View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repositório Institucional da UFSC

UNIVERSIDADE FEDERAL DE SANTA CATARINA

PÓS-GRADUAÇÃO EM LETRAS/INGLÊS E LITERATURA CORRESPONDENTE

PRE-TASK PLANNING, WORKING MEMORY CAPACITY, AND L2 SPEECH PERFORMANCE

MARIA DA GLÓRIA GUARÁ TAVARES

Tese submetida à Universidade Federal de Santa Catarina em cumprimento parcial dos requisitos para a obtenção do grau de

DOUTOR EM LETRAS

FLORIANÓPOLIS

Maio 2008

Esta tese de Maria da Glória Guará Tavares, intitulada Pre-task planning, Working Memory Capacity and L2 Speech Performance, foi julgada adequada e aprovada em sua forma final, pelo programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente da Universidade Federal de Santa Catarina, para fins de obtenção do grau de

DOUTOR EM LETRAS

Àrea de concentração: Inglês e Literatura Correspondente
Opção: Língua Inglesa e Lingüística Aplicada

Dr. José Luiz Meurer
Coordenador

BANCA EXAMINADORA:

Dra. Mailce Borges Mota
Orientadora e Presidente

Dr. Pedro Henrique Lima Praxedes Filho
Examinador

Dr. Ricardo Augusto de Souza
Examinador

Dr. Celso Henrique Soufen Tumolo
Examinador

Dra. Lêda Maria Braga Tomitch
Examinadora

Florianópolis, 16 de maio de 2008

To my mother and best friend, Aracy,
who made all efforts for me
to be who I am today.

To my husband and best mate, Kiko,
who supported me unconditionally
throughout these years.

With love and gratitude

ACKNOWLEDGEMENTS

It is with great pleasure that I thank all of those who have contributed in many different ways to the development of this doctoral study. The following special ones have helped me throughout my path and have made my walk more beautiful and smoother in so many incredible ways. Many thanks go to:

God, for all the blessings in my life, for giving me strength, wisdom, inspiration, and protection, for guiding me and for being always in control. Although sometimes I may forget, deep inside I know that everything comes from you. ‘‘***You’re every word, you’re every line, you’re everything’’.***

My mother, Aracy Maia Guará, a brave northeastern woman who has always taught me the value of study and work: The best schools were always the greatest priority for her two children. You have taught me that the most important thing is how far we want to go and how willing we are to go for it. You have inspired me to first believe in GOD and in myself, and then go ahead. Yes, there have been hard times, but above all you have taught me that times can never be too hard when there is endless love, faith, will, caring and sharing. You’re my best teacher, you’re a woman ahead of your time***, “you’re once, twice, three times a lady” in the fullest sense of the word.***

My husband, Kiko, a great northeastern man with the biggest heart, for all his support throughout my MA and PhD journeys. Thanks for putting up with my messy room, desk, mood, and mind. Thanks for cheering me up, for cooking delicious meals, for taking care of me whenever I got sick, for leaving your job to accompany me all the way to New Zealand, for telling me that ‘every little thing is gonna be all right’, for sharing love, dreams, and faith in our life and future. I am so blessed to have you by my side. Throughout our life together you have been ***‘‘the wind beneath my wings’’.***

My mother, brother, and my mother-in-law for the unconditional support, for the delicious meals whenever I went home to visit, and for the encouragement in my academic endeavors.

My grandmother Dolores (*in memorian*), who always used to pray for me when I had my tests in my school times, thanks for your frequent words when I used to leave home for school: “Seja feliz, vá com Deus e volte com Deus! Divino Espírito Santo! Deus por mim nada contra mim!”. Maybe at the time I was too young to understand *how precious* these words really are, I certainly understand them a lot better now.

CNPq for the financial support during my doctoral studies at Universidade Federal de Santa Catarina, and CAPES for the financial support which allowed me to carry out data analysis and attend courses at the University of Auckland, NZ, on a sandwich program.

My adviser, Mailce Borges Mota, for being a brilliant professor, researcher, and supervisor, for all the theoretical and methodological support, for all the relevant questions and insightful comments, for all the attention and careful reading of my work, for being so attentive to details, for introducing me to the fascinating, and challenging world of human cognition, for the *excellent* courses on SLA, working memory, and L2 speech production, and for accepting me as your advisee twice. It has been a fruitful professional partnership throughout these years, and I have learned a lot from you. Thanks, Mailce.

My adviser of the sandwich doctoral grant, Rod Ellis, for accepting me as a visiting scholar at the University of Auckland. Thanks for the *excellent* advising and insightful comments; it was an honor to share my results with such a bright mind. Thanks for the *excellent* course on TBLT, for asking me to present my work in one of your classes and in one doctoral seminar. Thanks for the office and computer space, for introducing me to the friendly DALSL staff. Thanks for helping me learn so much, for your generosity, sense of humor, and for having accepted to be our dinner guest. I had a great experience. Thanks, Rod.

My great best forever friends from my very homeland Fortaleza-Ceará: Illona, Marcia, Karine, and Júlia for all the emails of encouragement, for the prayers, for all our trips together to Pipa, Fernando de Noronha, Jericoacoara, Canoa, and so many other places. For the times we shared our lives, for the long walks by the wonderful beaches in Fortaleza, for all the times we laughed and cried together. Thanks for being wonderful friends, great sisters actually. Thanks for supporting me from the simplest projects (e.g., cute guys at Cais Bar during the 90's) to the most ambitious projects in my life (e.g., leaving my city and coming for academic life in Florianópilis). It's great to know that we are always together no matter how far apart we may be geographically. Real friendship is a real blessing.

My special friends Raquel D'Ely and Donesca Xhafaj for being such great examples of academic discipline, for helping me so much, for caring and sharing, for encouraging students to be my participants, for being so funny and so sweet, for cheering me up all the time, for believing in me, and for being always there when I needed. You are great friends and great researchers who have inspired me *a lot*. Thanks for sharing the journey with me and making my walk more beautiful and smoother in so many incredible ways.

All my professors at Universidade Federal de Santa Catarina, who helped me learn a lot: Mailce Borges Mota, Lêda Tomitch, Glória Gil, Maria Lúcia Vasconcelos, Apóstolo Nicolacopulous, Rosana Koerich, Barbara Batista, Viviane Herbele, and José Carlos Paes de Almeida Filho.

Professor José Luiz Meurer, for your support when I applied for the scholarship from CAPES. The experience abroad has made such a difference in my academic and personal life. I will never be able to thank you enough.

Professors Pedro Henrique Lima Praxedes Filho, Ricardo Augusto de Souza, Lêda Maria Braga Tomitch, and Celso Tumolo, for kindly accepting to be members of the examining committee. Thanks for the precious time and attention you dedicated to the reading of my study.

All my participants, for your cooperation and for dedicating your precious time to this study, even with no financial reward. I will never be able to thank you enough.

Bete from the language lab, for being so helpful in finding rooms for me to carry out data collection, and for lending us your home for a few days before we moved to Auckland. You were so helpful.

All my raters and friends Janaina, Kyria, Donesca, Denise, and Raquel for reanalyzing data with commitment and for making my life so much more pleasant during my doctoral studies.

Professors Rosana Koerich and Adriana Kuerten, for giving me permission to carry out my research with students from Cursos Extracuriculares and Letras.

João Carlos da Silva, for all your help in the office, for answering so many questions patiently and for lightning up our days at UFSC with your typical '*Manezinho*' smile and accent.

A special thanks (again!) to professor Lêda Tomitch for kindly opening the doors of your house and inviting us to wonderful parties… I've had so much fun! I always felt more energized to go back to my readings the days after the parties. Thanks also for being always present in my academic life, being part of my MA examining committee, PhD proposal examining committee, and now in my PhD examining committee.

My friends Vilmar de Souza and Renata Jorge for your friendship and support during the 'in between' moments right before I went to New Zealand and right after I got back. Thanks for your friendly support, for promptly picking us up at the airport, for kindly providing such a *delicious* meal, and for all the words of encouragement.

Professor Jenefer Philp at the University of Auckland, for the *excellent* course on individual differences in SLA and for your humanistic approach towards the students, for encouraging me to learn statistics, for sharing ideas, and for being so friendly and kind.

Sarina Ibnul, the research assistant at the Psychology Department at the University of Auckland, for your help with statistics when you barely knew me, and professor Rosane Silveira at Universidade Federal de Santa Catarina for the *excellent* course on quantitative research and for introducing me to the challenging world of statistics.

Carmen Schwager, for all your help on the *very first* day and on *all the remaining* days at the University of Auckland, for your kindness and friendliness.

My sweet friends Yiqian Cao (Katherine) and Emilia Petrocelli, for your support and friendship, for the good times together, for caring and sharing. I am so glad that I have met you. Each one of us comes from different backgrounds, cultures, and values, but the fact that we get along so well just proves that "*ebony and ivory can live together in perfect harmony.*"

My classmates and friends Carole Gamet, Toon, and Kazumi, for your friendliness and for making my life more enjoyable in Auckland.

My friends at the Railway Campus for the fun times in the kitchen and in karaoke shows, and for making my life so much more fun even when I had so much data to analyze: Sofie and Kiara, Lise and Andre, Pablo, Abdullah, Victor, Yuan, and Wilson.

My Friends from the Baptist Tabernacle on Queen St: Peter and Elvina, Tc, Roger, Helen, Pinaman, Marjorie, Shelly, Ken, Becky, Matilda, and many others. Thanks for all the prayers, for sharing the word of God, and for making my husband and me feel so welcome in NZ.

Professors Vladia Borges and Manolisa Vasconcellos, for having introduced me to the academic world at Universidade Federal do Ceará in 2000, for *all* the emails of encouragement, and for believing in me, sometimes even more than I do. You have really helped me move on, maybe you have no idea *how much*.

My friends at Ibeu Ceará, who have taught me so much and with whom I had great times of fun and friendship in the teachers' lounge having cakes, esfirras do Habib's, and Pardal ice cream bars. Thanks for all the chats at Picanha do Cowboy, D. Pastel, and so many other places.

My homeland Fortaleza-Ceará, the memories of which fill my heart and soul with joy, where I became the most of who I am today, where I learned so much about friendship, humor, enthusiasm, hard work, authenticity, and willingness. I definitely agree with Raquel de Queiroz when she states: "Se eu não fosse do Ceará, o que seria de mim? Não posso me imaginar ou fazer projeções ante esta sugestão insólita. O Ceará está muito ligado a mim para que eu possa me imaginar fora dele. Ou pior, sem ele."

ABSTRACT

PRE-TASK PLANNING, WORKING MEMORY CAPACITY,
AND L2 SPEECH PERFORMANCE

MARIA DA GLÓRIA GUARÁ TAVARES

UNIVERSIDADE FEDERAL DE SANTA CATARINA

2008

Advising Professor: Dr. Mailce Borges Mota

Research on task-based planning provides evidence of trade-off effects among the goals of fluency, accuracy, and complexity of L2 performance in the context of learners' limited attentional resources (e.g., Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003). However, there is a lack of empirical investigation on the role of working memory within the effects of planning on L2 performance. The present study investigates the relationship among pre-task planning, working memory capacity, and L2 speech performance. More specifically, it addresses the question whether individual differences in working memory capacity plays a role in performance under planning conditions and in the processes learners engage in when they plan. A population of 50 students from *Letras Licenciatura*, *Letras Secretariado*, and *Cursos Extracurriculares at Universidade Federal de Santa Catarina* was divided in two groups: control and experimental. Participants in the control group performed a working memory test, two narrative tasks under a no-planning condition, and a retrospective interview. Participants in the experimental group performed a working memory test, two narrative tasks (one under a no-planning and one under a planning condition), a retrospective online protocol, and a retrospective interview. L2 speech performance was assessed in terms of fluency, accuracy, and complexity. In general, results show that under a no-planning condition, working memory capacity significantly correlates with L2 speech accuracy (for the control group) and L2 speech fluency (for the experimental group). Under a planning condition, working memory capacity significantly correlates with L2 speech fluency and complexity. As for the impact of planning on performance, there was a significant effect on L2 speech accuracy and complexity, but not on fluency. Results also show that learners engage mainly in organization of ideas, rehearsal, lexical searches, and monitoring when they plan an oral task. Moreover, higher spans employ significantly more metacognitive strategies during planning when compared to lower spans. Results were discussed in terms of the working memory capacity model proposed by Engle, Kane, and Tuholski (1999), according to which individuals differ in the capacity for controlled attention in face of interference; and studies on task-based planning (e.g.,Ortega, 1999, 2005; Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003). Pedagogical implications were also pointed out suggesting that planning is as a task implementation condition that can be employed in the L2 classroom with the aim of drawing learners' attention to form.

Number of pages: 226
Number of words: 60.056

RESUMO

As pesquisas sobre o planejamento dentro da abordagem de ensino baseado em tarefas evidenciam efeitos de troca atencional entre os aspectos da fluência, acurácia e complexidade do desempenho em L2 devido à limitação dos recursos atencionais dos aprendizes (Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003). Entretanto, há uma carência de investigação empírica sobre o papel da memória de trabalho mediante os efeitos do planejamento no desempenho oral em L2. O presente estudo investiga a relação entre planejamento pré-tarefa, capacidade da memória de trabalho e desempenho oral em L2. Mais especificamente, ele investiga se diferenças individuais na capacidade de memória de trabalho afetam tanto o desempenho oral em condições de planejamento quanto os processos nos quais os aprendizes embarcam quando planejam uma tarefa oral. Uma população de 50 alunos dos cursos de Letras Licenciatura, Letras Secretariado e Cursos Extracurriculares foi dividida em dois grupos: controle e experimental. Participantes do grupo controle foram submetidos à coleta de dados que consistiu de: um teste de memória de trabalho, duas tarefas narrativas sob a condição de não planejamento e uma entrevista retrospectiva. Participantes do grupo experimental foram submetidos à coleta de dados que consistiu de um teste de memória de trabalho, duas tarefas narrativas (uma na condição de não planejamento e outra na condição de planejamento), um protocolo verbal e uma entrevista retrospectiva. O desempenho oral foi medido através da fluência, acurácia e complexidade. Em geral, os resultados mostram que na condição de não planejamento, há correlação significativa entre capacidade de memória de trabalho e fluência (para o grupo experimental), assim como também, entre capacidade de memória de trabalho e acurácia (para o grupo controle). Na condiçãode planejamento, há correlação significativa da capacidade de memória de trabalho com fluência e complexidade. Os resultados mostram também que o planejamento levou a diferenças significativas em acurácia e complexidade e que os alunos focam principalmente em organização de idéias, buscas lexicais, ensaio e monitoramento quando planejam uma tarefa oral em L2. Os resultados indicam ainda que o número de estratégias metacognitivas utilizadas durante o planejamento é significativamente maior para participantes com maior capacidade de memória de trabalho. Os resultados foram discutidos com base no modelo de memória de trabalho proposto por Engle, Kane e Tuholski (1999), segundo o qual, indivíduos se diferenciam em termos da capacidade de controle da atenção em condições de interferência; e também com base nos estudos sobre planejamento pré-tarefa (Ortega, 1999, 2005; Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003). Implicações pedagógicas também foram apontadas sugerindo que o planejamento é uma condição de implementação de tarefas a qual pode ser adotada em sala de aula com o objetivo de chamar a atenção dos aprendizes para o foco na forma.

Número de páginas: 226
Número de palavras : 60.056

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

(On CD)

CHAPTER I

INTRODUCTION

1.1 Background

Since I launched into my journey as an EFL[1] learner at Instituto Brasil Estados Unidos/Ceará, nearly 25 years ago, I have developed a fascination for the English language. As a student, I was always eager to speak during my classes because I used to view them as one of the few and best opportunities to speak the English language.

As time went by, such fascination led me to become an English teacher at IBEU Ceará, and as a teacher, it bothered me to see that not all students were as keen to speak in classes as I used to be when I was a student. As a consequence, I was always trying to find effective ways to promote speaking among my students. The speaking skill then became my fascination within the teaching of English. As time passed by, one question that started to follow me as a teacher was: Why do some learners speak so well whereas others have so many difficulties and can barely say a word in class? It was then that, besides the speaking skill, the fact that learners are all different started to attract my attention as well.

---

[1] Ellis (1994) distinguishes the terms second and foreign language. As for second language learning, "the language plays an institutional and social role in the community. In contrast, foreign language learning takes place in settings where the language plays no major role in the community and is primarily learnt only in the classroom" (p.11). Ellis (1994) also claims for the need of a neutral term, which in line with common usage, he uses the term second language. Therefore, from now on, following Ellis (1994), both second and foreign language will be referred to as L2 in the present study. Whenever necessary, the distinction between 'second' and foreign' learning contexts will be made.

When I started my academic journey at the graduate program in EFL Teaching Methodology at Universidade Federal do Ceará, in 2000, my goal was to investigate individual learning styles and speaking in the L2 classroom, which for a few reasons related to time constraints, was not attainable at the time. Thus, I ended up focusing only on learning styles. At the end of the EFL Teaching Methodology program, I was strongly advised to pursue a master's degree in which I could expand my research on learning styles.

When I began my journey as a master student at Universidade Federal de Santa Catarina, in 2002, I started to find some answers for questions which had followed me concerning individual differences and speaking. I came to realize by reading Levelt (1989) that speaking is a multifaceted cognitive skill, and by reading Ellis (1994), that a number of factors of learners' individual differences may affect L2 learning. However, I did not think it would be wise to just leave my initial project on learning styles and venture into investigating speaking at that time. I decided to postpone such a challenge for my PhD study. As a graduate student, I was pleased to be learning theories and improving my knowledge about L2 learning. I was hoping to find a line of research which would be relevant not only for theorizing about L2 learning but also for establishing a connection with L2 pedagogy. In this sense, I planned to carry out research on speaking and individual differences in my future PhD research.

It was during the enthralling course of "L2 Speech Production: Theoretical and Instructional Issues" taught by my advising professor Dr. Mailce Borges Mota Fortkamp, that I was given the assignment to present an article by Mehnert (1998): "The Effects of Different Lengths of Time for Planning on Second Language Performance." This article introduced me to the study of tasks, an area of inquiry that is promising for

theorizing about L2 learning in terms of information processing and that allows a link between L2 research and pedagogy (Ellis, 2005).

After some time reading, searching and talking to my adviser, who is a researcher on the working memory construct, I was able to define the line of inquiry for my PhD study. My adviser and I defined some questions to be pursued in a way to put together the study of tasks (through the construct of pre-task planning), the speaking skill, and individual differences in working memory capacity. We were successful in finding a line of inquiry for which I have had great fascination and that seems to be relevant for both L2 learning research and pedagogy. My hope is that the present study will not satisfy only my own interest. I hope this study will also contribute to existing research in the field of Task Based Language Learning and Teaching by shedding some light on the relationship between pre-task planning, working memory capacity and L2 speech performance[2].

1.1 Overview

Over the last decades, there has been a substantial body of research on tasks (Ellis, 2005). Within the study of tasks, one construct which has attracted considerable attention is planning[3]. According to Ortega (2005), planning seems to have evolved into an area of inquiry in its own right and "has become a burgeoning area of investigation within task-based learning" (p. 77).

---

[2] In the present study 'speaking', 'speech production' and 'speech performance' are operationalized as the ability to perform an oral narrative task (Fortkamp, 2000; D'Ely, 2006; Weissheimer, 2007).

[3] The terms 'strategic planning' and 'pre-task planning' will be used interchangeably in the present study to refer to planning which takes place before a task is performed (Ellis, 2005). The terms 'on-line planning' and 'within task planning' will be used to refer to planning that takes place during performance (Ellis, 2005). The term 'online-planning' will *also* be used to refer to planning as a cognitive process inherent to the act of speaking (Levelt, 1989). The term 'task-based planning' will be used to refer to the field of research on task planning be it pre-task planning (Skehan, 1996, 1998) or on-line (within task) planning (Ellis, 2005). These constructs will be dealt with in the Review of the Literature.

Researchers have investigated planning from a variety of perspectives, including the different types of planning (Foster & Skehan, 1996; Sangarun, 2005); different amounts of planning time (Mehnert, 1998); the interaction between planning and different task types (Foster & Skehan, 1996), and the interaction between planning and levels of proficiency (Kawauchi, 2005). Ortega (1999, 2005) states that most studies on planning take a *product*-oriented approach whose focus is on its *impact* on L2 performance. Thus, she claims for a more *process*-product oriented approach in the attempt to reveal where the benefits of planning come from. That is to say, she claims for a focus on the processes learners engage in when they plan, which help performance

In general, studies have shown a positive impact of planning on L2 performance. Several studies have shown that planning leads to gains in fluency[4] (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999). Planning also leads to gains in accuracy, although results have been more mixed in this respect (Ellis, 1987; Mehnert, 1998; Ortega, 1999; Foster & Skehan, 1999). Finally, studies have also shown that planning enhances complexity (Crookes, 1989; Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan & Ellis, 2003).

One interesting finding of the studies on the impact of planning on L2 performance is the evidence of attentional trade-off effects among the goals of fluency, accuracy, and complexity. Foster and Skehan (1996), Menhert (1998), as well as Yuan and Ellis (2003) discuss results of their studies in terms of an attentional model of learning and performance. In this sense, these researchers propose that there are trade-off effects among the goals of fluency, accuracy, and complexity in the context of the use of learners' limited capacity attentional resources. In other words, because

[4] According to Skehan (1996, 1998), fluency is related to the temporal aspects of speech production; accuracy is related to grammatical correctness; complexity is related to language elaboration (e.g., subordination).

attentional resources are limited, planning benefits can not be achieved to the same extent simultaneously for fluency, accuracy, and complexity of L2 performance. The trend of research results shows that there are gains in fluency and complexity at the expense of gains in accuracy.

I take the perspective that working memory resources are attentional. In the present study, working memory is defined as "a system consisting of those long-term memory traces above a threshold, the procedures and skills to achieve and maintain that activation, and limited-capacity, controlled attention" (Engle, Kane & Tuholski, 1999, p. 102). Despite the fact that researchers in task-based planning (e.g. Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003) explain results of studies in terms of learners' limited capacity attentional resources, individual differences in working memory capacity have not been taken into account in any of these studies (e.g. Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003) as a feasible variable for affecting learners' performance under planning conditions.

Although planning is a means of helping learners overcome limitations in working memory and improving performance (Ellis, 2005), I believe individual differences in working memory capacity may still emerge in L2 performance under planning conditions. Planning is a problem solving activity (D'Ely, 2006), and it seems to assist performance by triggering a range of strategic, metalinguistic, and metacognitive behaviors (Ortega, 2005). It seems reasonable to argue that one's ability to engage in such strategic behaviors successfully may to some extent explain benefits achieved from planning. Since individuals with higher capacity tend to be more strategic (McNamara & Scott, 2001; Mendonça, 2002; Weissheimer, 2007), individual differences in working memory capacity may reflect differences on how successful one is in the process of planning.

In addition to that, the benefits of planning on performance may also depend on the ability to actually retrieve what was planned and implement it into online performance (Ortega, 2005). According to Rosen and Engle (1997), working memory plays a crucial role in retrieval, that is to say, individuals with higher capacity tend to retrieve information more effectively during the performance of complex cognitive tasks.

Following these lines of reasoning, working memory capacity may play a role on how successfully one is in engaging in planning as well as on how effectively one may retrieve and implement what was planned into online performance. Thus, the present study sets out to examine how individual differences in working memory capacity may affect both L2 performance under planning conditions, and the processes learners engage in when planning their performance of an oral task.

