Skehan, P., and Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction*, (pp. 183-205). Cambridge: Cambridge University Press.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In Slobin, D. I. (Ed.), *The crosslinguistic study of language acquisition. Vol. 2: Theoretical issues* (pp. 1157-1259). Hillsdale, NJ: Erlbaum.

Slobin (1993). Adult language acquisition: a view from child language study. In Perdue, C. (Ed.), *Adult language acquisition: Crosslinguistic perspectives Vols. 2: The results*, (p.p. 239-252). Cambridge: Cambridge University Press.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

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

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

Tavakoli, P., and Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 61*(1), 37-72.

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

Ting, S.C.C. (1996). Planning time, modality and second language task performance: The Effect of Narrative Type on Learner Output. *Language Learning, 58*(5), 439-473.

Tomlin, R and Villa, V (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183-203

Trebits, A. and Kormos, J. (2008). Working Memory Capacity and Narrative Task Performance. 33rd International LAUD Symposium – Cognitive Approaches to Second/Foreign Language Processing: Theory and Pedagogy. March 10-13. Landau / Pfalz, Germany.

Treisman, A. and Gelade, G. (1980). "A feature-integration theory of attention." *Cognitive Psychology*, *12*(1), 97–136.

Tulsky, D. S. (2003). *Clinical interpretation of the WAIS-III and WMS-III*. Academic Press.

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

Vasylets, O., Gilabert, R. and Manchón, R.M. (Forthcoming). The Effects of Mode and Task Complexity on L2 Production.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. (4th ed.) Baltimore: Williams & Wilkins.

Weissberger, G. H., Gollan, T. H., Bondi, M. W., Clark, L. R., and Wierenga, C. E. (2015). Language and task switching in the bilingual brain: Bilinguals are staying, not switching, experts. *Neuropsychologia*, *66*, 193-203.

Weissberger, G. H., Wierenga, C. E., Bondi, M. W., and Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging*, *27*(4), 959.

Wickens, C.D. (1989). Attention and skilled performance. In *Human Skills,* Dennis Holding (ed.), (pp. 71–105). New York: John Wiley.

Wickens, Christopher D. (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

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

Wong W.L.H., Dubey-Jhaveri, S., and Wong, L.T. (2015).The Dilemma of Matching Learning Styles and Teaching Styles in English Language Classrooms. In A. Dubey-Jhaveri and L.T. Wong (Eds.), *English Language Education in a Global World: Practices, Issues and Challenges.* New York: Nova. Pp. 121-133.

Yuan, F., and 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.

APPENDIX I

Name: ____

Age: ______________

Please indicate how easy or hard each exercise was for you to do by circling a number 1 (easy) through 9 (hard):

Task 1

Picture arrangement

The pictures were easy to arrange	1	2	3	4	5	6	7	8	9	The pictures were hard to arrange

---

Narration

The story was easy to tell	1	2	3	4	5	6	7	8	9	The story was hard to tell

Task 2

Picture arrangement

The pictures were easy to arrange	1	2	3	4	5	6	7	8	9	The pictures were hard to arrange

---

Narration

The story was easy to tell	1	2	3	4	5	6	7	8	9	The story was hard to tell

Task 3

Picture arrangement

The pictures were easy to arrange	1	2	3	4	5	6	7	8	9	The pictures were hard to arrange

---

Narration

The story was easy to tell	1	2	3	4	5	6	7	8	9	The story was hard to tell

Task 4

Picture arrangement

The pictures were easy to arrange	1	2	3	4	5	6	7	8	9	The pictures were hard to arrange

---

Narration

The story was easy to tell	1	2	3	4	5	6	7	8	9	The story was hard to tell

Task 5

Picture arrangement

The pictures were easy to arrange	1	2	3	4	5	6	7	8	9	The pictures were hard to arrange

---

Narration

The story was easy to tell	1	2	3	4	5	6	7	8	9	The story was hard to tell

Nombre: ____
Edad: ________________

Por favor, indique la dificultad de cada ejercicio puntuando de 1 al 9; siendo el 1 la puntuación más fácil y el 9 la más difícil.

Actividad 1

Ordenar las viñetas

Las viñetas eran fáciles de ordenar	1	2	3	4	5	6	7	8	9	La viñetes eran difíciles de ordenar

---

Narration

La historieta era fácil de explicar	1	2	3	4	5	6	7	8	9	The story was hard to tell

Actividad 2

Ordenar las viñetas

Las viñetas eran fáciles de ordenar	1	2	3	4	5	6	7	8	9	La viñetes eran difíciles de ordenar

---

Narration

La historieta era fácil de explicar	1	2	3	4	5	6	7	8	9	La historieta era difícil de explicar

Actividad 3

Ordenar las viñetas

Las viñetas eran fáciles de ordenar	1	2	3	4	5	6	7	8	9	La viñetes eran difíciles de ordenar