1.3 Significance of the study

This study adds to existing research on task-based planning in three major ways. First, few studies so far have taken a *process*-product oriented approach in the attempt to scrutinize the processes learners engage in during planning. Second, only two studies – Ortega (2005) and Kawauchi (2005) – addressed how individual differences may affect the impact of planning on L2 performance and both of these studies have focused only on differences in terms of proficiency level. Third, and most importantly, investigations in the field of task-based planning explain trade-off effects among fluency, accuracy, and complexity in terms of learners' limited attentional resources (Foster & Skehan, 1996; Menhert, 1998; Yuan & Ellis, 2003). However, to the best of my knowledge, no studies to date have provided empirical evidence for the role of working memory capacity neither on L2 performance under planning conditions nor on the mental processes learners engage in when they plan.

In addition, according to Ellis (2005), the study of planning is relevant both for its importance for theorizing about L2 acquisition and for its usefulness to L2 pedagogy once it is a condition that can be implemented in language classrooms.

Individual differences in working memory capacity may be a fruitful window through which to look at pre-task planning for at least two main reasons. First, working memory as a limited cognitive system is one of the tenets of information processing theory (McLaughlin & Heredia, 1996), hence being a relevant construct for theorizing about L2 acquisition. Second, working memory may constitute a central component of language aptitude (Miyake & Friedman, 1998), as a predictor of reading comprehension (Daneman & Carpenter, 1980; Harington & Sawyer, 1992) and speech performance (Daneman, 1991; Daneman & Green, 1986; Fortkamp, 1999, 2003).

1.4 Organization of the doctoral dissertation

For the purpose of reporting on an empirical study which was carried out to investigate the relationship among pre-task planning, working memory capacity, and L2 speech performance, the remainder of this dissertation was organized in five chapters.

Chapter II lays the theoretical groundwork that will inform the study by reviewing relevant literature on L2 speech production, working memory, and pre-task planning. It reviews models of L1 and L2 speech production, models of working memory, empirical studies on the relationship between working memory and L2 speech performance, and empirical studies on the impact of planning on L2 performance.

Chapter III presents the method used for data collection and data analysis. This chapter also poses the research questions and hypotheses, as well as reports the pilot study which informed the present study.

Chapter IV presents the results of the study and Chapter V presents the discussion of these results by addressing the research questions and hypotheses in light of the literature reviewed in Chapter II.

Finally, Chapter VI presents some conclusions derived from the results of this study in light of existing literature in the field. Moreover, limitations of the study, suggestions for future research, and pedagogical implications are also pointed.

CHAPTER II

REVIEW OF THE LITERATURE

2.1 Introduction

The purpose of this chapter is to lay the theoretical groundwork that will inform the present study, whose focus is on the relationship among pre-task planning, working memory capacity, and L2 speech performance. The chapter focuses on the following issues: (a) L1 and L2 speech production models, (b) the conceptualization of the working memory construct, and models of working memory, (c) empirical studies on the relationship between working memory capacity and speech production, (d) planning as a metacognitive process, (e) empirical studies on the impact of planning on L2 speech production, (f) the strategy framework proposed by O' Maley and Chamot (1990), and (g) verbal protocols. Finally, I point out a gap in the research on planning and L2 speech performance in order to nestle the current study.

2.2 L1 and L2 Speech Production Models

As regards L1 speech production, I will focus on Levelt (1989) who developed one of the most influential models of L1 speech production by mature speakers. According to Levelt (1989) "the dissection of the speaking skill is a scientific endeavor in its own right" (p. 1). Although the scholar acknowledges the role of spoken interaction to the understanding of speakers as interlocutors, he claims that, in order to understand the act of speaking as a complex cognitive skill, one must scrutinize its systems and subsystems.

Levelt (1989) proposes an information process blueprint to explain the processes of L1 speech production. Thus, the model is depicted within the main tenets of the information processing approach to human cognition: (a) complex behavior builds on simpler processes, (b) the processes are autonomous (c) processes take time and predictions about time reaction can be made, (d) the mind is a limited-capacity processor, thus, being able to attend only so much to the various components of a complex task, (e) the constructs of automaticity and control permeate the functioning of the whole process, in which some tasks require more attention and others require less attention (McLaughlin & Heredia, 1996).

Levelt's (1989) model has four specialized components, which underlie the speech production process: the concpetualizer, the formulator, the articulator, and the speech comprehension system. These components work in a highly automatic way, and automaticiy is what allows them to work in parallel, which is, in turn, "a main condition for the production of uninterrupted speech" (p. 2).

In the conceptualizer, the message content is planned by retrieving background knowledge, knowledge about the topic as well as knowledge of discourse patterns. In order to generate a message, macroplanning as well as microplanning take place. In macroplanning, the speaker retrieves information to convey his/her communicative intention, the content of the message; in microplanning, on the other hand, the speaker plans the form of the message, which encompasses fixing the appropriate speech act, marking the status of referents as 'given' or 'new', and assigning topic and focus. The processes of macroplanning and mircoplanning result in the preverbal message that is the input for the formulator.

In the formulator, the preverbal message turns into a linguistic structure through two processes: grammatical encoding and phonological encoding. Grammatical

encoding encompasses procedures for accessing lexical units and applying syntactic rules. According to Levelt (1989), a lexical unit consists of a lemma and a lexeme. The former encompasses the semantic and syntactic properties of the lexical unit. The latter encompasses the morphological and phonological properties of the lexical unit. The lemma will be activated by means of matches between its meaning and part of the preverbal message, which will, in turn, lead to syntactic availability. When all lemmas have been accessed and syntactic building has been accomplished, a surface structure of the message is generated.

Phonological encoding, which takes place through the morphological and phonological information in the lexeme, functions in order to build a phonetic or articulatory plan for the lemmas. The product of phonological encoding is, thus, a phonetic or articulatory plan, which Levelt (1989) refers to as "an internal representation of how the utterance should be articulated" (p. 12). The product of the formulator will be the input of the articulator. In the articulator, the phonetic plan is executed and results in overt speech. Finally, the speech comprehension system makes overt and internal speech available for monitoring.

According to Levelt (1989), speech production is lexically driven, that is, knowing words is the paramount condition for expressing communicative intentions. Thus, Bock and Levelt (1994) as well as Bock (1995) acknowledge grammatical encoding as the heart of the speech production system since it serves as a bridge from message (meaning) to phonological encoding (sound).

As previously stated, grammatical encoding involves the selection of lexical items and syntactic building. In the attempt to go further in specifying the complexities of this subcomponent of the formulator, which represents the heart of the system, Bock and Levelt (1994) propose that grammatical encoding encompasses two sets of

processes: functional processing and positional processing. The primary subcomponents of functional processing are (a) lexical selection, which involves the identification of lexical concepts that are suitable for conveying the speaker's meaning and (b) function assignment, which involves the assignment of grammatical rules or syntactic functions.

The primary subcomponents of Positional processing are (a) constituent assembly, which involves the creation of a control hierarchy for phrasal constituents that manages the order of word production and captures dependencies among syntactic functions, and (b) inflection, which involves the generation of fine-grained details at the lowest level of this structure, such as information about number, tense, and aspect that are bound to words, for instance -Work-ing, Work-s, Work-ed.

Although Levelt's (1989) is basically a stage model, the input a component receives is the output of the previous component, the model also allows for formulation and articulation to run in parallel, since the processor can start working on the still-incomplete output of the previous processor. This combination of serial and parallel processing is what Levelt (1989) calls incremental processing. The paramount condition for incremental processing to occur is automaticity once it allows the processing components to work in parallel and contribute to fluent speech production.

Having brought the complexities of L1 speech production depicted by Levelt (1989) into the present scenario, I turn now to the discussion of three major L2[5] speech production models- Green's (1986), De Bot's (1992), as well as Poulisse and Bongaert's (1994). It is important to highlight that these models of L2 speech production will not

[5] According to Selinker (1969, as cited in Praxedes Filho, 2007), L2 is a fully developed system and interferences from L1 to the L2 are theoretically impossible; interferences take place from L1 to interlanguage (IL), which is a separate subsystem distinct from learner's L1 and from the L2. Although, the concept of interlanguage (learner language) as proposed by Selinker, is accepted in the present study, the terms 'L2 speech production', 'L1 transfer to L2', 'L1 traces in L2' will be used instead of 'IL speech production', 'L1 transfer to IL', and 'L1 traces in IL'. This choice was made due to the fact that L2 is the term commonly used in speech production models and task-based research.

be as extensively referred to as Levelt's (1989) model during the discussion of results. However, because there are few models of L2 speech production available in the literature, these three models will be reviewed here in order to acknowledge the efforts of these authors in proposing their models as well as to provide references on L2 speech production models for future graduate students.

Green (1986) proposes a framework for explaining the way in which normal and brain-damaged bilinguals control the use of their two languages. Based on the fact that brain-damaged bilinguals may lose command of one language but not the other, Green (1986) proposes that languages are separate in different subsystems, which can be activated to different degrees. Green distinguishes three different levels of language activation. Selected languages have the highest level of activation and are the ones being currently spoken. Thus, selected languages control speech output. Active languages are less activated; they are regularly spoken, but are not selected for being spoken. It is important to highlight that active languages may affect in ongoing process and cause interference effects. Lastly, dormant languages are the least active ones and can not interfere in ongoing process.

Green (1986) explains his framework around the ideas of control, activation, and resources. Failure to exercise full control over intact language systems can explain impaired performance in brain-damaged bilinguals as well as speech errors in normal bilinguals. Assuming that speech production can be viewed as a skilled action in general, Green (1986) explains that control is exercised by the amount of activation. That is, in order to select a word, for instance, one has to assure that this particular word's level of activation exceeds the activation of its competitors. Any act of control, in turn, consumes resources (energy). Since the resources necessary to regulate the activation and inhibition of languages are limited, control may not be fully exercised

and speech errors may occur. The ideas of control, activation, limited resources, and inhibition in Green's model imply the role of working memory in L2 speech production.

According to Poulisse (1997), one advantage of Green's model is that it explains why beginner L2 learners present L1 interference more frequently than advanced ones. With the ideas of control, activation, and limited resources in mind, it may be assumed that because beginners' L2 production has not been automatized they need to devote much energy when speaking their L2. As a consequence of such energy effort, they may have fewer resources left available for the suppression of their L1, thus leading to interferences between the two languages.

Although Green's (1986) model may be helpful, it does not go without its drawbacks since it does not offer detailed accounts of message generation, formulation, and articulation. De Bot's (1992) model is much more comprehensive in this sense. According to De Bot (1992), there are several reasons to propose a model of L2 speech production based on Levelt's (1989). The model is based on decades of psycholinguistic research and, thus, has gathered much empirical data. Therefore, De Bot (1992) made only few necessary adaptations.

The first assumption of De Bot's (1992) model is that the speaker has, first of all, to decide what language to speak. This decision takes place in the conceptualizer, more specifically, during macroplanning. Then, the speaker undergoes microplanning, which is language-specific in nature. The speaker uses information to convey the message in the language at play in order to bring the language appropriate lexical units in the formulator.

As far as the formulator is concerned, De Bot (1992) proposes that it is language-specific; thus, different procedures are applied to the grammatical encoding of L1 and L2 speech. Following Green (1986), De Bot (1992) suggested that bilinguals

produce two speech plans simultaneously, one for the language spoken at the moment (selected language) and one for the language not being spoken at a given moment, but in regular use (active language). Because bilinguals have two speech plans available, it makes it possible to stop the encoding of one of them and start the other (code switching).

With regard to De Bot's (1992) proposal for the organization of the mental lexicon, he assumes that this is language independent, that is, there is only a single lexicon, which is divided into different subsets that undergo activation to different extents, according to the language being spoken.

Finally, as regards the articulator, De Bot (1992) suggests only one for both languages. The formulator is assumed to store a large amount of sounds and pitch patterns from both L1 and L2. Since De Bot (1992) assumes only one articulator in which sounds and pitch patterns of both languages are stored together, L1 interferences in L2 can be explained.

Poulisse (1997) raises her voice in claiming that, although useful, De Bot' s (1992) model seems problematic mainly as regards language choice and code switching. She argues that if language choice is made in the conceptualizer so as to raise activation of the language being spoken, it seems fuzzy how both speech plans can be formulated in parallel. Although it is possible that the other language (the suppressed one) is still activated as a result of previous use, it is not clear how the speaker is able to keep both languages (the selected and the active one) apart. In addition, Poulisse (1997) claims that De Bot's (1992) model is uneconomical since more than one speech plan can be overtly produced. Rather than having to cope with two speech plans, attentional resources could be allocated directly at the speech plan of the selected language.

Poulisse and Bongaerts (1994) propose an account of L2 speech production that is also based on Levelt (1989). Similarly to De Bot (1992), they propose that bilingual speakers are able to manage separation or mix of different languages if they intend to do so and that speakers' language choice also takes place in the conceptualizer.

For Poulisse and Bongaerts (1994), L1 and L2 are stored in one single network. Thus, words must contain information that specifies which language they belong to. Following Green's view, Poulisse and Bongaerts (1994) postulate that "lemmas are tagged with a language label" (p. 216). They claim that lexical selection takes place through spreading activation. The lemmas receiving the most activation are the ones selected by the bilingual speakers.

According to Poulisse (1997), since lemmas are tagged for languages and lexical selection takes place through spreading activation, there is no need to have speech plans for L1 and L2 concurring simultaneously. Therefore, Poulisse and Bongaerts' (1994) model is more economical.

In brief, despite the fact that De Bot's (1992) model is uneconomical, he gives an elegant account of L2 speech production processes and is able to explain L1 phonological interference in L2. Poulisse and Bongaerts (1994) propose a model of L2 speech production that is in line with De Bot's model as regards the proposal that language choice takes place in the conceptualizer. Nevertheless, following Green (1986) they explain lexical access in terms of spreading activation.

The models of L1 and L2 speech production reviewed above provide insights concerning the complexity of L1 and L2 speech production processes. While L1 speech production is highly automatized, Poulisse (1997) postulates that: (a) L2 knowledge is not complete, (b) L2 is more hesitant, has shorter sentences and slips of the tongue, (c) L2 may carry traces of L1 and (d) proficient speakers can keep one or more languages

apart when they wish to do so. Thus, the high degree of automatization in L1 does not apply to L2. For this reason, in many circumstances, L2 learners may need to creatively construct plans for communicative situations since ready-made chunks may not be available, and this activation of procedures demands high degrees of cognitive control (Mehnert, 1998). These control processes take place under a limited capacity cognitive system, working memory. This construct will be the focus of the next section.

2.3. From short-term memory to working memory

Before focusing on the construct of working memory itself, I shall firstly focus on memory. What is memory? What do we need memory for? Memory is the system which brings sense and meaning to our existence (Ashcraft, 1994); it is the constant connection of our experiences (Baddeley, 1990). We need memory to make sense of the world around us, of who we are, of what we do. According to Ashcraft (1994), "any past event that is currently recalled is evidence for memory" (p.11). Therefore, most of what we experience in the present relies strongly on memory (Weissheimer, 2007).

Ashcraft (1994) defines memory as "the mental processes of acquiring and retaining information for later retrieval, and the mental storage system that enables these processes" (p.11). Hence, memory encompasses a system and the processes within such a system. Baddeley (1990) states that, even though philosophers have reflected upon memory for about two thousand years, the systematic study of memory is considered to be new since the relevant body of work has only started nearly a hundred years ago.

Ebbingghaus is referred to as the first one to propose a reasonably scientific method to the study of memory whereas William James was the first one to propose that memory consists of two parts: an immediate available one and a larger one that keeps

past experiences (Ashcraft, 1994), a proposal that is fully accepted today. Although William James was the first one to propose the division of memory into two parts, it was with the work of Atkinson and Shiffrin (1968, as cited in Atkinson & Shiffrin, 1971) that the division of memory into long-term and short-term memory was fully acknowledged.

According to Ashcraft (1994) and Miyake and Shah (1999), most current views of working memory have originally developed from the traditional concept of short-term memory. Initially, short-term memory was conceived as a passive unitary system which could temporarily hold information for retrieval after a brief period of time (Miller, 1956; as cited in Tomitch, 1996; Norman & Waugh, 1965, as cited in Tomitch, 1996).

Atkinson and Shiffrin (1968, as cited in Atkinson & Shiffrin, 1971) proposed a general model of memory which included a sensory buffer, a short-term storage and a long-term storage. According to Atkinson and Shiffrin's three stage model, information would first enter a range of sensory buffers simultaneously. Then, information would enter a short-term store where some control functions such as rehearsal would take place. Rehearsal would then enable information to be transferred to the long-term store.

According to Atkinson and Shiffrin's model, the longer an item was kept in short-term memory for rehearsal the more likely it would be transferred to long-term store. Atkinson and Shiffrin's assumption that short-term store was the path to long-term store, however, could not be supported due to evidence showing that patients wih damaged short-term memory could still have intact long-term store. Moreover, a few studies (e.g., Craik & Watkins, 1973, as cited in Fortkamp, 2000) suggested that rehearsal did not necessarily lead to long-term retention.

Baddeley and Hitch (1974) challenged Atksinson & Shiffrin's (1968, as cited in Atksinson & Shiffrin, 1971) view of short-term memory as a passive unitary system responsible for storing information and set out to examine whether there was a unitary short-term memory system or separate subsystems. They used a dual-task technique which employed concurrent tasks. Participants would have to remember a digit sequence of six items while performing a cognitively demanding task such as reading. If there were a unitary limited-capacity system responsible for all cognitive performance, performing the simple digit task would load the unitary system, and thus, participants would not be able to perform a cognitively demanding task such as reading concurrently. However, if there were different subsystems for simple and complex tasks, memorizing a digit load would not impair reading comprehension. They found that the digit load task caused interferences but not enough to entirely impair the performance of complex tasks.

These results along with the evidence that short-term memory damaged patients still showed intact long-term memory systems led Baddeley and Hitch (1974) to propose a multicomponent model of short-term memory which they termed working memory, as it is conceived until the present. Working memory is presently conceived as the limited-capacity human cognitive system responsible for simultaneous temporary storage and processing of information in the performance of complex cognitive tasks (Baddeley, 1990; Daneman, 1991; Daneman & Carpenter, 1980; 1983; Engle, 1996; Fortkamp, 1999; 2000).

Working memory is a dynamic system, and its limitations lie in the resources available for storing and processing information simultaneously; whereas short-term memory is a fixed set of slots that store information, and its limitations lie in the number of items it can retain while computing a mental activity (Ashcraft, 1994; Tomitch, 1996;

Torres, 2003). This view of working memory as an active system responsible for simultaneous storage and processing of information as opposed to the traditional concept of short-term memory as a passive buffer is widely accepted in the current days.

Nevertheless, there are still controversies in the field as regards the nature, structure, and function of working memory despite the whole body of research on this construct (Baddeley, 1990, 1999, 2000; Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Cowan, 1999; Engle, Kane, & Tuholsky, 1999, among many others). As a consequence of these controversies, the construct of working memory is, according to Miyake and Shah (1999), "one of the hottest topics in cognitive psychology and cognitive neuroscience" (p. xii), and models of working memory abound in the literature. They will be the focus of the next section.

2.3.1 Models of working memory

As previously said, there are several models of working memory available in the literature (see Miyake & Shah, 1999, for a review). However, only two working memory models will be reviewed in the present study. First, I will review the seminal model proposed by Baddeley and Hitch (1974) for its historical importance to the study of working memory. Next, I will review the model proposed by Engle, Kane, and Tuholski (1999), which is the model chosen for the present study since I take the perspective that working memory resources are attentional.

According to Fortkamp (2000), "the connotation with which the phrase 'working memory' is used nowadays was first introduced by the model proposed by Baddeley and Hitch (1974)". Thus, I agree with Fortkamp (2000) when she states that "any

discussion on the concept of working memory must start with Baddeley and Hitch's (1974) model" (p. 16).

According to Baddeley and Logie (1999), the original model proposed by Baddeley and Hitch (1974) consists of a central executive, which is a supervisory system, and two specialized slave systems, the phonological loop and the visual spatial sketchpad. According to Baddeley and Hitch (1974), the central executive coordinates the slave systems, controls attention, and activates information from long-term memory. The phonological loop is responsible for storing and manipulating speech-based information, and the visual spatial sketchpad controls visual and/or spatial material.

As regards the central executive, Baddeley (1990) postulates that this component parallels the supervisory attentional system proposed by Norman and Shallice (1986, as cited in Baddeley, 1990). The supervisory attentional system is responsible for the control of actions. Norman and Shallice (1986, as cited in Baddeley, 1990) propose that well-learned actions are triggered automatically through schema activation whereas actions involving novelty require attentional control by the supervisory attentional system in order to inhibit reflex like behavior. In the original model, besides attributing the coordination of the slave systems, control of attention and activation of information from long-term memory, Baddeley and Hitch (1974) also attributed storage functions to the central executive.

Later on, Baddeley and Logie (1999) propose that any increase in storage beyond that of the two slave systems can only take place by accessing long-term memory or other subsystems. More recently, Baddeley (2000) propose a fourth component to the model, the episodic buffer, which is responsible for integrating information from the two slave systems and from long-term memory. Baddeley and Logie (1999) view working memory and long-term memory as comprising two

functionally separate cognitive systems and, according to them, the view of working memory as a an activated portion of working memory is an uninformative oversimplification. Baddeley and Logie (1999) attribute to working memory the roles of retrieval of stored long-term knowledge relevant for the task being performed, manipulation and recombination of material, which allows the interpretation of novel stimuli. Moreover, they attribute to working memory the role of encoding into long-term memory the results of its operations.

According to Baddeley and Logie (1999), each component of working memory has different limitations according to the specialist function that each one of the components holds. However, they assume that, in each component, activation is a source of limitation and that both amount and duration of activation are limited. Working memory limitations may stem from capacity for activation or capacity for rehearsal, or from capacity for the complexity of material, or from the extent to which components are supported by acquired strategies and/or prior knowledge. While Baddeley and Hitch (1974) and Baddeley and Logie (1999) propose a multi-component model of working memory and focus on describing the different components of working memory, Engle et al. (1999) propose a unitary model of working memory and focus on investigating individual differences in the mechanism of controlled attention.

Engle et al. (1999) view working memory as a cognitive system comprising (a) a store in the form of long-term memory traces active above a threshold, (b) processes for achieving and maintaining this activation, and (c) controlled attention. Nevertheless, when they refer to 'working memory capacity', it is the limited capacity of the element of *controlled attention* that is being referred to. More specifically, for Engle and his associates the term *working memory capacity* refers to "attentional processes that maintain task-relevant information activated in an accessible state, or to retrieve that

information under conditions of interference, conflict, and competition" (Kane, Conway, Hambrick, & Engle, 2003, p.23). When referring to working memory capacity, Engle and his associates mean the limited capacity of the mechanism that Baddeley and Hitch (1974) as well as Baddeley and Logie (1999) call central executive, which is, in turn, similar to the supervisory attentional system proposed by Norman and Shallice (1986, as cited in Baddeley, 1990; Engle et al., 1999).

According to Engle et al. (1999, p. 104), "working memory is not about storage or memory per se, but about *the capacity for controlled sustained attention in the face of interference*" (emphasis in the original). They view the nature of working memory limitations in terms of the capacity for controlled attention, which will emerge in situations that require controlled processing. In a controlled processing activity, it is attention that is controlled, and the cognitive mechanisms that encompass a controlled processing activity include activation, suppression, serial search and retrieval, and monitoring (Engle, 1996; Engle & Oransky, 1999; Engle, et al.). In other words, individual differences in working memory capacity reflect differences in the capacity for sustaining, maintaining, and shifting attention among the various aspects of task performance (e.g. activation, suppression, monitoring), which also leads to differences in the ability to maintain and to inhibit activation of irrelevant information.

Although Engle et al. (1999) recognize that people may also differ in knowledge and in the skills for manipulating knowledge, the bulk of the research conducted by Engle and his colleagues has focused on the element of controlled attention (Conway & Engle, 2005; Engle & Oransky, 1999; Rosen & Engle, 1997; Unsworth & Engle, 2007, just to mention a few), and these researchers have consistently provided evidence that individual differences in working memory capacity reflect differences in humans' general ability to control attention.

While Baddeley and his associates have approached research on working memory by focusing on its different components, mainly the slave systems, and by establishing the biological implementation of these components (Baddeley, 2000; Baddeley, Gathercole, & Pagano, 1998;), Engle et al. (1999) state they have addressed two main issues in working memory research. First, they have sought to establish the processes that are tapped by the working memory tests that are also tapped by the higher-order cognitive tasks. For instance, in studies investigating working memory capacity as a potential source of individual differences in reading comprehension, it is important to establish what processes account for the correlations between reading performance and working memory performance. In other words, it is important to establish what is tapped by reading comprehension that is also tapped by the memory test. Second, they have sought to establish what results of the studies on individual differences have informed about the general nature of working memory. Within this aim, Engle and his associates have focused on issues concerning the generality of working memory (Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004; Turner & Engle, 1989), and the relationship of working memory capacity to short-term memory and general fluid intelligence (Conway, Kane & Engle, 2003; Kane & Engle, 2002; Unsworth & Engle, 2006).

Clearly, Baddeley and Engle have pursued different lines of research on working memory. Baddeley and his associates have focused on describing the components of working memory whereas Engle and his associates have focused on individual differences. Their models diverge in some aspects of working memory. Baddeley and Logie (1999) do not view working memory as an activated portion of long-term memory, whereas Engle et al. (1999) state that working memory consists of long-term memory traces activated above a threshold. Moreover, for Baddeley and Logie (1999),

working memory limitations may stem from capacity for activation or capacity for rehearsal, or from capacity for the complexity of material, or from the extent to which components are supported by acquired strategies and/or prior knowledge, whereas for Engle et al. working memory limitations stem from the capacity to control attention in face of interference.

Earlier, Cantor and Engle (1993) argued that individual differences in working memory capacity reflected differences in overall activation limits. However, later, Conway and Engle (1994) concluded that the view of working memory as reflecting differences in overall activation limits could not be sustained since the cognitive tasks used in Cantor and Engle (1993) involved both automatic retrieval and effortful retrieval. Thus, results of the former study by Cantor and Engle (1993) failed to provide specific accounts for individual differences in working memory capacity as being reflected in levels of activation limits. Conway and Engle (1994) concluded that higher and lower spans did not differ in terms of automatic retrieval. Rather, individual differences emerged when retrieval took place under conditions of interference, conflict, distraction, and competition.

Despite the aforementioned differences, the model proposed by Baddeley and Logie (1999) also bears similarities with the one proposed by Engle at al. (1999). First, Baddeley and Logie (1999) propose that the central executive is responsible for attentional control, which is compatible with the proposal of a limited capacity mechanism responsible for controlled attention advocated by Engle et al. Second, Baddeley and Logie (1999) agree with Engle et al. in the sense that they do not attribute storage functions to the central executive. These functions are attributed to the slave systems, according to Baddeley and Logie (1999), and to short-term memory, according to Engle et al.

After reviewing these two models of working memory, it seems now high time to establish and justify the perspectives I take on working memory for the purposes of the present study. This will be the target of the next section.

2.3.2 The perspectives on working memory for the purposes of the present study

As briefly stated in the Introduction, I take the perspective that working memory resources are attentional, that is to say, I take Engle's (1999) attention-view perspective of working memory capacity. Following Fortkamp (2000), it seems reasonable to characterize working memory resources as attentional for several reasons. First, it is compatible with Baddeley and Hitch's (1974) conceptualization of the central executive as responsible for controlling attention: "Approached from an attentional viewpoint, working memory (the central executive) could be named *working attention"* (Baddeley & Logie, 1999, cited in Fortkamp, 2000, p. 165). Second, it is also compatible with the notion of attention as a limited mental energy (Green, 1986; Just & Carpenter, 1992). Finally, Fortkamp (2000) also postulates that the view of working memory as a limited mental resource is accepted in the field of L2 acquisition by VanPatten (1990; 1996) and Skehan (1996; 1998).

In addition to Fortkamp's (2000) reasons, I should as well highlight that limitations in attentional resources are also frequently brought into play in the field of task-based planning research in order to account for trade-off effects in L2 performance (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003). In my viewpoint, the attention-view perspective of working memory capacity as the ability to sustain, control, and switch attention in face of interference, conflict, and competition (Engle et al., 1999) is compatible with Skehan's (1998) proposal that fluency, accuracy, and complexity compete for learners' limited attentional resources.

Based on what has been said, when I refer to individual differences in the present study, I mean that learners may differ in the capacity for sustaining, maintaining, and shifting attention, and in the ability to maintain and to inhibit activation in L2 speech performance under planning and/or no planning conditions, *and* in the processes learners engage in when they plan performance of an oral task.

So far this section has focused on the construct of working memory by reviewing its background, theoretical models, and also on the perspective I take towards working memory for the purposes of the present study. Now I turn to empirical studies on working memory and L2 performance.

2.3.3 Working memory and L2 performance

Although there is evidence which suggests that there is a relationship between working memory and L2 acquisition (Daneman & Case, 1981; Ellis & Sinclair, 1996; Mendonça, 2003; Miyake & Friedman, 1998; Mackey, Philp, Egi, Fugii, & Tatsumi, 2002, just to mention a few), I will focus on studies concerning the relationship between working memory and L2 performance since this is the focus of the present study.

According to Baddeley (1992), research on working memory has developed from two approaches, namely, the dual-task neuropsychological approach and the psychometrical correlational approach. The dual-task approach is concerned with explaining the structure of the three-component model of working memory proposed by Baddeley and Hitch (1974, as cited in Baddeley, 1990; Baddeley & Hitch, 1994). The psychometric correlational approach postulates that individual differences in the performance of complex cognitive tasks may reflect differences in working memory capacity.

A test was developed by Daneman and Carpenter (1980) to tap both storage and processing functions of working memory, the Reading Span Test. This test encompasses the two components - storage and processing - by joining the demands of sentence comprehension and the storage and retrieval of final words of sentences. The Reading Span Test was the first valid measure of working memory capacity (Engle, 1996), and, according to Fortkamp (1999), it has been the basis of most of the research on individual differences in working memory capacity and reading comprehension.

Although there is a bulk of research providing evidence for the relationship between working memory and language comprehension (Daneman & Carpenter, 1980, 1983; Tomitch, 1996, 2000, 2003; Torres, 2003, Whitney, Ritchie & Clark, 1991; just to mention a few), a mounting body of research has also gathered evidence for the role of working memory on language production (Bergsleithner, 2007;Daneman, 1991; Daneman & Green, 1986; Fontanini, Weissheimer, Bergsleithner, Perucci & D'Ely, 2005; Fortkamp, 1999, 2003; Guará-Tavares, 2006; Mizera, 2006; Weissheimer, 2006; 2007; Xhafaj, 2006).

Daneman and Green (1986) developed the Speaking Span Test (hereafter SST) in order to investigate whether working memory capacity would be a good predictor of learners' ability to use textual context to both comprehend and produce words in their L1. They tested whether there was a relationship between working memory capacity and the ability to produce synonyms for words presented in context. They found a correlation between working memory capacity scores and the synonym lexical test scores.

Later, Daneman (1991) investigated whether working memory capacity could account for individual differences in verbal L1 fluency. However, Daneman (1991) focused on fluency at a more comprehensive level. She used a speech generation task

(picture description) and investigated whether there was a relationship between WM capacity scores and speech rate. In Daneman's study, a significant correlation between the scores of the SST and speech rate in L1 was found.

Fortkamp (1999) expanded Daneman's (1991) study in order to investigate whether working memory would be a good predictor of *L2* verbal fluency. Fortkamp (1999) also found significant correlations between WM capacity as measured by the SST and L2 speech rate in the speech generation task. Fortkamp (2003) went further in her investigation on the relationship between WM capacity and L2 speech production and expanded the measures used to assess speech production in her 1999 study. She investigated whether WM capacity would predict individual differences in L2 fluency, accuracy, complexity, and weighted lexical density[6]. Results indicated that individual variation in the amount of attentional resources is related to variation in L2 speech performance. In this sense, results revealed that individuals with higher working memory capacity, as measured by the SST, tend to be more fluent, accurate, and complex in L2. Interestingly, the study provided evidence of trade-off effects since accuracy, fluency, and complexity of speech tended to be achieved at the expense of weighted lexical density.

Fontanini et al. (2005) report on a study which investigated the relationship between working memory and L2 performance in several domains being L2 speech performance (fluency and accuracy) one of them. Working memory capacity was assessed by the same measure of Fortkamp (1999, 2000), namely, the SST. However, participants in Fontanini et al. were beginners and the SST was adapted to this level of proficiency. Surprisingly Fontanini et al. did *not* find a significant correlation between

---

[6] Lexical density refers to the proportion of new and repeated words in a text (O'Loughlin, 1995). Weighted lexical density is a measure which provides a relationship between the number of words produced with lexical properties and the number of words produced with grammatical properties (O'Loughlin, 1995).

the measures of working memory capacity and fluency. A significant correlation was found between the measures of working memory capacity and accuracy. In the attempt to explain results as for the lack of correlation between WM capacity and fluency, Fontanini et al. claim that because participants were all beginners in this study, perhaps speech rate was not the most sensitive measure to assess fluency in this case. Due to their lack of proficiency in the L2, speakers may have devoted a great deal of attention to lexical searches and grammatical mappings, thus limiting fluency. In addition they raised the possibility of a methodological flaw.

Mizera (2006) also investigated whether WM capacity plays a role in L2 oral fluency. In his study, three measures of fluency were used: (a) speed of delivery in a monologic narrative task, (b) scores in a word translation task, and (c) scores in an imitation grammaticality task. Likewise, three measures of working memory capacity were employed: (a) Speaking span test, (b) a Math Span Test, and (c) a Non-word Repetition Test. The hypothesized strong correlation between WM capacity and fluency was not supported. According to Mizera (2006), the complexities involved in L2 speech performance may involve factors other than WM capacity. Thus, he claims that personal and affective factors may also play a role in fluent L2 speech.

It is important to highlight, however, that the only significant correlations found in Mizera's study were between the SST scores and fluency scores as measured by speed of delivery, and between the SST scores and the Imitation grammaticality task scores. There were no significant correlations between the other two WM capacity tests (Math Span Test and Non-word Repetition Task) and speech rate.

Moreover, the Imitation Grammaticality Task used by Mizera (2006) actually involved an element of grammatical accuracy since participants were supposed to imitate and also correct any errors they detected in samples of exchanges in Spanish.

Therefore, as regards the correlations between SST scores and fluency as well as correlations between SST scores and accuracy, Mizera's results seem to corroborate those of Fortkamp (1999, 2000).

Xhafaj (2006) reports on a study which investigated differences in silent pause distribution in L1 and L2 speech production of Brazilian speakers in the attempt to address the relationship among pause distribution, L2 fluency, and working memory capacity. Results suggest that (a) the first languages (English in the case of American speakers and Portuguese in the case of Brazilian speakers) did not differ in terms of pause distribution or mean length of run, (b) the L2 (English) speech of Brazilian presented more pauses and shorter mean length of run than the two first languages (English in the case of American speakers and Portuguese in the case of Brazilian speakers), and (c) significant correlations were found between working memory capacity (as measured by the Speaking Span Test) and frequency of within boundary pauses and mean length of run.

Xhafaj's (2006) results suggest that within boundary pauses seem to be more effective than at boundary pauses in distinguishing more and less fluent L2 speakers, as suggested by Skehan and Foster (2005). Moreover, speakers with higher working memory capacity tend to be more able to sustain L2 fluency and thus present fewer within boundary pauses and longer speech runs. Therefore, individuals' limited attentional resources seem to play a role in fluent L2 speech performance.

Bergsleithner (2007) reports on a study that investigated the relationship among working memory capacity, noticing of L2 forms, and L2 speech production in terms of grammatical accuracy. Results revealed significant correlations among working memory capacity (as measured by the Speaking Span Test), noticing of L2 forms (measured through the use of verbal protocols), and grammatical accuracy of L2 speech

performance. Bergsleithner's (2007) results suggest that the ability to notice L2 forms as well as the use of noticed grammatical forms in L2 speech seem to be mediated by individuals' working memory capacity.

Weissheimer (2007) reports on an exploratory study that investigated the relationship between working memory capacity (as measured by the Speaking Span Test) and the development of L2 speech production (as measured by fluency, accuracy and complexity). Results suggest that both lower and higher spans experience gains in speech production measures from phase I to phase II of the study. However, only lower spans experience statistically significant gains in working memory scores (as measured by the Speaking Span Test). In addition to that, results indicate that the Speaking Span Test was related to L2 speech development in terms of complexity, but not in terms of fluency and accuracy.

In the attempt to explain why only lower spans experienced significant gains in working memory scores, Weissheimer (2007) proposes that a variation in the working memory scores of lower spans may be due to an improvement in domain specific processes - strategies, encoding, and rehearsal. Higher spans, on the other hand, had already been more efficient in controlling attention and in strategy use since the beginning of the experiment and thus did not show any increase. Based on the power law of learning, Weissheimer (2007) proposes that lower capacity individuals have more room for improvement and thus tend to respond more to treatment.

As for the relationship between working memory capacity and L2 development in terms of complexity, Weissheimer (2007) suggests that higher spans are better able to manipulate language, particularly grammatical items, and thus may be more willing to take risks and use more cutting edge language.

2.3.5 Concluding remarks

A few conclusions can be drawn from the studies on the relationship between working memory capacity and L2 speech performance:

1. In general, these results suggest that there is a relationship between working memory and L2 speech performance, and this relationship is a complex one which merits far more scrutiny.

2. Results are not clear cut as for what aspects of L2 performance are related to working memory capacity: (a) Fortkamp (1999) provides evidence for the relationship between working memory capacity and fluency; (b) Fortkamp (2003) provides evidence for the relationship between working memory capacity, fluency, accuracy, and complexity, at the expense of weighted lexical density; (c) Fontanini et al. (2005) found evidence for the relationship between working memory capacity and accuracy, but not fluency; (d) Mizera (2006) found evidence for the relationship between working memory capacity, fluency and accuracy, but only when working memory was measured by means of the Speaking Span Test; (e) Xhafaj (2006) found evidence for the relationship between working memory and fluency when measured by *within boundary* pauses and mean length of run, but not when measured by *at boundary* pauses; (f) Bergsleithner (2007) found evidence for the relationship between working memory and accuracy; and (g) Weissheimer (2007) found evidence for the relationship between working memory and complexity of L2 development.

3. Results across studies seem to suggest that the relationship between working memory and L2 performance may be mediated by a range of other factors such as: (a) level of proficiency, as Fontanini et al. (2005) suggest; and (b) emotional factors as Mizera (2006) suggests.

4. The different measures used across studies to assess L2 speech performance and working memory capacity seem to influence the overall picture of results and thus make comparisons among studies sometimes difficult. Although most studies assessed L2 speech by means of fluency, accuracy and complexity, not always the same dimensions of fluency, accuracy and complexity were used (Fortkamp, 2000; Weissheimer, 2007).

In this section, I focused on empirical studies about the relationship between working memory capacity and L2 speech performance. Now I turn to the construct of pre- task planning as a metacogntive process, which will be followed by a review of empirical studies on the impact of planning on L2 performance.

2.4 Pre-task planning

Skehan (1996) proposes a framework for the implementation of task-based instruction. Within the task-based approach, the main assumption is that "psychological factors and processing conditions are highly relevant to second language learning and second language performance" (Skehan, 1998, p. 93). In this sense, three issues are central as regards task analysis and implementation (Skehan, 1996). First, attention and noticing are essential for L2 learning (Schmidt, 1990). Second, attentional resources are limited (Van Patten, 1990, 1996). Third, in L2 learning and performance, learners draw upon a dual-mode processing system consisting of the exemplar-based system and the rule-based system (Skehan, 1998). The exemplar-based system emphasizes meaning and regards learning in terms of the accumulation of chunks. The rule-based system emphasizes analyzability leading to the development of an open form-oriented system, according to which learning regards growth, change, and complexity of the underlying

system. Interestingly, Feldman-Barrett, Turgade, and Engle (2004) also acknowledge a dual-mode processing system: in associative processing (exemplar-based), information is processed automatically. Thus, associative processing is not under the constraints of working memory imitations. On the other hand, they claim that rule-based processing is subjectively effortful, strategically coordinated to individuals' goals. Thus, rule-based processing is more harshly under the constraints of working memory limitations.

In his framework, Skehan (1996) proposes a cycle of tasks which encompasses pre, mid, and post task activities. Pre-task activities are aimed at enhancing task performance. Mid-task activities focus on the ways in which the tasks are done and are aimed at balancing, reducing or enhancing task difficulty in order to balance learners' attention among the goals of fluency, accuracy, and complexity. Post-tasks activities are aimed mainly at raising awareness for a focus on form. Pre-task activities (e.g., planning) are used to introduce new language, mobilize language, recycle language, ease processing load, and to push learners to interpret the task in more demanding ways (Skehan, 1998).

Based on what has been said, pre-task planning is originally a type of activity which belongs to the pre-task stage in Skehan's (1996) framework to task-based instruction. As quoted earlier, due to an increasing interest in and, consequently, a growing body of research on planning, it has evolved into an area of inquiry in its own right and has become "a burgeoning area of research within task-based language learning" (Ortega, 2005, p.77).

Since planning is a crucial construct of this study, I find it necessary to start by operationalizing the term planning itself. First, it is important to highlight the difference between planning as a subprocess of speech production (Clark & Clark, 1977; Levelt,

1989), and strategic planning (Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996; Menhert, 1998; Ortega, 1999).

As referred to earlier, according to Levelt (1989), the message content is planned in the conceptualizer through the processes of macroplanning and mircoplanning. In macroplanning, the speaker retrieves information to convey his/her communicative intention, the content of the message, whereas in microplanning, the speaker plans the form of the message, which encompasses fixing the appropriate speech act, marking the status of referents as 'given' or 'new', and assigning topic and focus. The processes of macroplanning and mircoplanning result in the preverbal message, which will be, then, turned into a linguistic structure.

In a similar fashion, Clark and Clark (1977) also postulate that speech production is seen as a planned process; thus, speaking may be subdivided into planning and execution. In the planning stage, the language user activates linguistic resources, selects rules as well as items, and gathers them into plans at more complex levels, which will control the execution phase in order achieve a communicative goal (Faerch & Kasper, 1983).

Daneman (1991) also views speaking as involving a highly complex coordination of storage and processing functions, that is, as a skillful coordination of planning and execution requirements. She states that "speakers must *plan* what to say and temporarily *store* the plans until ready to execute them in words, phrases and sentences" (p.446). In this sense, speakers may be planning an utterance while articulating what was previously planned (Clark & Clark, 1977; Daneman, 1991). Therefore, the role of working memory is crucial within the mediation of the planning and execution functions of speaking.

Following these lines, planning is a *cognitive* subprocess of speech production (Clark & Clark, 1977; Daneman, 1991; Faerch & Kasper, 1983; Levelt, 1989). In L1 speech production, planning is subconscious and highly automatic, thus, in most situations, L1 speakers have a considerable amount of ready-made plans or chunks available which contributes to reduce the processing load (Mehnert, 1998). On the other hand, L2 speakers' 'ready-made' plans are more limited, and L2 language users need to construct plans in most communicative situations (Mehnert, 1998), which means that a high degree of cognitive control is demanded. In this sense, planning as a pre-task activity, as proposed by Skehan (1996), may have a positive impact on L2 speech performance for it reduces the amount of online planning during task performance, thus, reducing cognitive strains or processing load.

D'Ely (2004) argues that although planning is essentially a cognitive process which is inherent to the act of speaking, it evolves into a metacognitive process when it is used strategically by the learner. Therefore, D'Ely (2004) defines strategic planning as a problem-solving activity that provides learners the opportunity "to exert some control over what they know towards achieving gains in oral performance" (p.17).

Ortega (2005) highlights the fact that most of the research on planning is product oriented in the sense that it focuses on the impact of planning on performance. She claims for a *process*-product oriented approach in the attempt to focus not only the impact of planning on performance but also on *how* planning assists performance. She claims for more research in the attempt to scrutinize the processes learners engage in when they plan.

In order to state the perspective I take towards strategic planning in this study, I find it necessary to bring the concepts of *strategies* and *processes* into the present scenario. As put by Faerch and Kasper (1980), strategies are utilized by the learner,

which implies agentiveness. Processes, on the other hand, take place within the learner. Later, Faerch and Kasper (1983) state that the term "strategy may refer to a specific subclass of processes" (p.29). Still on the dichotomy between *strategies* and *processes,* Berardi-Coletta, Dominowski, Buyer, and Rellinger (1995) argue that it is not only a strategy per se which enhances performance but the metacognitive processes that may be triggered as a response to a strategy.

Based on the pilot study[7] in which I have also attempted to scrutinize the processes learners engage in during planning, I believe it is difficult to detach 'strategies' from the 'processes' they may trigger in learners. The best way to justify why I believe such a division is difficult to be drawn is by asking a question: What is a strategy when it is not being used by a learner for whatever purpose? It may be a fuzzy concept in a field of study, a possibility for learners on facing some learning tasks and so on.

In my own view, strategies take place when in use by learners and, when in use, they can only be described in terms of learners' behavior. Thus, for the purposes of the present study, I will make no distinctions between strategies and processes. Following Ortega (2005), learners' processes during planning will be operationalized in this study in terms of the strategies[8] employed by them as they plan an oral task.

On taking a *process*-product oriented approach to the study of task-based planning, I envisage planning as encompassing both its impact on L2 speech performance and the processes which are invoked within the learner. Therefore, drawing on D'Ely's (2004) metacognitive perspective on planning and following Ortega's (2005) *process*-product oriented approach to the study of planning, in this

---

[7] The pilot study will be reviewed in the Method chapter.

[8] The framework by O'Malley and Chamot (1990) will be used to report the strategies used by learners. This framework will be reviewed in Section 2.4.3 of this chapter.

study, I shall define strategic planning as a problem-solving activity in which learners may exert some control over their knowledge, and whose outcomes may be assessed in terms of its impact on performance and/or the processes learners engage in.

As for the definition of task, I follow D'Ely (2004) who asserts that "task is a tool devised for teaching/learning and research purposes, the performance of which allows learners to undergo metacognitive processing convey meaning for communicative/learning aims"(p.21).

In this section, I have focused on the construct of pre-task planning as a metacognitive process and stated my perspective towards planning. I turn now to the empirical studies on planning and its effects on L2 speech performance.

2.4.1 The impact of planning on L2 speech performance

Although there is evidence as for the impact of planning on L1 speech performance (Greene, 1984; Greene & Capela, 1986), most of the research on planning has focused on its impact on L2 performance. The seminal study on the impact of planning on L2 performance was carried out by Ellis (1987). He investigated whether planning would have an effect on style shift of three past tense forms (regular past, irregular past and past copula) in narrative discourse.

Data were collected with 17 intermediate L2 learners from various L1 backgrounds under three conditions: planned writing, planned speech and unplanned speech. Ellis (1987) reports mixed results. In relation to the regular past tense, accuracy decreased from condition 1 to 3 (more planning to less planning). As regards the irregular past tense, accuracy remained more or less constant across all conditions. Finally, concerning the past copula, accuracy levels were almost identical for conditions

1 and 2 and markedly lower for condition 3. Thus, the effects of planning may depend on the nature of the linguistic item being investigated. In general terms, however, results indicated that both planning conditions were beneficial to accuracy.