---

Narration

La historieta era fácil de explicar	1	2	3	4	5	6	7	8	9	La historieta era difícil de explicar

Actividad 4

Ordenar las viñetas

Las viñetas eran fáciles de ordenar	1	2	3	4	5	6	7	8	9	La viñetes eran difíciles de ordenar

---

Narration

La historieta era fácil de explicar	1	2	3	4	5	6	7	8	9	La historieta era difícil de explicar

Actividad 5

Ordenar las viñetas

Las viñetas eran fáciles de ordenar	1	2	3	4	5	6	7	8	9	La viñetes eran difíciles de ordenar

---

Narration

La historieta era fácil de explicar	1	2	3	4	5	6	7	8	9	La historieta era difícil de explicar

APPENDIX II

Consentiment per participar en la investigació

Nom de projecte: *Working memory, task complexity, and competition for cognitive resources during L2 oral task performance.*

Investigador: James Pownall Tel: 610.165.217 Email: JamesWilliamPG@blanquerna.url.edu

Sponsor: Departament de Filologia Anglesa i Alemanya. Universitat de Barcelona

---

Introducció

L'estudi consisteix en observar com influeix la memòria del treball i la capacitat d'atenció en la manera amb què les persones fan servir l'anglès per fer una sèrie d'activitats que es diferencien entre si en el seu nivell de dificultat.

Per començar, es demanarà que facin unes proves cognitives i de nivell de coneixements de l'anglès amb l'objectiu de poder classificar els participants segons els resultats obtinguts.

L'activitat principal consisteix en posar en ordre una sèrie de dibuixos per tal que formin una historieta lògica per després narrar-la mentre que són gravats, quedant la seva veu enregistrada. Hi ha cinc historietes en total per gravar. Una vegada gravades les historietes, passaran a ser transcrites per després poder analitzar les seves característiques lingüístiques. Quan s'han obtingut totes les dades, es durà a terme una anàlisi estadística.

El que es demana:

1. Completar una prova de nivell d'anglès, els resultats dels quals es faran servir exclusivament per raons estadístiques de l'estudi i no tindran vigència oficial en cap altre cas.
2. Participar en una sessió de obtencióde dades. La sessió tindrà una durada de 1,5 hores aproximadament. La sessió inclourà:
   - Obtenció de dades personals (5 minuts)
   - Una segona prova amb ordinador de nivell de coneixements d'anglès (15 minuts)
   - Dues provesambordinador de mesures cognitives (45 minuts total)
   - Narració i gravació de les historietes amb l'investigador (15 minuts)

Beneficis

- Els/les voluntaris/es tindran una oportunitat de fer servir l'anglès en un entorn fora de l'àmbit d'una classe normal.
- Els/les voluntaris/es tindran l'oportunitat de participar en un estudi formal de lingüística cognitiva amb implicacions a nivell internacional en l'àrea d'investigació d'adquisició d'idiomes.

Confidencialitat

Tota la informació recollida es mantindrà confidencial i s'utilitzarà només per a fins de recerca. La seva identitat es mantindrà anònima i cap altra persona, a part del investigador, tindrà accés a la informació vinculada amb els noms dels voluntaris. En el cas que els resultats de l'estudi siguin publicats, els noms dels/les voluntaris/es no es faran servir. Les dades es guardaran dins d'un fitxer digital al qual tindrà accés només l'investigador.

Participació

Participació en l'estudi és completament voluntària. Si, en qualsevol moment, canvia d'opinió i decideix no participar en l'estudi, s'ha de comunicar-ho al'investigador i no continuarà com a participant en l'estudi. Per qualsevol consulta, contacti amb l'investigador per telèfon, correu electrònic.

Declaració del investigador

El/la voluntari/a ha estat informat de l'estudi i de tots els seus detalls. He contestat de manera satisfactòria totes les preguntes que el/la voluntari/a ha tingut.

Nom del investigador:______James Pownall_____________________ data_____________

Firma:

Declaració del/de la voluntari/a

He llegit tota la informació recollida en aquest document i estic d'acord a participar de manera voluntària en l'estudi descrit.

Nom del/de la voluntari/a ___________________________________ data __________

Firma:

Consentimiento para participar en la investigación

Nombre de proyecto: *Working memory, task complexity, and competition for cognitive resources during L2 oral task performance.*

Investigador: James Pownall Tel: 610.165.217 Email: JamesWilliamPG@blanquerna.url.edu

Patrocinador: Departamento de Filología Inglesa y Alemana. Universidad de Barcelona

---

Introducción

El estudio consiste en observar cómo influye la memoria del trabajo y la capacitad de atención en la manera con qué las personas utilizan el inglés para hacer una serie de actividades que se diferencian entre si en su nivel de dificultad.

Para empezar, se pedirá que hagan unas pruebas cognitivas y de nivel de conocimientos del inglés con el propósito de poder clasificar a los participantes según los resultados obtenidos.