Crookes (1989) reports on a study on the impact of planning on the performance of two groups of 20 Japanese learners of English in two monologic production tasks (Lego and map tasks). Planning was operationalized under two conditions: minimal planning in which participants were not allowed time to plan prior to performance and 10-minute planning. Performance was assessed in terms of accuracy (number of words per error-free T-units, target like use of plural-s and concord, target like use of definite (the) and indefinite (a) articles and complexity (number of words per utterance, number of subordination per T-unit, number of subordination per utterance). In contrast to Ellis (1987), planning did not lead to gains in accuracy but led learners to achieve more complex language.

Foster and Skehan (1996) carried out a study on the influence of planning time and task type on L2 speech performance. They investigated the effects of three different tasks – personal information exchange, narrative and decision making – under three different implementation conditions –unplanned, non-detailed planning, and detailed planning. Participants were 32 pre-intermediate EFL students from different L1 backgrounds. Results indicated that planning led to gains in fluency and complexity and the relationship between the degree of planning and complexity was linear. Nevertheless, the relationship between planning and accuracy was found to be a more complex one. Students' performance was more accurate in the less detailed planning condition. Furthermore, results also indicated stronger effects of planning on the narrative and decision making tasks. Results are discussed in terms of a limited

attentional model of learning and performance, and a trade-off effect among the goals of accuracy, fluency and, complexity is emphasized.

Mehnert (1998) carried out a study on the effects of different amounts of planning time on L2 speech performance. Participants were 31 intermediate learners of German and two tasks were performed – an instruction task and an exposition task – under four conditions – no planning (control group), 1 minute planning, 5 minute planning and 10 minute planning (experimental groups). Results indicated that fluency and lexical density of speech increased in line with the amount of planning time. As regards accuracy, it increased with 1 minute planning only; however, it did not increase as planning time was increased to 5 or 10 minutes. Concerning complexity, results indicated that the most complex speech was achieved in the 10 minute planning condition. Similarly to Foster and Skehan (1996), Mehnert (1998) also discusses results in terms of a limited attentional model of learning and performance, and a trade-off effect among the goals of accuracy, fluency, and complexity is emphasized.

Ortega (1999) investigated the impact of planning on L2 performance and also focused on the processes learners engage in during planning. Retrospective interviews were used in order to document what learners did when they planned their speech. The participants were 64 advanced Spanish learners, and they were all native speakers of American English. Results from learners' performance and interviews indicated that planning time may lead learners to focus on form and produce more fluent and complex language. However, results were mixed as regards accuracy. Planning led to significant gains in the use of the noun-modifier agreement but not in the use of the article system in Spanish. Results corroborate Ellis (1987) in which planning effects on accuracy were also different according to the linguistic item being tested.

Ortega (2005) went further in her investigation of what learners do when they plan. As previously mentioned, she states that most of the research on planning is product oriented, that is, focused on the impact of planning on performance and claims for a more *process*-product approach to the research on planning to reveal what processes learners engage in when planning in order to understand how planning enhances performance. Using interview data from two previous studies, Ortega (1995) and Ortega (1999), Ortega (2005) scrutinized the issue of what processes learners engage in when they plan their performance by analyzing participants' metacognitive responses that provided insights into the cognitive processes associated with learners' strategic planning of tasks.

A variety of strategies – metacognitive (advanced planning, performance evaluation, production monitoring) and cognitive (writing for retrieval, avoidance, translating) – were reported by the learners. Overall, the most frequent strategies were writing, outlining, summarizing, production monitoring, organizational planning, lexical compensation, translating, emphasizing with the listeners, and rehearsing. These strategies highlight the high frequency of retrieval and rehearsal operations within strategic planning (Ortega, 2005).

As regards learners' perceptions about planning, Ortega (2005) reports that most learners view planning as beneficial. They used extra time mainly to organize and formulate thoughts, solve lexical problems, practice/rehearse, and write notes mainly to formulate thoughts, retrieve lexical items, improve lexical choice, help grammatical retrieval and monitoring, and improve overall content. Learners' perceptions of planning also point to the centrality of retrieval and rehearsal operations.

Following Crookes (1989), Ortega (2005) also claims that one of the main benefits of strategic planning is that it enables learners "to access the upper limits of

their interlanguage systems without time pressure, thus, making a wider linguistic repertoire available for subsequent on-line use" (p. 90). In other words, strategic planning reduces the cognitive pressure of online performance.

Despite the evident benefits of strategic planning on performance, Ortega (2005) also reports that some of her participants did not perceive strategic planning as advantageous. Some of the limitations of strategic planning identified by the learners were performance conditions, language expertise, and learner preferences.

As regards performance conditions, low task complexity was mentioned by learners as a reason for planning not being necessary. In relation to language expertise, learners reported lack of transfer to online performance and lack of retrieval. They also reported that planning did not help because what they did not know they could not plan. Concerning learner preferences, some learners are more oriented towards communication while others towards accuracy, regardless of having time to plan. Ortega's (2005) findings are relevant once they provide insights on how planning assists performance as well as point limitations to the effects of planning, thus suggesting that future research is needed on individual differences within the effects of planning.

Also following a *process*-product oriented approach to the study of planning, Sangarun (2005) reports on a study in which 40 Thai Grade 11 EFL participants at the intermediate level performed monologic tasks under four different planning conditions: (a) 10 at the minimal strategic planning condition in which they had no time for strategic planning, (b) 10 at the meaning-focused strategic planning condition in which they were given 15 minutes for strategic planning and were instructed to plan the meaning of their performance, (c) 10 at the form-focused strategic planning condition in which they were given 15 minutes for strategic planning and were instructed to plan the form of their performance, and (d) 10 in the form-meaning strategic planning condition

in which they were given 15 minutes for strategic planning and were instructed to plan both the meaning and form of their performance. Besides investigating the effects of the different planning conditions on performance, Sangarun (2005) also examined participants' actual application of their plans.

In general, results show that participants focused primarily in planning *meaning* in the meaning-focused planning, form-focuesed planning and meaning-form planning conditions. Results are mixed concerning learners' application of their plans. Positive effects of the meaning-form planning condition were revealed for the instruction task, and of the meaning-focused planning condition for the argumentative task in the application of planned meaning. Moreover, positive effects of the meaning-form planning condition were revealed in the application of form for both tasks. Therefore, strategic planning that is aimed at a balance between meaning and form seems to be more effective. These results seem to be in line with VanPatten (1990), who claims that learners will primarily attend to meaning and they will attend to form which is necessary to convey meaning.

Finally, as regards the impact of planning on the quality of oral performance, results indicate that all three strategic planning conditions (meaning-focused planning, form-focused planning and meaning-form planning) lead to better results than the minimal planning condition in terms of accuracy, fluency, and complexity. However, the meaning-form planning condition led to greatest effects on speech performance when compared to the meaning-focused planning and form-focused conditions. Therefore, planning is more effective when it is aimed at leading learners to balance attention between meaning and form.

While Sangarun (2005) focused on differences in planning conditions in terms of planning aimed at form, meaning and form-meaning, Kawauchi (2005) focused on

differences in planning conditions in terms of the activity carried out during planning – rehearsal, writing and reading – and also on individual differences in proficiency levels within the effects of planning on L2 performance.

In Kawauchi (2005), 39 Japanese learners of English with different levels of proficiency participated in the study: (a) 16 low intermediate EFL, (b) 12 high-intermediate EFL, and (c) 11 advanced ESL learners. It was a 'within subjects' study in which learners completed both the unplanned and planned tasks. Learners carried out three narrative tasks that consisted of sets of pictures (library, jogging, and hiking). All tasks were carried out twice, being the first time under the unplanned condition, and the second time either in the planning 'reading', planning 'rehearsal' or planning 'writing' condition.

In the unplanned condition, learners had two minutes to describe their stories based on the set of pictures. Then, learners did the same task again but were allowed ten minutes to plan their stories either through writing, reading or rehearsing. In the writing activity, they were told to write out what they wanted to say. In the reading activity, they were provided with a model passage of the picture story to read and think about how they would tell the story. Finally, in the rehearsal activity, they were told to rehearse by saying aloud what they had tried to produce in the unplanned condition. Learners' performance was assessed in terms of fluency (rate of speech and repetitions), complexity (the number of clauses per T-unit, T-unit length, subordinate clauses, and the number of word types), and accuracy (the past tense markers for copula be, auxiliary verbs, regular and irregular verbs).

Kawauchi (2005) reports no significant differences in results concerning the effects of different strategic planning activities – writing, reading and rehearsal – on performance. Therefore, there seems to be no distinctive benefits among the three

planning conditions. However, there seems to have been a qualitative difference between the input (reading) and output (writing and rehearsal) planning activities concerning the use of low frequency lexical items and problematic structural items. These items showed more target-like use in the reading activity.

In relation to the role of proficiency level on performance after strategic planning, results revealed that there were sgnificant effects for proficiency as well as planning on both fluency measures. In the unplanned condition, there were significant differences among the proficiency groups for number of words with advanced learners showing the greatest performance followed by high intermediate, and low intermediate ones, respectively. However, in the planned condition there were no significant differences between high intermediate EFL and advanced ESL learners, the only significant differences were between these proficiency levels and the low intermediate EFL learners.

Concerning repetitions, the advanced learners were the ones showing the fewest repetitions. The results for the advanced ESL group were significantly different from those for the low and high EFL groups. Despite the fact that the advanced learners showed fewest repetitions, their repetitions in the planned task were significantly more frequently than in the unplanned task.

Significant effects of proficiency and planning were also found for complexity with no interactions on the measures of the number of clauses per T-unit and number of words per T-unit. The low EFL group differed significantly from the high EFL group, which also differed from the advanced ELS group, with the advanced group showing the greatest performance followed by high and low intermediate groups, respectively. The results for T-units also showed that planned performances were significantly more complex than unplanned performances. As for subordination and word types there was a

significant interaction between proficiency and planning. The low intermediate EFL group differed significantly from both high intermediate EFL group and advanced ESL group, but there was no significant differences between the latter two groups under the planned condition.

In addition, there were no significant differences between the unplanned and planned performances of advanced learners, which means that strategic planning did not lead advanced learners to use more complex language as far as subordination is concerned (Kawauchi, 2005). As for word types, both unplanned and planned conditions revealed that advance learners showed the greatest performance followed by high and low intermediate learners, respectively.

Finally, as regards accuracy, significant effects of proficiency and planning were found for correct use of past tense, and there was also an interaction. The low EFL group differed significantly from both high EFL and advanced ESL groups in both unplanned and planned conditions, but there were significant differences between the latter two groups. Moreover, accuracy also varied according to the verb categories, with the use of past irregular verbs showing the highest levels of accuracy in both planned and unplanned conditions whereas the past copula tended to show the lowest levels. These results are in line with Ellis (1987), for which accuracy levels also varied according to the linguistic item under investigation.

In brief, while the high intermediate EFL group tended to show highest gains in fluency and complexity, the low intermediate EFL learners showed the highest gains in accuracy. The advanced ESL learners presented the greatest performance under the unplanned condition when compared to the two other groups; however, the performance of advanced learners tended to be similar to the performance of high intermediate ones under planned conditions. These results provide evidence for a role of proficiency

within the effects of strategic planning on performance, that is, “there seems to be a level beyond which planning will have only a limited effect” (Kawauchi, 2005, p. 164).

While the studies reviewed so far have dealt with strategic planning, Yuan and Ellis (2003) bring the issue of within-task (online) planning into play. Yuan and Ellis (2003) theorize online planning as involving a type of speech production which encompasses both ‘careful’ production and monitoring. In order to propose the concept of on-line planning, they draw on Levelt (1989), who proposes that internal speech is available for monitoring before production while overt speech is available for monitoring after production. Moreover, they also draw on Krashen’s (1991) proposals for monitoring that focus on editing immediately before production.

They define online planning as “the process by which speakers attend carefully to the formulation stage during speech planning and engage in pre-production and post-production monitoring of their speech acts” (Yuan & Ellis, 2003, p.6). Yuan and Ellis (2003) acknowledge online planning as being required in all types of speech; however, they highlight that they use the term to refer to ‘careful’ speech production in which learners may plan and replan message conceptualization and formulation as opposed to ‘rapid’ speech production, which involves greater extents of improvisation.

Yuan and Ellis (2003) report on a study that investigated the effects of pre-task and online planning on L2 speech performance. A population of 42 undergraduate Chinese learners of English performed oral narratives based on pictures. Participants were randomly divided in three groups according to the three planning conditions: no planning, pre-task planning and online planning.

In the no planning condition, participants were required to start their performance immediately after studying the set of pictures for 0.5 minute. They had 5 minutes to perform the task itself. In the pre-task planning condition, participants had 10

minutes to plan their task performance and as in the no-planning condition, they had to perform the task within 5 minutes. In the online planning condition, participants were required to start performance of the task immediately after studying the pictures for 0.5 minute but they had no time limit to the performance of the task.

According to Yuan and Ellis (2003), setting a time limit would restrict on-line planning in both no planning and pre-task planning conditions, whereas unlimited time would allow ample opportunities for on-line planning to take place. Performance was assessed in terms of fluency (number of syllables per minute), complexity (syntactic complexity, syntactic variety, and mean segmental type-token ratio), and accuracy (error-free clauses and correct verb forms).

Results revealed that online planners spent significantly longer on tasks than both non-planners and pre-task planners, which indicates that the unlimited time for task performance may actually have been used in planning speech online. As regards the impact of planning on fluency, the pre-task planners showed the greatest effects followed by non-planners and on-line planners, respectively. In other words, online planners presented the slowest speech rate and the highest number of repetitions and/or reformulations. Thus, online planning seems to be detrimental for fluency, that is, not having a time pressure seems to engage learners in monitoring their performance at the expense of their speech rate.

As for complexity, mixed results were reported. Both online and pre-task planning groups outperformed the non-planning group in syntactical complexity, but no significant differences were found between online and pre-task planners. Similar results were obtained for syntactical variety but differences among groups did not reach significance. As for lexical variety, pre-task planners showed the best performance, but only the differences between pre-task and online planners were statistically significant.

Thus, lexical variety seems to increase when there is a time pressure and learners are allowed to plan ahead (Yuan & Ellis, 2003). Finally, concerning accuracy, online planners had the best performance in both measures: error-free clauses and error-free verb forms, followed by pre-task planners and non-planners, respectively.

Therefore, Yuan and Ellis (2003) conclude that pre-task planning led to higher scores than online planning in fluency, whereas online planning led to higher scores in accuracy. Results were mixed for complexity, with no differences in grammar but with significant differences as for lexical variety in favor of the pre-task planners.

Yuan and Ellis (2003) discuss their results in terms of a limited attentional model of learning and performance and emphasize trade-off effects among the different aspects of speech production. Foster and Skehan (1996) and Mehnert (1998) have also proposed that there are trade-off effects among the different goals of speech performance: fluency, accuracy, and complexity.

Moreover, Yuan and Ellis (2003) add that there may be a dual trade-off. First, the competition for attentional resources involves fluency and accuracy. If learners are allowed unlimited time to engage in online planning during performance, they will focus on accuracy at the expense of fluency. On the other hand, if learners are allowed to engage in pre-task planning, they will focus on fluency at the expense of accuracy during task performance.

The second trade-off, according to Yuan and Ellis (2003), occurs between grammatical accuracy and lexical variety. As evident in their results, pre-task planners tended to show more lexically varied but less grammatically accurate performance, whereas online planners tended to show more grammatically accurate but less lexically varied performance.

As previously stated, Yuan and Ellis (2003) conceptualize online planning as encompassing careful production and monitoring and operationalize on-line planning by providing unlimited time for task performance. In this sense, Yuan and Ellis (2003) suggest that speakers will attend predominantly to the formulation stage during online planning.

Skehan and Foster (2005) argue that, since planning is an unobservable activity, it has to be treated as a construct. In this sense, in both pre-task and online planning, researchers are left to infer what operations learners undergo based on the ways task conditions are manipulated and on the effects of planning on performance. With online planning, for instance, they claim that there is no *direct* evidence that learners will be engaged in psycholinguistic operations concerned with ongoing planning such as planning the form of future utterances.

Skehan and Foster (2005) claim that the concept of online planning needs further supplementation with more direct manifestation and evidence of the psycholinguistic processes it may involve. They view online speech compensation measures such as filled pauses and mid-clause pauses as promising in the attempt to provide evidence for on-line planning.

Skehan and Foster (2005) report on a study which that investigated whether (a) different forms of strategic planning- detailed and non-detailed- (see Foster & Skehan, 1996) impact differently upon performance, (b) length of time on task influences performance, and (c) introduction of surprise information during task influences performance.

Sixty-one English learners from a variety of L1 backgrounds and attending to one of six different intermediate EFL classes were the participants of the study. Participants of each class performed the task that consisted of choosing an appropriate

sentence for a list of people found guilty in crimes in which the victims died or were seriously injured. The classes performed the tasks under different strategic planning conditions.

Classes A and B were not allowed strategic planning time and were instructed to start performing the task after reading the description of the crimes. Classes C and D were allowed to plan their performance for 10 minutes before the task began. Classes D and F were also allowed to plan their performance for 10 minutes and they were also given some guidance notes on how to use their planning time, on how to focus on what to say, and on how to say it. Moreover, classes A, C, and E learners were interrupted after 5 minutes of task time and were given surprise information about their crimes such as further details about the victims and killers. These pieces of surprise information were designed to take learners away from any strategic planning they had carried out and incorporate the new details through on-line planning.

Performance was assessed for both the first five minute and second five minute period in terms of accuracy, fluency, and complexity. Accuracy was measured by the percentage of error-free clauses and by the clauses that were greater than four words and error-free. Fluency was expressed through measures of breakdown fluency (number of pauses greater than one second, total silence per five minutes that were divided in end-clause pauses and mid-clause pauses; filled pauses, length of run) and measures of repair fluency (reformulations, false starts and repetitions). Complexity was measured by dividing the data into syntactic clauses and AS-units and expressed as the ratio of clauses to AS-units[9].

[9] As-unit is defined as a single speaker's utterance consisting of na independent clause, or sub-clausal unit, together with any subordinate clause (s) associated with either (Foster et al. 2000, p. 365).

Differently from Foster and Skehan (1996), which indicated that undetailed planning generated the highest levels of accuracy, it was detailed planning which yielded the best accuracy results in Skehan and Foster (2005). As for complexity, the detailed planning condition also led to greater results than the no planning and undetailed planning conditions.

As regards fluency, the end-of clause measure was the only one that reached significant differences in both intervals (first and second five minute period) and the significance was the same with the no planning condition showing more pauses than both strategic planning conditions. There were no significant differences between the two strategic planning conditions, but the detailed planners exhibited more filled pauses than the other two conditions. In addition, time seems to have effects on performance. Results revealed decreases in performance in the last five minutes. Finally, no evidence was found as for the impact of surprise information on performance.

Skehan and Foster (2005) shed some light on the issues of pauses once they did not treat all pauses in the same way. According to them, pauses at the end of clauses are more natural whereas mid-clause pauses seem to be an indication of learners' inability to deal with the pressures of real time communication. Thus, they argue that mid-clause and filled pauses may be an indication of on-line planning in order to deal with the loads of real time performance.

D'Ely (2006) reports on a study that was, to the best of my knowledge, the first one in task-based planning carried out in Brazil. She investigated the impact of different planning conditions – no planning, strategic planning, repetition, strategic planning *plus* repetition and strategic planning *for* repetition – on learners' oral performance. A population of 47 intermediate Brazilian learners of English divided in

five groups performed narrative tasks under these different conditions. L2 speech performance was assessed in terms of fluency, complexity, accuracy, and lexical density.

In general, results show that repetition, strategic planning *plus* repetition, and strategic planning *for* repetition yielded significant gains in some aspects of performance: (a) lexical density and accuracy for the repetition group, (b) lexical density for the strategic planning *plus* repetition group, and (c) accuracy, complexity, and lexical density for the strategic planning *for* repetition group. Surprisingly, the strategic planning condition did not yield gains in oral performance.

In face of these dismissive results concerning the impact of strategic planning (particularly on fluency), D'Ely (2006) emphasizes the role of linguistic knowledge and suggests that there may be a great tension between what learners know, the conditions under which they perform, and the metacognitive processing these performing conditions may evoke. She also raises the possibility that learners in the strategic planning condition may have felt as if they were being evaluated; thus, planning did not impact on performance, which corroborates results of the study by Elder and Iwashita (2005).

Moreover, D'Ely (2006) argues that a range of factors – the nature of the task, learners' focus of attention during planning, learners' effectiveness in implementing, and retrieving of planned information – seem to influence the effects of planning on performance. She also highlights that, when too much attention is devoted to form (as in the case of learners performing under the planning *for* repetition condition), fluency may be penalized as a consequence of monitoring.

Up to this point, all studies reviewed dealt with planning and performance of *adult* learners with proficiency ranging from *intermediate to advanced* levels. However, there have also been attempts to investigate the impact of planning on the performance

of *children* (Philp, Oliver, & Mackey, 2006) as well as on the performance of *beginners* (Mochizuko & Ortega, 2008).

Philp et al. (2006) report on a study which focused on the impact of planning on the performance of children in interactional contexts. The study investigated whether different amounts of planning time would lead to more provision of feedback and more quality of speech in terms of fluency, accuracy and complexity. In their study, 21 dyads of ELS learners from ages of five to twelve years old performed three communicative tasks under the conditions of 0 min, 2 min and 5 min of planning time.

In general, results show that children's provision of feedback was enhanced when they had no time or a short time for planning. As for fluency and accuracy, results suggested no differences across different amounts of planning time. Concerning complexity, it was enhanced under the 5 min planning condition when compared to the 0 min and 2 min conditions.

Therefore, in terms of provision of feedback, planning did not increase learning opportunities; in terms of fluency and accuracy, planning did not yield gains. Philp et al. (2006) also found that learners perceived the time of 5 min planning as too long, which suggests that the impact of planning on performance may be mediated by age, in addition to factors related to task type, structure, and/or complexity.

According to Armsbruster (1983), the development of metacognition appears to be related to proficiency in learning. In other words, learner characteristics (e.g. strategies) are age and experience dependent. Following these lines, interesting questions to be pursued on the impact of planning on the performance of children seem to be: (a) In what mental processes do *children* engage when they plan? and (b) How are the processes children engage different from the processes adults engage? Efforts in

this direction may help explain why planning yielded to rather small effects in Philp et al. (2006) when compared to planning studies examining performance of adults.

Mochizuki and Ortega (2008) report on a study that investigated whether guided pre-task planning involving a specific grammatical feature (relative clauses) would be an appropriate pedagogical tool to be used with beginning levels of proficiency in foreign language contexts. A population of 56 high-school students in Japan was divided in three groups and each group performed a narrative task under a different planning condition: no planning, 5 minutes of unguided planning, and 5 minutes of guided planning that included a grammar handout on relative clauses. Speech performance was analyzed in terms of task essentialness, use of relativization (amount and quality), fluency, and complexity.

Overall results show that guided planning led to greater relativization both in the amount and accuracy of use when compared to the no planning and unguided planning conditions. Moreover, the guided planning led to levels of fluency and complexity that were similar to the no planning and unguided planning. Based on these results, Mochizuki and Ortega (2008) advance the proposal that guided planning that involves specific grammatical features may be a suitable pedagogical tool to be used with beginning levels in foreign language classrooms since this type of guided planning may lead to a balance between communication and grammar. In terms of task essentialness, these researchers found that the design of the study made relative clauses useful for task completion, however not essential.

The studies reviewed so far have provided evidence for the effects planning may have upon task performance in classroom and laboratory contexts. However, there have been also efforts to provide evidence for the impact of planning on task performance in a testing situation.

Wigglesworth (2001) reports on a study in which she focused on the impact of task variation on learners' performance in informal classroom assessments. She operationalized three variables in her study: (a) the cognitive difficulty of the task (5 types of task were used), (b) type of interlocutor (native or non-native speaker), and (c) presence or absence of strategic planning. The planning condition encompassed manipulation of task structure (either structured or unstructured task) and task familiarity.

Speech performance was assessed qualitatively by external, experienced raters who evaluated performance in terms of grammar, fluency, cohesion, vocabulary, and intelligibility. Task difficulty was also evaluated by external raters and learners. Overall results suggest that structure makes the task easier in most cases (task types 2, 3 and 4). As for familiarity, it also appeared to make the task easier. However, less familiar tasks also appeared to be easier when manipulated in conjunction with non-native speakers as the interlocutors. In addition, planning led to more complex performance, at the expense of fluency and accuracy. Moreover, results revealed a complex interaction between task characteristics and task conditions, with both affecting learners' performance in testing situations.

Elder and Iwashita (2005) set out to investigate the effects of strategic planning on monologic performance in the context of a tape-based test of speaking proficiency. Participants were 197 ESL learners performed narrative tasks based on a sequenced set of pictures. Participants were also asked to answer questionnaires after each task in order to gather data about their perceptions of the tasks conducted under the planning and no planning conditions.

Under the planning condition, participants had 3 minutes to plan plus 75 seconds to read the instructions for the task, whereas in the no planning condition participants had only the 75 seconds to read the instructions. Performance was assessed qualitatively

through the use of rating scales for fluency, accuracy, and complexity; 14 experienced raters were selected for the assessment of the speaking tasks. Moreover, a subset of 36 subjects was randomly selected for quantitative analysis in terms of fluency (repetitions, false starts, hesitations, and pauses, divided by the total amount of speech), accuracy (percentage of error-free clauses), and complexity (number of clauses per c-units).

Overall results revealed that planning time made no significant differences as to the scores of the candidates in terms of fluency, accuracy, and complexity in qualitative assessment. In the quantitative analysis, results revealed a higher number of pauses, reformulations and repetitions, and a lower number of error-free clauses in the no planning condition. However, these differences did not reach significance. As regards participants' perceptions of taking tests under planning or no planning conditions, results indicated that the planning condition was perceived to make the task easier, whereas telling the story under the no planning condition was perceived to be more enjoyable. However, there were no statistically significant differences in task difficulty and task enjoyment between the planning and no planning conditions.

These results are not in line with previous research (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999) in which planning led to benefits in performance. Elder and Iwashita (2005) raise some possible reasons for the results obtained such as: (a) lack of task complexity, the simple narratives used may not have been conductive to elicit complex language; (b) simple task instruction ("you will have three minutes to think about the story") may not have evoked a focus on form; and (c) length of planning time (3 minutes) was insufficient to enhance performance.

Tavakoli and Skehan (2005) report on a study which set out to investigate the effects of task structure, strategic planning, and proficiency level on test performance. Pre-task planning and level of proficiency were operationalized in a between-participant

design whereas task structure was operationalized in a within-participant design with all participants performing all four tasks.

The tasks used were narratives based on sets of pictures. Task structure was operationalized in terms of type and degree of structure. The tasks varying the type of structure were (a) a problem-solution, the football task which was a picture series with a transparent problem-solution structure a well presented sequential organization, and (b) The picnic task which presented a clear sequential organization but the problem was implicitly stated and revealed only in the last picture of the set. Thus, the picnic task did not present a clear problem-solution structure. On the other hand, the tasks varying the degree of structure were (a) The Unlucky man which had a loosely presented sequential organization, and (b) the walk-man task which did not contain any sequential organization. Participants were 80 language learners from two different levels of proficiency, that is, elementary and intermediate.

Participants had 5 minutes to plan the tasks under the strategic planning condition and 30 seconds under the no planning condition. Performance was assessed quantitatively in terms of fluency (false-starts, reformulations, replacement, speech rate, length of run, number of pauses, total pausing time), accuracy (error-free clauses), and complexity (ratio of clauses to AS units). Perceptions of task difficulty were also assessed through questionnaires.

Overall results suggest that for number of pauses and speaking time, the two structured task generated significantly more fluent language. Moreover, the two structured tasks generated significantly more accurate language than the unstructured ones. As for complexity, the picnic task (containing a clear organization sequence but implicit problem-solution structure) yielded significantly more complex language use.

Concerning the effects of strategic planning on performance, fluency improved significantly in the measures of total silence, length of run pause length, speaking time, and speech rate under the planning condition. In addition, measures of temporal fluency are significantly higher in the performance of intermediate proficiency learners when compared to elementary learners. It is important to highlight that the impact of planning on total silence, speaking time, and pause length is greater than the impact of proficiency. In other words, it seems more advantageous to be an elementary proficient planner than an intermediate proficient non planner.

Similarly, planning led to significant gains in accuracy, and language performed by intermediate proficiency learners is significantly more accurate than elementary proficiency learners' language; however, differently from fluency, the effects of proficiency on accuracy are greater than the effects of strategic planning. Finally, planning also led to more complex language use and similarly to accuracy, the effects of proficiency of complexity are also greater than the effects of strategic planning. As far as task difficulty is concerned, answers to the questionnaires revealed that unstructured tasks were perceived as more difficult in both planning conditions. These results raise an interesting issue as regards the roles of strategic planning and proficiency level in performance, suggesting that strategic planning will enhance fluency regardless of proficiency levels; as for accuracy and complexity there might be limits for strategic planning benefits beyond which it is proficiency level that will play a greater role.

2.4.2 Concluding remarks

A few conclusions can be drawn from empirical studies conducted to date that have examined the impact of planning on L2 performance:

1. Overall results suggest a stronger impact of planning on fluency (e.g. Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999), and on complexity (e.g., Crookes, 1989; Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan & Ellis, 2003), whereas results are more mixed for accuracy (e.g., Ellis, 1987; Foster & Skehan, 1999; Mehnert, 1998; Ortega, 1999).

2. Most studies so far have targeted adult populations with levels of proficiency ranging from intermediate to advanced levels, except for the study by Philp et al. (2006), which examined the performance of children, and the study by Mochizuki and Ortega (2008), which examined beginning levels of proficiency.

3. Research targeted at performance in testing situations has yielded mixed results across studies, which may be explained in terms of the differences in the testing contexts (Ellis, 2005).

4. Most of the studies have taken a product oriented approach focusing on the impact of planning on performance. Only three studies (Ortega, 1999, 2005; Sangarun, 2005) have taken a *process*-product oriented approach in the attempt to scrutinize learners' processes during planning.

5. Most studies have focused on different types of (a) tasks, (b) planning and (c) amounts of planning time (Foster & Skehan, 1996; Mehnert, 1998; Sangarun, 2005, among others). Only two studies – Kawauchi (2005) and Ortega (2005) – have examined individual differences, and in both of these studies the focus of individual differences was on different levels of proficiency. No other aspect of individual differences (e.g., motivation, anxiety, working memory) has been investigated this far.

6. The trend of research on planning suggests that there are trade-off effects among fluency, accuracy, and complexity in the context of learners' limited attentional resources. However, to the best of my knowledge, no study to date has investigated

working memory capacity as a feasible variable affecting L2 performance under planning conditions.

These conclusions suggest a potential gap in the research on pre-task planning, which I will define in the next section.

2.4.3 A gap in the research on pre-task planning

Based on the few conclusions drawn from the research on the impact of planning on L2 performance, three issues merit to be highlighted. First, most of the research on planning is product oriented. Little research has been carried out in the attempt to scrutinize learners' processes so that one can reach firmer grounds on *how* planning assists performance. Second, little research has examined how individual differences play a role within the effects of planning on performance; Furthermore, the few studies on individual differences focused solely upon differences in proficiency levels. Third, several studies explain trade-off effects in terms of learners' limited attentional resources. Nevertheless, no study to date, to the best of my knowledge, has examined individual differences in working memory capacity as a feasible variable for affecting learners' L2 performance in planning conditions.

Bearing that in mind, I believe that individual differences in working memory capacity constitute a potential gap in the research on task-based planning. Thus, following a *process*-product oriented approach, the study reported in this dissertation seeks to examine how individual differences in working memory capacity affect both the impact of planning on L2 speech performance *and* the processes learners engage when they plan an oral task.

In this section, I have reviewed empirical studies on planning and have pointed out a gap in this field of research. In the next section, I will briefly review the framework used to report the strategies learners employ when they plan an oral task.

2.4.3 The framework by O' Malley and Chamot (1990)

Definitions and models of learning strategies abound in the literature (O'Malley & Chamot, 1990; Oxford, 1990; Rubin, 1975; Wenden, 1991). Although I acknowledge that there are several perspectives on strategies and, as a consequence, there are several strategies and frameworks of strategies available in the literature, these issues are beyond the scope of the present study.

As previously stated, I take a *process*-product oriented approach in this study as an attempt to scrutinize learners' processes, and these processes will be focused in terms of the strategies learners employ during planning. In other words, strategies will be the means to appreciate learners' processes. Following Ortega (2005)[10] in one of her seminal studies on this perspective, I will adopt the framework of strategies proposed by O'Malley and Chamot (1990) in order to allow this study to be comparable to hers.

O'Malley and Chamot (1990) define strategies as "ways of processing information that enhance comprehension, learning, use or retention of the information" (1990, p.1). In their framework[11], strategies can be divided in three main categories: metacognitive strategies, cognitive strategies, and socio-affective strategies. Metacognitive strategies require planning, thinking of a task as it takes place, monitoring one's production or comprehension and evaluating performance or learning after an activity is completed (O' Malley & Chamot, 1990). Examples of metacognitive

---

[10] Ortega (2005) adopted O'Malley and Chamot (1990) and Oxford (1990) since the only purpose of her study was to scrutinize learners' processes. However, for the purpose of simplification, I will focus only on O'Malley and Chamot (1990).

[11] See O'Malley and Chamot (1990) for an extensive, detailed description and explanation of the framework.

strategies are organizational planning, problem identification, monitoring, evaluation, selective attention, and rehearsal.

Following O' Malley and Chamot (1990), the following metacognitive strategies were defined:

1. *Organizational planning* concerns the planning of parts, sequence, and main ideas to be expressed. In the present study, it included overall organization carried out before the actual planning of oral performance started. It included sequencing the pictures, making sense of the pictures, defining a main idea for the content of the story and so on.

2. *Problem identification* concerns awareness of a problem to be solved, which may not be restricted to language problems but also when learners have doubts on what to do in general such as which picture should be the beginning or the end of the story, what they should do if they forget the pictures and so on.

3. *Monitoring* concerns production checking while it takes place. However, since the strategies are taking place during planning, in the present study, monitoring concerns checking and correcting language production during the process of planning performance.

4. *Evaluation* regards judging how well one has accomplished the task. In other words, judging how well one is planning oral performance.

5. *Selective attention* regards attending to or scanning key words, phrases, sentences, linguistic markers, sentences or types of information. This strategy is more related to reading and listening comprehension. During planning, instances of selective attention were commonly classified as other strategies. For instance, when a learner is attending particularly to the pictures to make sense and sequence then, this strategy, although implies selective attention, was classified as organizational planning. When

learners focus on grammar mistakes of specific linguistic features, this strategy, although implies selective attention, was classified as monitoring.

6. *Rehearsal* concerns practicing the language to be used. For the purposes of the present study, rehearsal regards practicing the planning of the oral narrative either by reading what was planned or by practicing the narratives mentally.

According to O' Malley and Chamot (1990), cognitive strategies are more limited to a specific task and involve more direct manipulation of material. Examples of cognitive strategies are writing, summarizing, outlining, grouping, lexical search and compensation, translating, imagery, contextualization, elaboration, and avoidance.

Following O' Malley and Chamot (1990) and Ortega (2005), the following cognitive strategies were defined:

1. *Writing/summarizing/ outlining* were grouped together for the purpose of simplification. This concerns all types of written production during planning: writing words, sentences, paragraphs, outlines, and summaries.

2. *Grouping* regards classifying words, terminology, number, and concepts according to their attributes.

3. *Imagery* regards using visual images (either mentally or by drawing) to understand and/or remember information.

4. *Lexical compensation* regards substituting words unknown whereas *avoidance* concerns circumventing an intending meaning/idea of being expressed. An example of lexical compensation is when a learners does not know how to say 'peas' and decides to substitute the unknown word by a familiar one 'beans'. An example of avoidance is when a learner wants to express that 'a man is not brave' and decides to change this idea by expressing the idea that the 'man doesn't like to argue and never answers to what his wife says'. Lexical compensation and avoidance seem to interact.

6. *Lexical search* was added to the analysis of the present study to refer to instances when learners explicitly verbalize to be searching for words and lexical searches which are solved by means of successful retrieval of the lexical item being searched. O' Malley and Chamot (1990) do not include this strategy in their framework, and Ortega (2005) does not include a category simply called *lexical search* to her analysis. For them, lexical compensations and avoidance are all instances of lexical searches. Although I agree that compensation and avoidance imply lexical searches, I believe it is also important to highlight the instances in which learners verbalize only the search itself (e.g., without mentioning how they are going to solve problems *if* they don't remember certain lexical items) as well as lexical searches that end up being solved by successful retrieval of the proper lexical item since learners of the present study frequently searched and retrieved the proper lexical items.

7. *Elaboration* concerns improving one's performance by relating new information to prior knowledge, by making meaningful personal associations with the new information, and by attempting to improve and/or embellish performance.

Finally, socio-affective strategies are related to social-mediating activities and interacting with others (O' Malley & Chamot, 1990). Examples of socio-affective strategies are cooperation, question for clarification, appeal for help, and lowering anxiety. Following O'Malley and Chamot (1990), in the present study the following socio-affective strategies were defined:

1. *Question for clarification* (or appeal for help) refers to instances when learners are not able to cope with the demands of a task by themselves and ask others for help. In the present study, it refers to instances when learners ask the help of the present researcher.

2. *Lowering anxiety* concerns using mental techniques that helps one feel comfort or competent.

In the present study, strategies employed by learners will be analyzed *qualitatively* in order to establish overall processes learners engage when they plan, and also *quantitatively* in order to examine whether individuals with higher and lower working memory capacity differ in the processes they engage in when they plan. Having described the framework by O' Malley and Chamot (1990), I turn now to the means used to assess strategies employed by learners – the verbal protocols.

2.4.4 Verbal protocols

In order to document the processes learners engage in when they plan an oral task, verbal reports were carried out. Ortega (1999, 2005) carried out retrospective interviews for the purpose of documenting learners' mental processes. However, the disadvantage of retrospective protocols is the possibility of memory constraints, that is, participants may forget what they did during planning. Sangarun (2005) carried out think aloud protocols in order to document what learners plan. Leow and Morgan-Short (2004) suggest that introspective protocols, such as think aloud protocols, should be employed in order to avoid memory constraints.

Introspective protocols have been extensively used in the realm of SLA to investigate L1 and L2 strategies and also in problem-solving tasks (Leow & Morgan-Short, 2004). Despite the fact that verbal protocols may offer benefits, they are not without risks. According to Leow (2002), a potential criticism to introspective protocols concerns the issue of reactivity. In other words, when thinking aloud, participants'

internal processes may differ from their internal processes when they are not verbalizing what they think.

In the attempt to scrutinize the issue of reactivity, Leow and Morgan-Short (2004) report on a study that investigated whether thinking aloud would be detrimental to learners' performance on a reading and written production task. Results revealed that reactivity was not an issue affecting learners' performance. Likewise, Simon and Ericsson (1993) also found no reactivity effect for problem solving tasks.

Based on Leow and Morgan-Short (2004) as well as Ericsson and Simon (1993), Guará-Tavares (2005) employed think aloud protocols in the attempt to document what learners plan, and they were also revealing in terms of documenting learners' mental processes. However, results also revealed that participants in Guará-Tavares (2005) used the think aloud as performance itself, they somehow rehearsed performance during the think aloud procedures, thus planning overlapped with task rehearsal. Therefore, it seems that retrospective protocols have the disadvantage of memory constraints whereas introspective protocols seem to lead to an overlapping between pre-task planning and task rehearsal, making it hard to distinguish whether the gains in performance would be due to pre-task planning or rehearsal.

Leow and Morgan-Short (2004) distinguish two types of retrospective protocols: (a) retrospective on-line, carried out after some sort of processing has taken place during specific breaks in the actual task, (b) retrospective off-line, carried out immediately after a whole task has taken place. In the attempt to avoid memory constraints as well as an overlapping between planning and rehearsal, *retrospective on-line* protocols were selected for the pilot study (Guará-Tavares, 2006). These protocols are carried out after some sort of processing during specific breaks during the actual performance of the task; thus, they would be more effective for avoiding memory

constraints since participants would not have to complete the whole task before verbalizing what they were planning. In addition to that, retrospective online protocols also seemed to be more effective in avoiding the overlapping between pre-task planning and task rehearsal since participants were not required to verbalize what they were planning during the whole time.

Bearing the possible advantages of retrospective online protocols in mind, the pilot study (Guará-Tavares, 2006) employed this type verbal protocols in order to document the processes learners engage in when they plan. Results revealed that these protocols were effective in eliciting learners' processes during planning. Consequently, *retrospective online* protocols were selected for the present study.

In brief, studies in other fields (Ericsson & Simon, 1993) and in SLA (leow & Morgan-Short, 2004) have shown no reactivity effects. Results of the pilot study (Guará-Tavares, 2006) have also suggested that retrospective online protocols seem effective to document what learners plan. However, it is important to highlight that the issue of reactivity, claim Leow and Morgan-Short (2004), still needs further empirical scrutiny.

Along this chapter, I have reviewed relevant literature for this study. In the next chapter, I will describe the methodology used for data collection and data analysis.

CHAPTER III

METHOD

3.1 Introduction

In order to investigate the relationship among pre-task planning, working memory capacity, and L2 speech performance, a cross sectional, experimental and quantitative study was conducted. Although the data analysis was predominantly quantitative, there was also qualitative analysis in the attempt to focus on *how* planning assists performances by scrutinizing the mental processes learners engage in when they plan an oral task.

This chapter describes in detail the method for conducting the study and analyzing the data. First, it presents the objective, questions, and hypotheses which motivated the study. Then, it presents information about the participants involved, the procedures for the selection of these participants, and the instruments of data collection. Finally, it describes the procedures for data transcription, data analysis, and reliability analysis.

3.2 Objectives

The main objective of the study is to investigate the relationship among pre-task planning, working memory capacity and L2 speech performance. In addition to that, the study also aims at examining how planning assists L2 speech performance by

scrutinizing the processes learners engage in when they plan. With these broader objectives in mind, the following specific objectives are pursued:

1. To investigate the relationship between working memory capacity scores and measures of L2 speech performance in no planning conditions.
2. To investigate whether planning leads to significant differences on L2 speech performance.
3. To investigate the relationship between working memory capacity scores and measures of L2 speech performance in planning conditions.
4. To investigate whether planning leads to significant differences on L2 speech performance.
5. To investigate the mental processes learners engage when they plan.

3.3 Research Questions

Based on the objectives just mentioned, the following research questions were generated:

1. Does speech performance under no planning condition significantly correlate with learners' WM capacity?
2. Does pre-task planning opportunity significantly increase fluency, accuracy, and complexity of L2 speech performance?
3. Does L2 speech performance under pre-task planning condition correlate significantly with learners' WM capacity?
4. Do higher working memory span participants significantly outperform lower working memory span participants in terms of L2 speech performance under pre-task planning condition?

5. What mental processes do learners engage in when they plan an oral task?
6. Do higher and lower span individuals differ in terms of the mental processes they engage in when they plan?

3.4 Hypotheses

Drawing on the objectives and research questions, the following hypotheses were formulated.

Research question 1 has generated Hypotheses 1, 2, and 3:

- **Hypothesis 1**: Participants' working memory capacity scores will significantly correlate with **fluency** measures of L2 speech performance under no planning condition.
- **Hypothesis 2**: Participants' working memory capacity scores will significantly correlate with **accuracy** measures of L2 speech performance under no planning condition.
- **Hypothesis 3:** Participants' working memory capacity scores will significantly correlate with **complexity** measures of L2 speech performance under no planning condition.

Research Question 2 has generated Hypotheses 4, 5, and 6:

- **Hypothesis 4**: Under planning condition, there will be greater **fluency** for the experimental group when compared to the control group.
- **Hypothesis 5**: Under pre-task planning condition, there will be greater **accuracy** for the experimental group when compared to the control group.
- **Hypothesis 6**: Under pre-task planning condition, there will be greater **complexity** for the experimental group when compared to the control group.

Research Question 3 has generated Hypotheses 7, 8, and 9:

- **Hypothesis 7:** Participants' working memory capacity scores will significantly correlate with **fluency** measures of L2 speech performance under pre-task planning condition.
- **Hypothesis 8**: Participants' working memory capacity scores will significantly correlate with **accuracy** measures of L2 speech performance under pre-task planning condition.
- **Hypothesis 9**: Participants' working memory capacity scores will significantly correlate with **complexity** measures of L2 speech performance under pre-task planning condition.

Research Question 3 has generated Hypotheses 10, 11, and 12:

- **Hypothesis 10**: Within the experimental group, under pre-task planning condition, higher working memory spans will significantly outperform lower working memory spans as regards **fluency** of L2 speech production.
- **Hypothesis 11**: Within the experimental group, under pre-task planning condition, higher working memory spans will significantly outperform lower working memory spans as regards **accuracy** of L2 speech production.
- **Hypothesis 12**: Within the experimental group, under pre-task planning condition, higher working memory spans will significantly outperform lower working memory spans as regards **complexity** of L2 speech production.

Research Question 5 and 6 have generated Hypotheses 13 and 14:

- **Hypothesis 13**: When planning an oral task, learners will engage in the following processes: (a) organization of ideas, (b) lexical-grammatical search, (c) task rehearsal, and (d) monitoring.

- **Hypothesis 14**: Higher and lower span individuals will differ in terms of the mental processes they engage in when they plan.

3.5 Research design

In order to test the hypotheses aforementioned, the study employed a between-subject design, in which participants in the control group completed both first and second narrative tasks under a no-planning condition, and participants in the experimental group completed the first task under a no-planning and the second task under a planning condition. In this section, the data collection procedures will be described, followed by the description of the instruments of the study, measures of L2 speech production, and measures of working memory capacity. The research design is summarized in Tables 1 and 2.

Table 1
Data collection procedures for the experimental group

PHASE	***SETTING***
1. Control of proficiency level / task 1 (no-planning condition)	Whole groups/Language Lab
2. Speaking Span Test (SST)	Individually with the researcher /room
3. Planning (with verbal protocols) and performance of Task 2 (planning condition)	Individually with the researcher /room

Table 2
Data collection procedures for the control group

PHASE	***SETTING***
1. Control of proficiency level / task 1 (no-planning condition)	Whole groups/Language Lab
2. Speaking Span Test (SST)	Individually with the researcher /room
3. Performance of task 2 (no- planning condition)	Individually with the researcher /room

Data collection of the present study was divided in three phases as displayed in Tables 1 and 2. The first phase was the selection of participants which aimed at controlling for proficiency level. Participants performed the proficiency trial task at the language laboratory, and all students of the same class did the task together. Prior to task performance, students signed a consent form (Appendix I) and answered a biographical data questionnaire (Appendix II). Due to participants' time constraints, the task used for selecting participants also served as the first sample of L2 speech performance under no planning condition.