La actividad principal consiste en poner en orden una serie de dibujos para que formen una historieta lógica para después narrarla mientras que son grabadas, quedando su voz registrada. Hay cinco historietas en total para grabar. Una vez grabadas las historietas, pasaran a ser transcritas para después poder analizar sus características lingüísticas. Cuando se hayan obtenido todos los resultados, se llevará a cabo un análisis estadístico.

Lo que se pide:

3. Completar una prueba de nivel de inglés, los resultados de los cuales servirán exclusivamente para fines estadísticos del estudio y no tendrán vigencia oficial en ningún otro caso.
4. Participar en una sesión de obtención de datos. La sesión tendrá una duración de 1,5 horas aproximadamente. La sesión incluirá:
   - Obtención de datos personales (5 minutos)
   - Una segunda prueba con ordenador de nivel de conocimientos del inglés (15 minutos)
   - Dos pruebas con ordenador de medidas cognitivas (45 minutos total)
   - Narración y grabación de las historietas con el investigador (15 minutos)

Beneficios

- Los/las voluntarios/as tendrán una oportunidad de utilizar el inglés en un entorno fuera del ámbito de una clase normal.
- Los/las voluntarios/as tendrán la oportunidad de participar en un estudio formal de lingüística cognitiva con implicaciones a nivel internacional en el área de investigación de adquisición de idiomas.

Confidencialidad

Toda la información recogida se mantendrá confidencial y se utilizará solo para fines de investigación. Su identidad se mantendrá anónima y ninguna otra persona, a parte del investigador, tendrá acceso a la información vinculada con los nombres de los voluntarios. En el caso que los resultados del estudio sean publicados, los nombres de los/las voluntarios/as no se utilizarán. Los datos se guardaran dentro de un fichero digital al cual tendrá acceso solo el investigador.

Participación

La participación en el estudio es completamente voluntaria. Si, en cualquier momento, cambia de opinión y decide no participar en el estudio, se tiene que comunicarlo al investigador y no continuará como participante en el estudio. Para cualquier consulta, contacte con el investigador por teléfono, correo electrónico.

Declaración del investigador

El/la voluntario/a ha estado informado del estudio y de todos sus detalles. He contestado de manera satisfactoria todas las preguntas que el/la voluntario/a ha tenido.

Nombre del investigador:___James Pownall_______________________ fecha ____________

Firma:

Declaración del/de la voluntario/a

He leído toda la información recogida en este documento y estoy de acuerdo a participar de manera voluntaria en el estudio descrito.

Nombre del/de la voluntario/a: ______________________________ fecha ____________

Firma:

Ref: ____________________

Personal data

Age: ____________

☐ Male ☐ Female Gender:

Nationality: ____

University studies (major): ____

Socio-linguistic information

Which language do you consider your native language?

☐ Catalan ☐ Spanish ☐ Other (Specify) ____

Do you speak other languages? ☐ Yes ☐ No

- If you answered 'yes' above, please specify. ____________________________________

If you have studied other languages apart from English, which is the language that have studied the most? ____

How many years did you study / have you studied it? ______________________

Do you still study it? ☐ Yes ☐ No

For how many years have you studied English? ______________________

How old were you when you began studying English? ______________________

What do you consider your level in English to be?

☐ Advanced ☐ Upper Intermediate ☐ Intermediate ☐ Pre-Intermediate ☐ beginner

Ref: ________________

Oxford University Press
and
University of Cambridge Local Examinations Syndicate

Name:…………………………………………………………………..……

Email:…………………………………………………………………..……

Quick Placement Test

Please complete the whole test (part 1 and part 2) or as much as possible within a period of time not greater than 30 minutes.

Time: 30 minutes

IMPORTANT: Els resultats de la prova es mantindran confidencials. No tindran cap implicació en les notes obtingudes a les classes de la universitat i s'utilitzaran només per a fins de recerca.

Declaro haver complert la prova sense cap mena de recursos, apart dels meus propis coneixements de la llengua anglesa per contestar a les preguntes de la prova, inclosos diccionaris de cap mena ni d'altres medis de referència.

Declaro també no haver estat més que 30 minuts per fer la prova, encara que no hagi pogut acabar de contestar totes les preguntes.

IMPORTANTE: Los resultados de la prueba se mantendrán confidenciales. No tendrán ninguna implicación en las notas obtenidas en las clases de la universidad y se utilizarán sólo para fines de investigación.

Declaro haber completado la prueba sin ningún tipo de recursos aparte de mis propios conocimientos de la lengua inglesa para contestar a las preguntas de la prueba, incluidos diccionarios de ningún tipo ni de otros medios de referencia.

Declaro también no haber empleado más de 30 minutos para hacer la prueba, aunque no la haya podido acabar de contestar todas las preguntas.

Firma ....

Fecha ....