The second phase consisted of the Speaking Span Test (Appendix III). Participants of the experimental and control groups carried out the speaking span test individually in a room. First, I gave the instructions for the procedures of the test (see Appendix IV). I read the instructions aloud and the participants followed me, reading it silently. After the instructions, I clarified any doubts they had on the procedures of the test. Then, I carried out a training phase in order to help participants get familiar with the procedures of the test. The training phase consisted of a short version of the

Speaking Span Test containing 20 words organized in *one* set, which started with two words and finished with six words, in the same way the sets are organized in the actual span test. After the training phase, I checked whether participants still had any doubts about the procedures. The actual span test would only start when participants reported having no doubts about the test.

The third phase of data collection consisted of the second narrative task. Participants of the control group carried out the second narrative task under the same condition as the first narrative task, that is to say, under a no planning condition. On the other hand, participants of the experimental group carried out the second narrative task under a planning condition. Following Mehnert (1998), participants had 10 minutes to plan the second task prior to performance.

In order to document the processes learners engage in when they plan an oral task, retrospective on-line protocols were carried out during planning time. After time for planning was over, learners performed the second narrative task. After the performance of the second narrative task, an interview[12] was conducted for the purpose of complementing information of the retrospective online protocols. Having described the general research design, I turn now to the subsection on the context and participants, which will be followed by the procedures for the selection of participants.

[12] Interviews carried out after a complete process are classified as retrospective off-line protocols (Leow & Morgan-Short, 2004)

3.5.1 Participants and context

The participants of the present study were 50 intermediate learners from the Letras[13] Licenciatura, Letras Secretariado[14], and also from the Extracurricular Language Courses at the Federal University of Santa Catarina (UFSC). Participants were selected[15] from semesters 2 and 4 of Letras Licenciatura and Letras Secretariado Programs, and from semesters 7 and 8 of the Extracurricular Language courses. Participants of the Extracurricular Language courses were all undergraduate students at the Federal University of Santa Catarina from a variety of backgrounds (Biology, Engineering, Law, and History, among others). Out of the 50 participants, 30 were female, 20 were male, and their ages ranged between 18 and 29 years old, being an adult population.

The participants from the Letras Licenciatura Program had from 8 to 10 hours of English classes per week. *New Interchange II* by Jack Richards, Jonathan Hull and Susan Proctor, and *Passages I* by Jack Richards and Chuck Sandy are the course books adopted for the second and fourth semesters respectively. The participants from the Letras Secretariado Program had also from 8 to 10 hours of English classes per week. For these participants oral skills are developed specifically for business purposes. The course book adopted is *Business Class* by Cotton and Robbins. The participants from the Extracurricular Course, at semesters 7 and 8, had three hours of English per week focusing on the four skills, totaling a number of forty-five hours per semester. The course book adopted for both levels is *Passages I* by by Jack Richards and Chuck Sandy.

---

[13] Undergraduate Language Teaching and Literature program

[14] Undergraduate Bilingual Secretary program

[15] The procedures employed for the selection of participants will be described in detailed in the next section of the method.

Participants were invited to take part in the study, and no financial reward was given. First, I contacted the participants' teachers in order to explain the purposes of the study. After the teachers' permission, I visited the classrooms and invited the students to take part in the study. I told them there would be a first phase in order to select a homogeneous sample, and, after the first phase, some participants would be selected for the second phase of the study. At this first contact, I also asked students concerning their availability for taking part in the research. A few students reported having no time to take part in the study, but most of them reported being able to meet me twice for the purposes of the study. Students were encouraged by their teachers to participate in the study, but participation was voluntary in all phases. The only compensation for taking part in the study was the feedback on their performance, which was given by the present researcher in writing (see Appendix V for the feedback card).

3.5.2 Procedures for Selection of participants

The first phase of data collection of the study was the selection of participants, which was conducted using the rating scale proposed by D'Ely and Weissheimer (2005). The level of proficiency chosen for participation in the study was the intermediate level. The choice of the intermediate level was due to two reasons. First, the level of participants in most planning studies range from pre-intermediate to advanced levels (Foster & Skehan, 1996; Kauwachi, 2005; Mehnert, 1998; Ortega, 1999; Sangarun, 2005). Thus, choosing participants from these levels would allow for comparisons between my study and previous studies on planning.

Second, the intermediate level classes of the context of the current study (especially at the Extracurricular Language Courses) tend to have a larger number of

students than the advance level classes. Since the current study is quantitative and I needed a homogeneous sample in terms of proficiency level, it would be more feasible to select the amount of 50 participants at the intermediate level than at the advanced one.

In order to select the participants of the present study, I invited ninety-nine participants from three classes of semester 7 and two classes of semester 8 of the Extracurricular Language Courses, one class of semester 2 and one of semester 4 from Letras Licenciatura Program, and one class of semester 4 of Letras Secretariado Program. The choice of classes from these semesters was based on a previous study, carried by D'Ely (2006), which selected participants from the same semesters and reported that among these classes it was possible to select a considerable amount of intermediate learners for the purposes of a quantitative study.

Students were briefly told about the general purposes of the study (investigation of speaking skill), but they were not told which level I was interested in (intermediate level). I briefly told them that since my study was quantitative, I needed to have a homogeneous sample in terms of proficiency level. The proficiency trial took place from August 21st to September 26th , 2006, according to the days students attended classes and the days the language laboratory was available. All participants of the same class did the first task together in the language laboratory.

The first task consisted of a picture-cued narrative. Participants received the instructions in writing (see Appendix VI for Task 1 instructions). I read them aloud as they followed me silently. Participants were instructed to: (a) look at the set of pictures for fifty seconds, (b) put the pictures away when I signal that time was over, (c) tell a story about the pictures. Participants were also told that there were no restrictions as regards the time length for telling the story, there was no correct or incorrect sequence for the story and they were free to organize the pictures into a story they way they

wanted to, and in case they forgot a picture, they were also free to use their imagination to fill any gaps in the story.

The choice of 50 seconds for looking at the set of pictures aimed at minimizing pre-task planning as much as possible in the no planning condition. According to Mehnert (1998), one minute planning may be enough for gains in accuracy to take place. Thus, I gave participants less than one minute to look at the set of pictures.

After reading the instructions and checking whether participants had any questions about the procedures, I gave them the 50 seconds to look at the pictures. I said "turn the picture around now and look at it for 50 seconds, please. I will tell you when time is over". I used a chronometer to count the 50 seconds. When time was finished, I said, "Stop, put pictures away, do not look at them anymore, and start telling and recording your stories, please".

Participants' oral production was recorded on tapes then compiled into CDs using Sound Forge 6 Sophtware®. Participants' speech samples were given to three raters who were instructed to evaluate their performance according to the rating scale proposed by D'Ely and Weissheimer (2004) (see Appendix VII for the rating scale). According to D'Ely (2006), the scale was, in fact, an adaptation of the First Certificate in English speaking test assessment scale (Cambridge Examination), the Iwashita, McNamara and Elder's (2001) scale and the Royal Society of Arts (RSA) test (in Hughes, 1989). The scale is assessor-oriented (Luoma, 2004), that is to say, the rating scale adopts an analytical approach in order to provide detailed guidance to raters and help them make consistent rating decisions (D'Ely, 2006; Luoma, 2004;).

According to this scale, participants are assessed in terms of fluency, accuracy, and complexity on a scale from 0 to 5. According to D'Ely (2006), score 1 determines the criteria for the beginner level, score 3 determines the criteria for the intermediate

level and score 5 determines the criteria for the advanced level. There are also scores in between the three main levels, which according to D'Ely (2006), "allow for nuances of performance in between these levels" (p. 58). In other words, there is a range of three scores between 1 and 3, that is, 1.5, 2.0 and 2.5. The 1.5 score, for instance, shows that the speech sample contains more characteristics of the beginner level than of the intermediate one. Likewise, the 2.5 score shows that the speech sample has more characteristics of the intermediate rather than the beginner level. The 2.0 score shows that the speech sample presents some features of the beginner and intermediate levels in comparatively equal amounts. The same range of scores is present between 3 and 5, and the scores 3.5, 4.0, and 4.5 show that speech samples contain characteristics of performance in between the intermediate and advanced levels.

Since the target proficiency level for the present study was the intermediate, the learners to be selected were those who obtained a score of 3 (with a variation from -0.5 to +0.5) as a result of the average score of the sum of the scores in each of the descriptors (accuracy, complexity, and fluency). According to D'Ely (2006), scores ranging between 2.5 and 3.5 are the ones which show that the participants' speech samples contain more features of an intermediate level than features of either the beginner or advanced level.

As regards the raters, they were all experienced teachers of English who had been trained and had used the scale previously in the pilot study by Guará-Tavares (2006). At the time of data collection, one of the raters was an MA student and the other two raters were PhD students in the graduate program in Letras/ Inglês at the Federal University of Santa Catarina. Raters had approximately a week to assess participants' performance.

In order to estimate the degree of interrater reliability, the statistical procedure selected was the Cronbach Alpha coefficient of reliability found in Statistical Package for Social Science (SPSS)®. The Cronbach Alpha Coefficient is widely used in reliability analyses (Field, 2005). It allows finding the degree of interrater reliability, the means and the standard deviation of participants' performance. Reliability estimates (see Appendix XIX for statistics on Cronbach Alpha analysis) for the rating procedure were .84, which is considered good level of reliability (Field, 2005).

As can be seen in Figure 1, the group means was 2. 95. According to the rating scale proposed by D'Ely and Weissheimer (2004), this score (2. 95) is almost at the score that is considered intermediate (3.0).

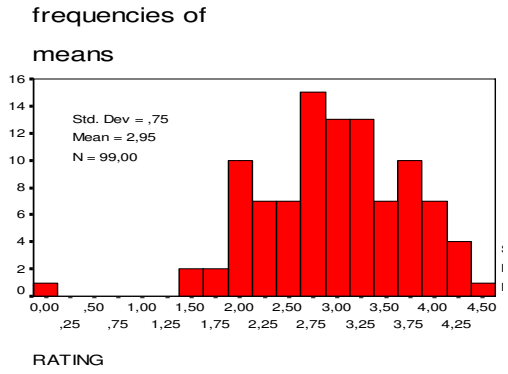

Figure 1 **-** Selection of participants –rating procedure

In order to select participants who would present features of an intermediate level speech performance, students whose means varied from 2.54 to 3.5 were selected for the present study. Therefore, out of the 99 students who took part in the proficiency trial, 55 were selected for the present study. Out of the fifty-five participants selected, five did not participate in the study. One of them did not want to take part, two of them accepted to participate but did not show up for data collection, and two of them I was

not able to find after trying to contact them either by phone or email several times. Therefore, out of the 55 participants selected, 50 participated in the present study.

After participants were selected, they were randomly divided into a control group and an experimental group. Due to their time constraints and for the sake of practicality, the same narrative task performance used for selecting participants was also used as the first sample of speech performance under the no planning condition for both control and experimental groups. As previously mentioned, participants had 50 seconds to look at a set of nine pictures. Then, they were required to put the pictures away, and start telling their stories immediately. After the first task was carried out, participants answered a questionnaire on their perceptions about the task (see Appendix VIII for the questionnaire after Task1). Having explained the procedures for the selection of participants, I will turn now to the instruments of data collection.3.6 Instruments

3.6.1 Materials and equipment

As previously explained, the experiment consisted of three tasks: one task aimed at measuring working memory capacity and two narrative tasks aimed at eliciting speech production in the L2. The working memory task was conducted using an ACER 3620 laptop computer. The software SOUND FORGE 6® was used to record participants' responses in the working memory task and in the narrative tasks. The software Praat® was used to analyze speech pauses.

3.6.2 Questionnaires

When participants came to the language laboratory to perform the proficiency trial, they answered a biographical data questionnaire (Appendix II) and signed a consent (Appendix I) form prior to task performance. The biographical questionnaire consisted of three questions: (a) one concerning participants' age and gender, (b) one concerning their university courses, and (c) one concerning the length of time studying English.

After the performance of Task 1, participants answered a questionnaire on their perceptions of the task (Appendix VIII). The questionnaire consisted of three questions. The first question concerned the degree of difficulty of the task for them, and the second and third concerned their procedures while looking at the pictures for 50 seconds. The aim of this questionnaire was to have an overall idea of how students perceived the task under no planning conditions, and to check whether the time of 50 seconds for looking at the pictures was brief enough to avoid strategic planning as much as possible.

3.6.3 The Speaking Span Test

In the second phase of data collection, the Speaking Span Test (hereafter SST) was administered to measure participants' working memory capacity. Due to participants' time constraints, only one test was used to measure working memory capacity. The Speaking Span Test was chosen because it has been previously used in seminal studies on the relationship between working memory capacity and speech production in L1 (Daneman, 1991; Daneman & Green, 1986), and in L2 (Fortkamp, 1999; 2000).

Weissheimer's (2006) version of Daneman and Green (1986) and Daneman (1991) SST was used in the present study. This version of the SST had been previously piloted with intermediate Brazilian learners in studies carried out by Weissheimer (2006, 2007) and Guará-Tavares (2006).

The test contained 60 unrelated words organized in three sets. According to Weissheimer (2006), the criteria and procedures for the selection of words that composed the SST were: (a) words should be familiar to all participants and therefore were selected from intermediate level course books; (b) the high frequency levels of the words selected to constitute the test were attested by checking two specialized websites[16], (c) only monosyllabic words were included; (d) words semantically and phonetically related were avoided within each sequence in order to prevent participants from making associations between words, which could assist memorization and, thus, influence performance on the SST.

In each set, the number of words increased progressively from two to six. Each word was presented one at a time on a computer monitor for one second. After the last word in each set disappeared from the computer screen, participants viewed question marks on the screen. The number of question marks corresponded to the number of sentences to be produced. As the following examples from Set 1 of the test that show the sequencing:

[16]Sites:http://www.paulnoll.com/China/Teach/English-3000-common-words.html,and http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists

Figure 2 - SST examples from Set 1

Participants were instructed to read each word aloud. This reading aloud procedure differed from the procedures in Daneman and Green (1986) and Daneman (1991), and it was meant to avoid participants reading only some of the words in each set. The decision for having participants read the words aloud was taken after the pilot study (Guará-Tavares, 2006) when some participants reported having read only some words in each set to make it easier to remember them. The use of idiosyncratic strategies may blur the relationship between performance on working memory span tests and performance on complex cognitive tasks (Friedman & Miyake, 2004).

By having participants read the words aloud, I could make sure all of them were reading all the words in the test. This procedure may have aided their memorization of the words since vocalization is a retention strategy (Fortkamp, 2007, personal communication). However, I believe any possible effect of vocalization was minimized since *all* participants read aloud.

Participants were instructed to use the words in the exact form and order they appeared on the screen to generate syntactically and semantically acceptable sentences,

aloud, in English. There were no restrictions concerning the length or complexity of the sentences. For instance, after being presented a set of three words: guy point train, a participant produced the following sentences:

"I am a guy"

"What's your point?"

"The train was dirty"

Figure 3 - SST production sample

Following Daneman (1991), Daneman, and Green (1986) and Fortkamp (1999), participants' responses, which were recorded, transcribed, and analyzed, generated two different speaking span scores: (a) a speaking span *strict score*, when all the sentences the subject produced contained the target word in the exact form and order of presentation, and (b) a speaking span *lenient score*, when credit (1.0 point) was given for sentences that contained the target word in a form other than that of presentation (e.g., target word being 'guy' and the word in the produced sentence being 'guys').

Weissheimer (2006, 2007) claims that, when no credit is attributed to words recalled in a different order, participants who recall words in a different order and produce correct sentences with these words are equated with participants who recall no words at all. Thus, in order to grasp individual differences more effectively in the performance of the span test, half credit (0.5) was given to words recalled in a different order. This procedure was adopted following Weissheimer (2006, 2007). No credit was given to ungrammatical sentences in terms of syntax and semantics.

The following excerpts from my data illustrate the procedures for obtaining the speaking span test *strict and lenient* scores. First, I counted all instances in which words were recalled in the exact order and form of presentation, and grammatically and semantically sentences were produced, and attributed one point (1), as illustrated in the following figure.

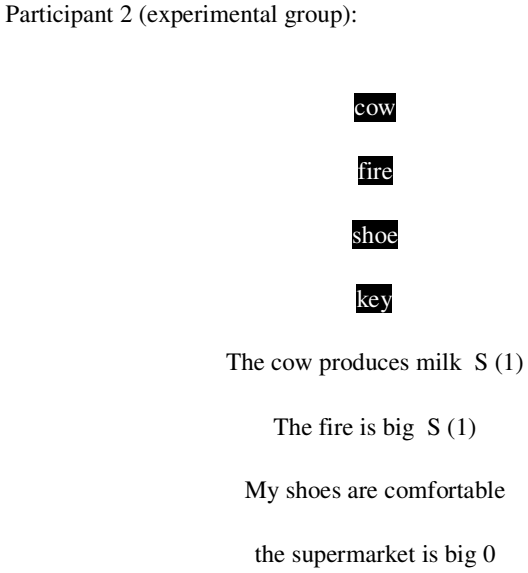

Figure 4 - SST scoring

After assigning the points relative to the strict score- S (1), I read the sets a second time and added 0.5 point for words recalled in a different order of presentation-L and 1 point to words recalled in a different form of presentation-L (1), as the following figure shows:

The cow produces milk S (1)

The fire is big S (1)

My shoes are comfortable L (1.0) word recalled in a different form of presentation

Figure 5 - SST scoring

As for the lenient scores, all points were counted whereas for the strict scores only the points yielded by strict scoring were counted. In excerpt 3, the partial strict and lenient scores of the participant were 2 and 3 respectively.

In addition to the lenient and strict scores, the time participants took to perform the Speaking Span Test was also calculated. This methodological procedure was adopted based on the findings of Friedman and Miyake (2004). These researchers verified through three experiments that differences in the time taken to perform span tests may lead to strategy use which may blur results regarding the relationship between working memory capacity and the task under performance, oral narratives in the present study.

It is important to highlight that investigating any effects of time on the overall span performance or on L2 speech performance is beyond the scope of the present study. Response time was calculated *only* for the purpose of checking whether the control and experimental groups were homogeneous in terms of the time taken to perform the test. This procedure will be further explained in the data analysis section of the current chapter.

3.6.4 The speech generation tasks: 'there-and-then' narratives

The two tasks were both 'there- and- then' picture cued narratives (Robinson, 1995). In 'there-and-then' tasks, participants have no access to the visual stimuli of the task while telling the stories. In both tasks, participants had fifty seconds to look at the set of pictures and then put the picture away. The choice of 'there-and-then' narrative tasks for the present study was based on the following criteria:

1. Monologic narratives, claims Ejzenberg (1992), are more efficient in assessing oral ability than dialogue tasks.
2. Narrative tasks are considered to be cognitively demanding, thus, the effects of pre-task planning are more likely to be revealed (Foster & Skehan,1996; Skehan & Foster, 1995, 1997).
3. According to Robinson (1995),'there-and-then' tasks are more complex than 'here- and-now', since in the former, the participants are not allowed access to the visual stimuli of the tasks during performance. In order for individual differences in working memory capacity to emerge, the task performed has to be complex (Fortkamp, 2000; Just & Carpenter, 1992; Tomitch, 1996).
4. Following Skehan and Foster (1999), the sequence of the events in the narrative tasks of the present study was not completely clear and/or predictable. Several sequences of events would be possible; participants were supposed to organize the events in order. This relative openness of possibilities seem to turn the task into a more complex one when compared to a task which has a completely clear and predictable sequence of events.

5. According to Elder and Iwashita (2005), narrative tasks based on a set of pictures are widely used in the Test of Spoken English (TSE).
6. Narrative tasks have been widely used in previous studies on task based planning (D'Ely, 2006; Ellis, 1987; Ellis & Yuan, 2005; Foster & Skehan, 1996, 1999; Kawauchi, 2005; Ortega, 1999; Wendel, 1997), thus, allowing for comparison between the present study and previous ones in the field.
7. The tasks used in the present study were previously piloted (Guará-Tavares, 2005, 2006; Weissheimer, 2005, 2007) and showed to be feasible to be performed by *intermediate* learners, which is the proficiency level of the participants in the present study.

One of the picture cued narratives used in the present study displayed a series of pictures of a couple at a restaurant (Appendix IX). During the meal the man kept imagining things he would like to do to the woman. The other picture cued narrative (Appendix X) displayed a series of pictures of a couple in a living room. The man kept giving several gifts to a woman who seemed to refuse all of them. As previously stated, there was no fixed order of events, participants were told to look at the series of pictures and organize them into a sequence in order to tell a story.

The order of tasks was counterbalanced among participants for the purpose of controlling practice effects. In other words, half of the participants carried out Task 1 as their first task (no-planning condition for both control and experimental groups) and Task 2 as their second task (planning condition for experimental group and no planning condition for control group). The other half of the participants carried out the opposite procedure; they performed Task 2 as their first task and Task 1 as their second task.

3.6.5 The verbal reports: retrospective online protocols and retrospective interviews

Following retrospective online procedures (Leow & Morgan-Short, 2004), which were reviewed in Chapter 2, participants were given 10 minutes to plan and were required to verbalize what they were planning in breaks of every one minute. After every one minute of planning they were prompted with the question: '*What were you just thinking about?'* However, in some moments when participants stopped taking notes and seemed to be thinking hard or when they erased part of their notes, the present researcher asked different questions: "*What were you just thinking when you stopped writing?"* or "*What did you just erase from your notes*?" The 10-minute-planning time was counted with the aid of a chronometer which was stopped during the verbalization so that participants could have 10 minutes of actual planning. Also, the instances of verbalization were made the shortest possible so that they would not take participants away from the planning task itself. Basically, I asked the question and accepted whatever answer they gave me and instructed them to go on planning. In general, each verbalization was no longer than 40 seconds (including my question and the answer). The following figure illustrates a segment of the retrospective online protocol of participant 9:

R: What were you just thinking about?

P: I'm trying to put the things… the pictures together, everything together
R: What were you just thinking about?
P: about how could I use the grammar in the right way…and thinking about the tenses of the verbs… about the right words to say what I want to say
R: What were you just thinking about?
P: about grammar… and for example I don't know how to say 'garrafa' and the past of think is thought?

Figure 6 **-** Retrospective online protocol sample

It is important to highlight that all participants took a training session on these procedures and only started the verbal protocol itself when they had no doubts or questions about the procedures. The training session consisted of participants planning a narrative task (see Appendix XII for training session task) for three minutes. After every period of one minute, I prompted participants with the question: *What were you just thinking about?* A brief three minute training section was designed in order to prevent any fatigue effect and due to participants' time constraints. Training sessions are suggested by Ericsson and Simon (1993) and Leow and Morgan-Short (2004).

The retrospective online protocols were carried out in English. The decision for conducting the protocols in English was made after the pilot study (Guará-Tavares, 2006). The first two participants of the pilot study reported that it was hard for them to plan the task in English and then having to switch to Portuguese during the protocols. I decided to let the participants of the pilot study choose whether to respond to the protocols in English or Portuguese. All of them responded in English. Therefore, for the present study, I told participants they could answer the protocols in English or Portuguese, and, again, all of them responded in English. They did use Portuguese at some moments when they were searching for lexical items. The following figure illustrates one of these moments:

Researcher: What were you just thinking about?

Participant: I'm trying to find the adjective like 'uma pessoa chata que enche o saco'

Figure 7 - Retrospective online protocol sample

After the retrospective on-line protocols, participants performed Task 2. I left the room so that participants would be comfortable to tell their stories. After performance of

Task 2, an interview (Appendix XIII) was also carried out for the purpose of complementing the retrospective on-line protocols. The interview contained questions concerning perceptions of the task under planning conditions and learners' processes during planning. This combination of protocols is suggested by Leow and Morgan-Short (2004) and Ericsson and Simon (1993)[17].

Participants of the control group also answered to an interview after performance of Task 2, which was carried out under a no planning condition. The aim of the interview was to check whether Tasks 1 and 2 (both under no planning conditions) were equivalent in terms of difficulty for learners, and also whether the 50- second time for looking at the pictures was actually not enough for carrying out any sort of planning.

3.7 Measures of L2 speech performance

After the speech samples were collected, compiled into CDs, and fully transcribed, they were analyzed in terms of fluency, accuracy, and complexity. These measures have been extensively used in studies investigating the effects of planning on L2 speech performance (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1995, 2005; Yuan & Ellis, 2003, among others), and according to Fortkamp (2000), these measures "seem to give a global view of L2 speech performance since they are intended to capture complementary aspects of this multidimensional process" (p.87).

[17] See subsection 2.4.4 on Retrospective online protocols.

3.7.1 Fluency

In the present study fluency is conceptualized as a temporal phenomenon, which reflects the ability to cope with communication in real time (Skehan, 1996, 1998). Since fluency is a multifaceted concept (Tavakoli & Skehan, 2005), two temporal dimensions were used: (a) speed fluency (Freed, 2000) and (b) breakdown fluency (Skehan, 2003). Speed fluency concerns the speed in which speech is delivered and was assessed by means of two versions of speech rate –unpruned and pruned (Fortkamp, 2000; Lennon, 1990; Ortega, 1999).

- Speech rate unpruned and pruned- According to Freed (2000), speech rate regards how fast the language produced is. Following Fortkamp (2000), **speech rate unpruned** was calculated by dividing the total number of semantic units (complete and partial words) produced by the total time in seconds (including pause time), the resulting figure was multiplied by 60 to express the number of semantic units per minute; **speech rate pruned** was calculated in the same way but excluding: (a) the words that were abandoned before completion, and (b) words that were immediately repeated (except words repeated for rhetorical purposes).

Breakdown fluency, in turn, concerns silence which may lead to features of disfluency such as pauses and hesitations (Freed, 2000; Skehan, 2003), and was assessed by means of number of silent pauses per c-unit (D'Ely, 2006) and percentage of unfilled pausing time (Foster & Skehan, 1996; Lennon, 1990).

- Number of silent pauses per c-unit - this measure was operationalized by dividing the number of silent pauses in each subject's speech sample by the

number of c-units, as in D'Ely (2006). Following Foster and Skehan (1996), Mehnert (1998), D'Ely (2006), and Weissheimer (2007), a cut-off point of 1 second was considered optimal in determining silent pauses in L2 speech samples. Pauses were identified and measured using the computer software PRAAT® 4606. This software provides the precise location and length of speech pauses.

- Percentage of total silent pausing time- this measure was calculated by dividing the total silent pausing time by the total time participants took to complete the task, the resulting figure was multiplied by 100 (D'Ely, 2006; Foster & Skehan, 1996; Lennon, 1990).

3.7.2 Accuracy

According to Skehan (1996, 1998), accuracy is related to 'a learner's belief in norms' and, thus, concerns form in the sense of error-free performance. When assessing language acquisition, claims Ellis (1987, 2005a), specific measures of accuracy (such as tense morphemes or plural –s) seem to be more appropriate. However, when assessing performance, as in the present study, researchers claim for a more general approach to accuracy (Foster & Skehan, 1996; Mehnert, 1998). Therefore, two general measures of accuracy were used to assess participants' speech performance in the present study:

- Number of errors[18] per a hundred words- this measure was calculated by dividing participants' total number of errors by the total number of words produced and multiplying the result by 100 (Fortkamp, 2000; Mehnert, 1998).

[18] The criteria for defining errors was based on American English norms since this is the norm of the text books used by participants.

- Percentage of error free clauses- this measure was calculated by identifying the number of error free clauses, which was then divided by the total number of clauses produced, and the resulting figure was multiplied by 100 (Foster & Skehan, 1996; Mehnert, 1998).

3.7.3 Complexity

According to Skehan (1996, 1998), complexity, similarly to accuracy, concerns form but it is related to the utilization of more elaborated and structured language which emerges as a result of the willingness for risk taking from the part of the learners. According to Foster and Skehan (1996), subordination is considered a satisfactory measure to assess complexity. Quirk and Greenbaum (1973) define subordination as "a non-symmetrical relation, holding between two clauses in such a way that one is a constituent part of the other (p.309).

In the present study, complexity was measured by an index of subordination reflected by the number of clauses per c-unit. It was calculated by dividing the total number of clauses (dependent and independent) by the total number of c-units. The higher the index of subordination obtained the higher the complexity of the speech was.

Following Foster, Tonkin, and Wigglesworth (2000), a clause will be considered subordinate when it consists "minimally of a finite or non finite verb element plus at least one other clause element (subjects, objects, complement or adverbial)" (p. 326). According to Foster and Skehan (1996), a c-unit is defined as "each independent utterance providing referential or pragmatic meaning consisting of one single independent finite clause or else and independent finite clause plus one or more dependent finite or non finite clauses" (p. 310).

When analyzing c-units in the present study, I followed Foster et al. (2000) criteria. Utterances that were abandoned were not counted as a unit; phrases or full clauses that were repeated verbatim were counted once, with only one instance being considered as either a c-unit or belonging to a c-unit; verbatim repetition of words including those used for rhetorical purposes were considered as parts of the c-unit they belonged to; and whenever self-corrections took place, only the final version was counted as belonging to the c-unit.

3.8 Procedures for data transcription

The participants' speech samples were transcribed verbatim and fully analyzed. As for the transcription conventions, repeated words were underlined, silent pauses were indicated by parenthesis containing the time length within them and placed at the precise location at the speech sample. For instance, (1.5) indicates an unfilled of 1 second and 5 hundred milliseconds. Only pauses from 1 second or longer were included in the analysis. Grammatical and lexical errors were marked in bold. When analyzing complexity, the following transcription conventions were used: (a) C: clause, (b)1st C: first clause, (c) 2nd C: second clause, and so on; (d) I: independent, (e) D: dependent, (f) F: finite, (g) NF: non finite. The following figure illustrates the data transcription procedures:

(7.509) this is a story about a man (1.263) who is trying to **conquest** a woman that he loves (1ST C: I, 2ND C: DF, 3RD C: DF- 3 CLAUSES- 1 C-UNIT) (3.189) he's trying to conquest **conquest** her giving her gifts like expensive clothes (1.362) like a a ring and a necklace (1.296) but she doesn't care about about him (4TH C: I, 5TH C: DNF, 6TH C: I- 3 CLAUSES- 2 C-UNITS) (1.063) and (1.030) he really is trying but she doesn't care (7TH C:I, 8TH C:I – 2 CLUSES-2C-UNITS) and at the end he stops giving her presents or trying to **conquest** her and **decide** to go away in his beautiful car (1.163) with another woman (1.130) that I think is really more beautiful (1.362) (9TH C: I, 10TH C: DNF, 11TH C: I, 12TH C: DF, 13TH C; DF- 5 CLAUSES- 2 C-UNITS) so he **lets** her the hard woman that he was trying to **conquest** he **lets** her (1.761) behind going away with another woman (14TH C: I, 15TH C: DF, 16TH C: I, 17TH C: DNF- 4 CLAUSES- 2 C-UNITS)

Figure 8 - Procedures for data transcriptions

As shown in figure 8, the repeated words were underlined, the errors were in bold, and pause length was indicated in parenthesis at the pause location. Independent, dependent, finite, and non finite clauses were indicated by C, I, D, F, and NF, respectively.

3.9 Procedures for data Analysis

3.9.1 Analysis of the quantitative data

The quantitative data was submitted to statistical treatment. The first step was to verify whether the measures used to assess performance (fluency, accuracy, and complexity) actually underlie different constructs of L2 speech. For that purpose, a Principal Component Analysis was carried out. A Principal Component analysis is concerned with establishing which linear components exist within the data (Field, 2005)[19]. In studies on task-based planning, it is frequently used for the purpose of checking whether the measures of performance (e.g., fluency, accuracy, and complexity) actually underlie distinct constructs. Prior to the analysis, the suitability of the data for factor analysis was examined. The Kaiser-Meyer-Oklin value was .61; the recommended value is .6. The Barlett's Test of sphericity reached statistical significance .000.

The second step was to carry out descriptive statistics analyses in order to give an overview of the seven variables of speech production, working memory scores lenient and strict, and response time. Descriptive statistics provide the minimum, the

[19] See Field (2005) for the complete theory behind Principal Component Analysis.

maximum, and the mean values of general results in each of the measures previously mentioned, as well as the standard deviation for each group (control and experimental).

In the next step, the normal distribution of each group on all variables was tested by examining skewness and kurtosis. Then, Pearson Product Moment correlations were used to verify whether there were correlations between speech performance scores and working memory capacity scores. When the size of the population is relatively small (N = 25 for control group and N = 25 for experimental group, in the case of the present study), Spearman correlations can be used (Hatch & Lazaraton, 1993; Ellis, 2007, personal communication). Therefore, Spearman correlations were also computed as an attempt to further scrutinize the data. Since the Spearman correlations yielded results similar to the Pearson Product moment correlations, some of the results of the Spearman correlations were included in the appendixes only (Appendix XIV). The Results and Discussion chapters were based on the Pearson Product Moment correlations which are the standard procedures for normally distributed populations.

The next step was to verify whether planning led to significant differences in the performance of the experimental group when compared to the control group, and one-way ANOVAs were used for this purpose. The one-way ANOVA procedure produces a one-way analysis of variance for a quantitative dependent variable (the different measures for fluency, accuracy, and complexity) - by a single factor (independent variable - the different conditions – no planning for the control group and planning for the experimental group).

First, ANOVAs were run to compare the control and experimental groups in terms of a) performance of the *first* task, and (b) lenient and strict scores on the SST and c) time taken to perform the SST. These procedures were followed to verify whether the groups were homogenous in Task 1 performance and working memory capacity

scores so that any group differences in the performance of the *second* task could be attributed to pre-task planning and not to group differences in terms of speech performance a priori or to group differences in terms of memory or time to perform the memory test. Then, ANOVAs were run to compare the control and experimental groups in the second task in order to check whether planning led to significant differences in the performance of the experimental group when compared to the performance of the control group in task 2. Effect sizes were also calculated in the attempt to verify the magnitude of the effects of planning on L2 speech performance. Effect sizes were calculated using the formula by Cohen (1988, as cited in Norris and Ortega, 2000), (see Appendix VI for the effect size formula).

Following Conway et al. (2005) and Weissheimer (2007), the present study adopted an extreme group design in the attempt to scrutinize differences between lower and higher spans individuals more precisely. According to Conway et al., "extreme-group designs refer to contexts in which a continuous variable is categorized, and only the lower and upper ends of this variable distribution are represented" (p.782). Although extreme-group designs present problems, they are common in the working memory literature, and they may be useful in the attempt to scrutinize differences between lower and higher spans individuals (Conway et al.).

The first problem with extreme-group designs is that information is lost, since only the extremes of the population are included in the analysis. Second, they tend to overestimate effect sizes (Conway et al., 2005). One advantage in using extreme-group designs is that individuals are hardly ever misclassified as lower or higher spans since only the extremes are used. Moreover, it allows further scrutiny of differences between higher and lower spans (Conway et al.).

The most common type of extreme-group design is based on quartiles; however, tertiles can also be used when data samples are small (Conway et al., 2005). In order to conduct the extreme group design, the cut off point was established between two percentiles: 33,3% and 66,6%. Based on these percentiles, participants were categorized as having higher working memory span when they fell in the upper tertile (the ones above 66,6%), and lower working memory span when they fell in the lower tertile (below 33,3%). Out of the 25 participants who belonged to the experimental group, 8 were classified as lower spans, and 8t were classified ad higher spans. The remaining 9 participants were classified as intermediate spans and were not included in the analysis which focused specifically on comparing lower and higher spans.

In order to verify whether higher spans outperform lower spans in planning performance, one way ANOVAs were run for the measures of fluency, accuracy and complexity. Having reported the procedures for the analysis of quantitative data, I turn now to the procedures for analysis of the qualitative data.

3.9.2 Analysis of the qualitative data

As mentioned in the Review of Literature chapter, Ortega (2005) based the coding of her interviews on the taxonomies proposed by O'Malley and Chamot (1990) and Oxford (1990). The only purpose of Ortega's (2005) study was to examine learners' processes during planning. The present study has purposes other than investigating learners' processes during planning (examining the impact of planning and its relationship with working memory capacity); thus, for the sake of simplification only the framework of O' Malley and Chamot (1990) was adopted.

The analysis of the protocols consisted of three phases. For the first phase I had a first interrater. First, a content analysis of the protocols was carried out individually by the present researcher and the first interrater. This content analysis consisted of going through the protocols and writing down our first general impressions on them. We focused mainly whether (a) participants were focusing on form or meaning, (b) instances which were repeated in each protocol, (c) instances which were common among protocols.

After this first analysis, we carried out a second individual analysis in order to classify learners' protocols into strategies using the framework by O' Malley and Chamot (1990). When this analysis was over, we got together in order to compare our findings concerning our strategy classification. Whenever there was disagreement between our classifications, we discussed the samples of the protocols in order to reach consensus.

At the end of the discussion, there were still a few instances of the protocols in which we could not reach consensus because some answers were general and we had difficulties to associate them with strategies from the framework of O' Malley and

Chamot (1990). In the attempt to solve these shortcomings, I contacted Professor Lourdes Ortega from the University of Hawaii to ask for help. As shown in the Review of Literature, Professor Lourdes Ortega carried out the seminal studies on the processes learners engage during pre-task planning. She promptly replied the email message with clarifying answers. According to her answers to my questions, in order to be classified as a strategy, a comment from the protocol would have to be as concrete as possible and general comments on content should not be classified as strategies (Ortega, 2007, personal communication).

The interrater and I went over Ortega's answers to my questions in order to reach a final consensus on the analysis of the protocols. Together, we classified 15 strategies, which will be fully reported and discussed in Chapters IV and V. To illustrate, the following excerpts are instances of general comments about content throughout the protocols which were not classified as strategies:

Excerpts

"I'm thinking that the man is angry with his wife because he thought she lied to him so he tried to hit her with the lamp" (p21)

"I'm thinking of what the woman said to the man when he brought gifts" (p20)

Instances like the ones in the excerpts were classified as general comments (focusing on content) because they did not mention anything specifically or concretely related to language use or behavior. These instances illustrate general comments or thoughts about content of the narratives. Although learners could be thinking about more concrete or specific language problems they did not mention it in an explicit way which was concrete enough to be classified as a strategy.

The following excerpts, on the other hand, illustrate instances of more concrete comments throughout the protocols which were classified as strategies:

Excerpts

> "A mistake in my grammar, I wrote -didn't should- and should is a modal and I don't need to use the didn't" (p07)
>
> "I'm reading the story to remember it when I tell" (p11)

As it can be seen in Excerpts, participants' comments are more concrete. The first comment illustrates a self-correction, which was classified as monitoring. The second comment illustrates practice of the story through reading for the purpose of memorization, which was classified as rehearsal. After the first interrater and I finished the analysis of the protocols, I submitted 20% of the protocol data to a second interrater for the purpose of reliability.

After the qualitative analysis, protocol data were also submitted to statistical treatment. First, I counted the types of strategies employed by each participant. The counting was based on the different types of strategies reported by learners. When learners mentioned the same strategy twice or more, only one instance of the strategy was counted. This procedure was adopted because when participants repeated a strategy, it was hard to tell if they were referring to the exact same use of a strategy previously mentioned or to a different use of the same strategy. Thus, we counted one instance of each strategy verbalized in order to prevent the same instance of strategy use to be counted twice. Second, descriptive statistics analysis of the strategies reported by the participants was also carried out in the attempt to have an overall view of the strategies employed by higher and lower spans.

Third, an independent *t*-test was performed to verify whether there were any statistically significant differences in the strategies employed by lower and higher spans. In order to scrutinize the differences between lower and higher spans, I also followed

the extreme- group design procedures and excluded intermediate spans from the analysis.

For all statistical analyses, a probability level of $p < .05$ was used to determine statistical significance. Having described the data analysis procedures, I turn now to the reliability procedures.

3.9.3 Reliability analysis

3.9.3.1 Reliability of the analysis of speech performance measures

The procedures for reliability analysis varied according to time availability of the raters. Following Ellis and Yuan (2005), Pearson Product Moment correlations were performed to measure the relationship between my analyses and the raters' analyses.

Five raters reanalyzed different portions of the data. Rater 1, who holds a PhD from the Letras/Inglês Program at Universidade Federal de Santa Catarina, reanalyzed 20% of the data for accuracy (number of errors per a hundred words). Pearson Product Moment correlations were performed to check the strength of the relationship between our analyses of accuracy. There were positive significant correlations between our accuracy scores ($r = .910^{**}$, $p = 000$). These correlations indicate a strong relationship between our analyses of accuracy.

Rater 2 holds a PhD degree from the same graduate program. She reanalyzed 100% of the protocols and we discussed our findings in order to reach consensus. After this analysis, 24% of the protocols were submitted to Rater 3. She holds a Master degree from the same program. Her MA study was about learning strategies and she was familiar with the O' Malley and Chamot's (1990) framework.

It is important to highlight that 25% of the protocol data corresponds to protocols by six participants. Each participant produced ten verbalizations throughout the protocols; thus, six participants produced a total of sixty verbalizations. Out of the sixty verbalizations, there was agreement in fifty-four of them when comparing the analysis of rater 4 and the previous analysis carried out by rater 3 and the present researcher. Therefore, the final consensus on the protocols reached 90% agreement.

Rater 4 recalculated 100% of total pausing time analysis. Rater 4 holds a Bachelor's degree in Education from the State University of Santa Catarina. He added all pauses throughout the narratives using a calculator. Whenever there was disagreement, we recalculated the pauses until we reached the same results. Therefore, we were able to reach 100% agreement on total pausing time.

I could not find anyone available for reanalyzing speech rates unpruned and pruned, pauses, and complexity. I made the following decisions concerning the reliability analysis of these variables. The analysis of speech rate unpruned is an extremely simple, straightforward one. All it one has to do is count the total words spoken and divide by the total time spoken. The total time taken to speak was counted using SOUND FORGE 6® software and the total words were counted using the tool word counting of the computer. The analysis of speech rate unpruned is unlikely to present any problems, thus, I did not reanalyze any portions of this data.

The analysis of speech rate pruned is also a straightforward one. The procedures are very similar to the analysis of speech rate unpruned. The only difference is that the words that are abandoned before completion and words that are immediately repeated were excluded from the analysis (except words repeated for rhetorical purposes). Therefore, analysis of speech rate unpruned is also unlikely to yield any problems which may affect results.

The analysis of pauses was carried out with the aid of PRAAT® software which provides the precise location and length of pauses. Although the analysis of pauses is simple with the aid of the software, it is extremely time consuming. Due to my time constraints for data analysis, interpretation, and writing up of the dissertation, I could not reanalyze any portions of pauses.

As for the analysis of complexity, although I did not have any rater available to actually reanalyze any portions of the data, I had an interlocutor with whom I shared some of my doubts throughout the analysis. This interlocutor was also the rater for the first analysis of protocols. Whenever I had any doubts concerning the analysis of complexity, I sent her part of the speech samples and we exchanged ideas on the data in order to reach consensus.

After my analysis of complexity was over, I reanalyzed 20% of the data myself including five participants from the control group and five participants from the experimental group. Pearson Product Moment correlations were performed to check the strength of the relationship between the two analyses I carried out for 20% of the data on complexity. There was a positive significant correlation between the two analyses (r =.945**, p = 000), indicating a strong relationship between them.

In brief, the only measures for which I had no portions of data reanalyzed were speech rate unpruned and pruned, and pauses. I could not find anyone available to reanalyze this data and due to time constraints I could not reanalyze it myself. However, speech rates unpruned and pruned are simple and straightforward measures to analyze, thus, unlikely to yield any problems. As for pauses, I had the aid of a computer software, which allows the analysis to take place without major problems.

3.9.3.2 Reliability analysis of the Speaking Span Test

Rater 5, who is pursuing her PhD in Letras/Inglês at Universidade Federal de Santa Catarina, reanalyzed 100% of the data for the Speaking Span Test scoring. Pearson Product Moment correlations were performed to check the strength of the relationship between our analyses of the SST. There were positive significant correlations for the strict scores (r =.997**, p = 000) and also positive significant correlations for the lenient scores (r = 990**, p = 000). These correlations indicate a strong relationship between our scoring of the speaking span test. This analysis rater 5 carried out was conducted for the purpose of *inter-reliability* of the Speaking Span Test.

In addition to the procedures of *inter-reliability*, procedures of *intra-reliability* were also performed for the test. *Intra-reliability* estimates were calculated using Cronbach's Alpha coefficient formula for examining internal consistency, also used by Turner and Engle (1989), Engle et al. (1992), Fortkamp (2000), and Weissheimer (2007). Internal consistency measures the extent to which different parts of a test or different items of a test measure the same construct (Field, 2005).

As previously explained, the speaking span test consists of three sets of words which go from two to six words. For each set, the correct sentences were computed (for both strict and lenient scores) as one partial span. For instance, participant 3 of the experimental group obtained partial lenient scores of 9 in the first set, 5 in the second set and 8 in the third set, obtaining a total lenient score of 22. The same participant obtained partial strict scores of 9 in the first set, 4 in the second set and 7 in the third set, obtaining a total strict score of 20.

For the analysis of internal consistency, the partial scores of each set were used to compute the Cronbach Alpha Coefficient in order to check whether the sets measured

the same construct. Reliability estimates for internal consistency were .84 for both strict and lenient scores, which is considered a good degree of reliability (Field, 2005).

Still concerning the reliability of the Speaking Span Test scores, Pearson Correlations were also performed to check the relationship between participants' strict and lenient scores. Significant positive correlations (r = .975**, p = 000) between strict and lenient scores indicated that lower spans in the strict scores were also lower spans in the lenient scores, and higher spans in the strict scores were also higher spans in the lenient scores.

3.10 The Pilot study

I carried out a pilot study on the relationship between pre-task planning, working memory capacity and L2 speech performance from February to June 2006 in order to inform my methodological decisions and pilot the narrative tasks, the Speaking Span Test, and the verbal protocols used in the present study. Although the pilot study was mainly quantitative, there was also an attempt to scrutinize what learners do when they plan.

The pilot study also employed a between subject design, in which participants in the control group completed both the first and second narrative tasks under a no-planning condition. The participants in the experimental group, in turn, completed the first task under a no-planning and the second task under a planning condition.

The participants in the study were 25 Brazilian adult learners of English at the Extracurricular Language Courses offered by the Federal University of Santa Catarina. They were all undergraduate students from a variety of backgrounds (engineering, biology, and business, language teaching, among others). There were 16 female and 9

male, and their ages ranged from 18 to 27. They were all intermediate learners from semesters 7 and 8.

The main findings of the pilot study revealed no significant correlations between measures of L2 speech production and working memory capacity scores in the performance of the first narrative tasks for both control (N =13) and experimental group (N =12). In order to account for these results, the possibility of having overly complex task implementation conditions was raised. For individual differences in working memory capacity to emerge, the task under performance has to be complex (Fortkamp, 2000; Just & Carpenter, 1992; Tomitch, 1996;). Tasks which are either too complex or too easy seem to yield no individual differences in working memory capacity.

In addition to that, results also revealed that pre-task planning led to significant difference in fluency and accuracy of speech performance. Such a finding did not corroborate previous studies in which the greater impact of planning was on fluency and complexity (see Ellis, 2005a for a review). In order to provide a tentative explanation for these results, learner orientation was raised as an issue. It may be that participants in the present study were conservative and aimed at error free performance (accuracy) instead of taking risks in the attempt to achieve more complex performance. A need to further scrutinize the learning contexts in the attempt to expand our understanding about learners' orientation was also attested.

Results also revealed significant correlations between measures of L2 speech production and working memory capacity scores in the performance of the second narrative task of the control group (no planning) as well as between measures of L2 speech production and working memory capacity scores in the performance of the second narrative task of the experimental group (planning). In other words, individual

differences in working memory capacity emerged in the second tasks performed by both control and experimental groups. Hence, it was not possible to state whether the relationship between speech performance and working memory capacity was evident due to planning or task familiarity (by performing a task a second time). No conclusive evidence was reached as regards the relationship between pre-task planning, working memory capacity, and L2 speech performance.

As regards what learners do when they plan, results of the pilot study revealed that learners focused mainly on organizational, retrieval, rehearsal, and monitoring operations. The findings concerning retrieval and rehearsal operations during planning corroborate Ortega (2005). It was also found that these operations seem to have taken place in a chronological fashion. In other words, organization of ideas were mentioned by participants mainly in the beginning of planning time whereas monitoring seems to have been mentioned mainly at the end.

The pilot study had several limitations. The limited sample size (N =25) may have weakened the correlations between L2 speech performance and working memory capacity. Moreover, few variables of speech performance were investigated (only four measures). No strong claims can be made based given the small data set and such a small set of measures of speech production. Despite these limitations, the pilot study represented a step forward by taking individual differences in working memory into account. Most previous studies raised limitations in working memory capacity as a possible explanation for trade-off effects when results were discussed. In this sense, the pilot study went beyond, by suggesting that the relationship among pre-task planning, working memory capacity, and L2 speech performance is a complex one which merits further scrutiny.

Having described the methodology used for data collection and data analysis, the following chapter presents the results of the data analysis.

CHAPTER IV

RESULTS

4.1 Introduction

The aim of this chapter is to report the results of the present study whose purpose was to investigate the relationship among pre-task planning, individual differences in working memory capacity, and L2 speech performance. In the attempt to scrutinize this relationship, four main objectives were pursued: (a) to examine the relationship between working memory capacity scores and measures of L2 speech performance in no planning condition, (b) to examine whether planning leads to significant differences on L2 speech performance, (c) to examine the relationship between working memory capacity scores and measures of L2 speech performance in planning condition, and (d) to examine what processes learners engage in when they plan.

The remainder of this chapter is divided into six sections. Section 4.2 presents the factor analysis (Principal Component analysis) for the seven measures of L2 speech performance used in the present investigation: speech rate unpruned (SRU), speech rate pruned (SRP), number of silent pauses per c-unit (PCU), total % of silent pausing time (TPT), number of errors per a hundred words (ACCW), % of error free clauses (ACCC), and number of clauses per c-unit (COMP). Section 4.3 reports the correlational analyses between working memory capacity scores (lenient and strict) and the seven measures of L2 speech performance. Section 4.4 presents the descriptive statistics for the measures of working memory capacity (lenient and strict scores), the time taken to perform the test, and the seven measures of L2 speech performance. The

descriptive statistics is followed by the results of the ANOVAs performed to detect any significant difference between the control and experimental groups in the performance of the first narrative task (no planning condition for both groups), and in the performance of the second narrative task (no planning condition for control group and planning condition for the experimental group). Section 4.5 presents results of the ANOVAs concerning the extreme group design adopted in the attempt to further scrutinize performance differences between higher and lower working memory spans in the planning condition. Section 4.6 informs the results of the think aloud protocols in order to establish the processes learners engage in during pre-task planning. Section 4.7 reports on results of the t-tests performed in order to scrutinize any differences in the processes higher and lower spans engage in when they plan. For this analysis, again the extreme group design was adopted and only the upper and lower memory span tertiles were included. At the end of the chapter, a summary of results and hypotheses addressed is also presented.

4.2 Factor Analysis

As previously stated in Chapter III, L2 speech performance was assessed by means of fluency, accuracy, and complexity; thus, I start by showing that these measures underlie distinct dimensions of L2 speech. Table 3 shows the Principal Component Analysis carried out to check whether the different measures of L2 speech load on different components. This analysis was computed for the performance of the first narrative task in which both the control and the experimental group performed under the same condition, that is, a no planning condition, so that the analysis could be carried out with all the fifty participants of the study. Prior to the analysis, the suitability

of the data for factor analysis was examined. As said earlier, the Kaiser-Meyer-Oklin value was .61; the recommended value is at least .6. The Barlett's Test of sphericity reached statistical significance (p = .000).

Table 3
Factor Analysis of Performance Measures in Task 1

	Component		
	1	**2**	**3**
SRU	.920		
SRP	.944		
PCU	-.879		
TPT	-.640		
ACCW		-.923	
ACCC		.924	
COMP			.990

Note. The values below .30 are not reported simply to ease the interpretation of the output.

The findings of the factor analysis indicate that the fluency measures load highly on the first component. The fact that these measures load on the same component, being, thus, associated with each other, indicates that they are reflecting the same underlying construct, that is to say, fluency. As can be seen, participants who produced higher speech rates (unpruned and pruned) also produced fewer pauses per c-unit and lower percentage of total pausing time.

The accuracy measures load highly on the second component, and their association also indicates that they reflect the same underlying construct, namely, accuracy. As can be seen, participants who produced fewer errors per a hundred words also produced more error free clauses. Finally, complexity loads on the third component, underlying one more distinct construct of speech production. Having shown that the speech performance measures underlie distinct aspects of L2 speech, I turn now to the relationship between L2 speech performance and working memory capacity under the no planning and planning conditions.

4.3 Correlational analyses

In this section, I will report results of the correlational analyses. First, Table 4 displays the results of the correlations between working memory capacity and L2 speech performance in the *first* narrative task for the *control* group, and Table 5 displays the correlations between L2 speech performance in the *first* narrative task for the *experimental* group. As previously explained, the first narrative task was carried out under *no planning* condition for both control and experimental groups.

Second, Table 6 displays the correlations between working memory capacity and L2 speech performance in the *second* narrative task for the *control* group, and Table 7 displays the correlations between working memory capacity and L2 speech performance in the *second* narrative task for the *experimental* group. As previously explained, the second narrative task was carried out under a *no planning condition* for the control group and under a *planning condition* for the experimental group.

Table 4

Correlations Between Working Memory and Speech Performance in Task 1 (control group)

		SRU	SRP	PCU	TPT	ACCW	ACCC	COMP
WML	Pearson Correlation	.231	.215	.016	.027	**-.737****	**.798****	.240
	Sig. (2-tailed)	.267	.302	.940	.900	**.000**	**.000**	.249
	N	25	25	25	25	25	25	25
WMS	Pearson Correlation	.227	.236	.005	.041	**-.722****	**.785****	.222
	Sig. (2-tailed)	.275	.257	.981	.846	**.000**	**.000**	.286
	N	25	25	25	25	25	25	25

Note. WML= working memory lenient scores; WMS = working memory strict scores; SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

** p<. 01

As shown in Table 4, the correlations between working memory lenient and strict scores and fluency failed to achieve significance. These results corroborate those of the pilot study (Guará-Tavares, 2006) reported in the previous chapter. Interestingly, however, these results are at odds with those reported by Fortkamp (1999, 2003) and Mizera (2006), in which positive correlations between working memory and fluency measures of speech rate were found.

The picture is different when it comes to accuracy. Results in Table 4 show that there were significant correlations between working memory and accuracy. There was a negative correlation between working memory capacity lenient and strict scores and number of errors per a hundred words (r = -737, p = 000, and r = -722, p = 000). The magnitude of these correlations (-737 and -722) is relatively large (Cohen, 1988; 1992). These results indicate that individuals with higher working memory capacity are the ones who tend to make fewer errors per a hundred words.

Similarly, there is a positive correlation between working memory capacity lenient and strict scores and the percentage of error free clauses produced (r = 798, p = 000, and r = 785, p = 000). The magnitude of these correlations is relatively large

(Cohen, 1988; 1992). These correlations indicate that participants with higher working memory capacity are the ones who produce a higher percentage of error free clauses.

These results corroborate those reported by Fortkamp (2003), Bergsleitner (2007), and Weissheimer (2007), in which significant correlations between working memory and accuracy were also found. Interestingly, though, these results are at odds with results of the pilot study (Guará-Tavares, 2006), in which there were no correlations between working memory and accuracy in the performance of the first narrative task under no planning condition.

As for complexity, results in Table 4 reveal no significant correlations between working memory lenient and strict scores and complexity (as measured by number of clauses/c-unit). These results corroborate those of the pilot study (Guará- Tavares, 2006), in which there were no correlations between working memory and complexity in the performance of the first narrative task.

Nevertheless, these results do not corroborate those of Fortkamp (2000) and Weissheimer (2007), in which there were significant correlations between working memory and complexity. These results will be discussed in the next chapter of this dissertation.

Having reported the correlations between working memory and L2 speech performance for the *control group in task 1*, I turn now to the results of the *experimental group in task 1*.

Table 5

Correlations Between Working Memory and Speech Performance in Task 1 (Experimental group)

		SRU	SRP	PCU	TPT	ACCW	ACCC	COMP
WML	Pearson Correlation	.215	.237	-.163	**-.422***	-.183	.079	.243
	Sig. (2-tailed)	.302	.254	.437	**.036**	.382	.706	.243
	N	25	25	25	**25**	25	25	25
WMS	Pearson Correlation	.199	.212	-.130	**-.461***	-.140	.009	.267
	Sig. (2-tailed)	.341	.309	.535	**.020**	.506	.964	.196
	N	25	25	25	25	25	25	25

Note. WML= working memory lenient scores; WMS = working memory strict scores; SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

* $p<0.05$

According to the results shown in Table 5, there were significant correlations between working memory and fluency. There were negative correlations between working memory capacity lenient and strict scores and the percentage of total pausing time ($r = -.422$, $p = .036$, and $r = -.461$ and $p = .020$). The magnitude of these correlations is moderate (Cohen, 1988; 1992). These correlations show that participants with higher working memory capacity are the ones who produce a lower percentage of total pausing time.

These results corroborate those of Fortkamp (1999, 2003), in which there were correlations between working memory and fluency. It is important to highlight that Fortkamp (1999, 2003) used measures of speech rate, number of pauses per minute and mean length of run.

These results presented in Table 5 contradict results of the pilot study (Guará-Tavares, 2006), in which there were no significant correlations between working memory and fluency in the performance of the first narrative task. These results also contradict the results found for the *control group* in the *present* study. As previously shown in Table 4, there were no significant correlations between working memory

lenient and strict scores and *fluency* in the performance of the first narrative task for the control group.

There were no correlations between working memory and accuracy (as measured by number of errors/100 words and % of error free clauses). These results corroborate results of the pilot study (Guará-Tavares, 2006) in which there were no correlations between working memory and accuracy in the performance of the first narrative task.

However, this lack of correlations between working memory and accuracy is at odds with the results reported by Fortkamp (2003), Bergsleitner (2007) and Weissheimer (2007). These results also contradict the ones found for the *control group* in the *present* study. As previously shown in Table 4, there were significant correlations between working memory lenient and strict scores and *accuracy* in the first narrative task for the control group.

As for complexity, results in Table 5 show no significant correlations between working memory and complexity (as measured by number of clauses/c-unit). These results corroborate the findings of the pilot study (Guará-Tavares, 2006), in which there were no significant correlations between working memory and complexity in the performance of the first narrative task. These results also corroborate the ones found for the *control group* in the *present* study in which there were no correlations between working memory and complexity.

Nevertheless, these results are at odds with the ones reported by Fortkamp (2003) and Weissheimer (2007), in which significant correlations between working memory and complexity were found.

Having reported results of the correlations between working memory and L2 speech performance in the *first* narrative task for both the control and experimental

groups, I turn now to the correlations between working memory and L2 speech performance in the *second* narrative task, which are displayed in Tables 6 and 7. As previously explained in Chapter III, participants in the control group performed the second narrative task under a no planning condition, whereas participants in the experimental group performed the second narrative task under a planning condition.

Table 6

Correlations Between Working Memory and Speech Performance in Task 2 (Control group)

		SRU	SRP	PCU	TPT	ACCW	ACCC	COMP
WML	Pearson Correlation	.032	.038	-.007	.282	**-.703****	.740**	-.036
	Sig. (2-tailed)	.881	.858	.973	.172	**.000**	.000	.863
	N	25	25	25	25	**25**	25	25
WMS	Pearson Correlation	.016	.021	.008	.273	**-.676****	.696**	-.072
	Sig. (2-tailed)	.940	.921	.970	.186	.000	.000	.732
	N	25	25	25	25	25	25	25

Note. WML= working memory lenient scores; WMS = working memory strict scores; SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

** p<. 01

As displayed in Table 6, the same pattern of correlations between accuracy and working memory capacity found in the performance of the *first* narrative task of the control group was also found in the performance of the *second* narrative task. There were significant negative correlations between working memory capacity lenient and strict scores and number of errors per a hundred words (r =-.703, p = 000, and r = -.676, p = 000) and percentage of error free clauses (r = .740, p = .000, and r = .696, and r = .000). The magnitude of these correlations is relatively large (Cohen, 1988; 1992). These correlations indicate that participants with higher working memory capacity produce fewer errors per a hundred words and a higher percentage of error free clauses. These results corroborate previous results of previous studies in the literature (Fortkamp, 2003; Bergsleithner 2007; Weissheimer 2007) in which significant correlations between working memory capacity and accuracy were also revealed.

The correlations between working memory capacity scores and fluency (as measured by speech rate unpruned, pruned, number of pauses per c-unit, total pausing time), and complexity (as measured by number of clauses/c-unit) all failed to achieve significance. There results are, thus, at odds with previous studies reported in the literature (Guará-Tavares, 2006; Fortkamp, 1999; 2003; Weissheimer, 2007). Having reported the correlations between working memory and L2 speech performance for the *control* group in *Task 2*, I turn now to the results of the *experimental* group in *Task 2* under planning condition.

Table 7

Correlations Between Working Memory and Speech Performance in Task 2 (Experimental group)

		SRU2	SRP2	PCU 2	TPT2	ACCW2	ACCC2	COMP2
WML	Pearson Correlation	**.430***	**.442***	.159	-.294	-.371	.229	**.426***
	Sig. (2-tailed)	**.032**	**.027**	.448	.154	.068	.271	**.034**
	N	25	25	25	25	25	25	25
WMS	Pearson Correlation	**.481***	**.494***	.146	-.290	-.335	.223	.345
	Sig. (2-tailed)	**.015**	**.012**	.485	.160	.102	.284	.092
	N	25	25	25	25	25	25	25

Note. WML= working memory lenient scores; WMS = working memory strict scores; SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

* $p<0.05$

As shown in Table 7, there are significant correlations between fluency, as measured by speech rate unpruned, and working memory capacity lenient and strict scores respectively (r = .481*, p = .015, and r = .430*, p = .032). The magnitude of these correlations is considered moderate (Cohen, 1988, 1992); there are also significant correlations between fluency as measured by speech rate pruned and working memory capacity lenient and strict scores, respectively (r = .442, p =.027, and r = .494, p = .012). The magnitude of these correlations is also moderate.

The correlations between working memory capacity lenient scores and complexity, as measured by the number of clauses per c-unit are also significant (r =.426, p = .034). The magnitude of these correlations (.426) is also moderate (Cohen, 1988, 1992). The correlations between accuracy, as measured by number of errors per a hundred words, and working memory capacity lenient scores only approached significance (r = -.371 and p = .068). Taken together, these correlations show that higher span individuals speak significantly faster, produce significantly more clauses per c-unit, and tend to produce, although only marginally significant, fewer errors per a hundred words when performing a task under a planning condition. In general, these results tend to corroborate those of Fortkamp (2003) and those of the pilot study. However, these results contradict those of the *control group* in the *present* study. As shown in Table 5, in the performance of the second task for the control group, the only significant correlations were between working memory capacity and accuracy.

The correlations between working memory capacity and the other measures of L2 speech performance (number of pauses per c-unit, total pausing time, and percentage of error free clauses) all failed to achieve significance.

In brief, results of the correlations between working memory capacity and L2 speech performance under no planning and planning conditions show that:

1. there are significant correlations between working memory capacity and accuracy in the performance of Task 1 (no planning), and task 2 (no planning) for the control group.
2. there are significant correlations between working memory capacity and fluency in the performance of Task 1 (no planning) for the experimental group.

3. There are significant correlations between working memory capacity and fluency in the performance of Task 2 (planning) for the experimental group; there are significant correlations between working memory and complexity in the performance of Task 2 (planning) for the experimental group.

Reflecting upon these results, I can return now to some of the hypotheses, which were proposed earlier in this dissertation. Hypothesis 1, which predicted that there would be correlations between working memory capacity scores and fluency measures of L2 speech performance under *no planning conditions,* is only partially supported. Correlations between working memory and fluency (as measured by total percentage of pausing time) only achieved significance in the performance of the first narrative task of the experimental group.

Hypothesis 2, which predicted that working memory capacity scores would correlate with accuracy measures of L2 speech performance under *no planning conditions*, is only partially supported. Correlations between working memory and accuracy only achieved significance in the performance of the first and second narrative tasks of the control group; performance of the first narrative task of the experimental group yielded no significant correlations between working memory and accuracy.

Hypothesis 3, which predicted that there would be significant correlations between working memory capacity scores and complexity measures of L2 speech performance *under no planning condition*, is not supported.

Hypothesis 7, which predicted that working memory capacity scores would significantly correlate with fluency of L2 speech performance *under planning condition*, is partially supported. Significant correlations were found between working memory

and fluency as measured by speech rate unpruned and pruned; correlations between working memory and number of pauses per c-unit and total percentage of pausing time failed to achieve significance. Hypothesis 8, which predicted that there would be significant correlations between working memory capacity scores and accuracy measures of L2 speech performance under *planning condition*, is not supported.

Hypothesis 9, which predicted that there would be significant correlations between working memory capacity scores and complexity measures of L2 speech performance under *planning condition*, is supported.

In this section, I have reported the results regarding the relationship between working memory capacity and L2 speech performance in both planning and no planning conditions. These results will be discussed in the next chapter. In the next section, I will report the results of the impact of planning on L2 speech performance.

4.4 ANOVAs

In order to verify whether there were statistically significant differences between the control and experimental groups, ANOVAs were computed (as previously explained in Chapter III). First, I will report whether the two groups are homogeneous in terms of performance on the Speaking Span Test. Second, I will report whether there are any significant differences between the two groups in terms of the performance of the first narrative task (no planning condition for both groups). Third, I will report whether there are significant differences between the two groups in terms of the performance of the second narrative task (no planning condition for control and planning for experimental). Tables 8 and 9 display the descriptive statistics, and Table 10 displays the differences in working memory capacity and response time for the experimental and control groups.

Table 8

Descriptive Statistics - Working Memory Lenient and Strict Scores

Working memory capacity scores	group	N	Minimum	Maximum	Mean	Std. deviation
WML	Control	25	12	38.50	24.81	6.72
	Experimental	25	9	37	23.64	7.54
WMS	Control	25	10	38	23.84	7.08
	Experimental	25	9	35	24.92	7.56

Note. WML = working memory lenient scores; WMS = working memory strict scores.

Table 9

Descriptive Statistics - Response Time

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Response time	Control	25	4.01	18.70	8.36	3.72
	Experimental	25	5.30	12.30	7.94	2.14

Table 10

ANOVA - Working Memory Capacity Scores and Response Time

		Sum of squares	df	Mean square	F	Sig.
WMS	Between groups	.320	1	.320	.006	.939
	Within groups	2555.500	48	53.240		
	Total	2555.820	49			
WML	Between groups	.125	1	.125	.002	.961
	Within groups	2459.780	48	51.245		
	Total	2459.905	49			
RT	Between groups	2.264	1	2.264	.245	.623
	Within groups	443.835	48	9.247		
	Total	446.100	49			

*Note***.** WMS = working memory strict scores; WML = working memory lenient scores; RT = response time.

As can be seen in Table 10, there were no significant differences between the control and experimental groups neither in terms of working memory capacity scores nor in the time the groups took to perform the Speaking Span Test. Therefore, the control and experimental groups are homogeneous in terms of working memory capacity scores.

Tables 11, 12, and 13 display the descriptive statistics of speech performance measures in Task 1. Table 14 displays results of the ANOVAs comparing the performance of control and experimental groups in Task 1.

Table 11

Descriptive Statistics - Fluency Measures of Speech Performance in Task 1

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Speech rate unpruned	Control	25	46.66	118.75	75.73	15.03
	Experimental	25	39.09	115.47	72.83	20.60
Speech rate pruned	Control	25	36.11	113.75	71.54	16.80
	Experimental	25	35.40	114.33	70.13	20.59
Number of pauses per c-unit	Control	25	.38	3.40	1.49	.8669
	Experimental	25	.36	3.20	1.45	.7088
Total pausing time	Control	25	.11	.60	.2936	.1407
	Experimental	25	.09	.82	.3022	.1598

Table 12

Descriptive Statistics -Accuracy Measures of Speech Performance in Task 1

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Number of errors/100 words	Control	25	2.0	9.09	5.27	2.35
	Experimental	25	1.30	15.70	6.87	3.57
% of error free clauses	Control	25	.25	.81	.6368	.1589
	Experimental	25	.21	.88	.5792	.1995

Table 13

Complexity Measures of Speech Performance in Task 1

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Number of clauses/c-unit	Control	25	1.10	1.70	1.36	.1701
	Experimental	25	1.10	2.10	1.36	.2212

Table 14

Speech Performance in Task 1

		Sum of squares	df	Mean square	F	Sig.
SRU	Between groups	104.748	1	104.748	.322	.573
	Within groups	15618.769	48	325.391		
	Total	15723.517	49			
SRP	Between groups	24.865	1	24.865	.070	.792
	Within groups	169.62691	48	353.389		
	Total	16987.556	49			
PCU	Between groups	2.376E-02	1	2.376E-02	0.38	.846
	Within groups	30.092	48	.627		
	Total	30.116	49			
TPT	Between groups	9.245E-04	1	9.245E-04	0.41	.841
	Within groups	1.088	48	2.266E-02		
	Total	1.089	49			
ACCW	Between groups	31.840	1	31.840	3.018	.089
	Within groups	440.691	48	9.181		
	Total	472.531	49			
ACCC	Between groups	4.147E-02	1	4.147E-02	1.275	.264
	Within groups	1.561	48	3.253-02		
	Total	1.603	49			
COMP	Between groups	.000	1	.000	.000	1.000
	Within groups	1.869	48	3.893E-02		
	Total	1.869	49			

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

As shown in Table 14, there are no significant differences between the control and experimental groups in the performance of Task 1 in terms of fluency as measured by speech rate unpruned (f = .322, p = .573), speech rate pruned (f = .070, p = .692), number of pauses per c-unit (f = .038, p = .846), and percentage of total pausing time (f = .041, p =. 841).

In addition, there are no significant differences between the control and experimental groups in the performance of Task 1 in terms of accuracy as measured by

number of errors per a hundred words (f = 3.468, p = .089), nor as measured by the percentage of error free clauses (f = 1.275, p = .264).

Finally, there are no significant differences between the control and experimental groups in the performance in Task 1 in terms of complexity as measured by the number of clauses per c-unit (f = 000, p = 1.000). Therefore, there are no significant differences between the control and experimental groups in the performance of the first narrative task carried out under a no planning condition for both groups, which allows me to argue that statistically significant differences that emerge between the control and experimental groups, in the L2 speech performance of the *second* narrative task, can be attributed to planning.

Now I turn to the performance of the control and experimental groups in the second narrative task. Tables 15, 16 and 17 display the descriptive statistics of speech performance measures in Task 2, and Table 18 displays the results of the ANOVA comparing the performance of the control and experimental groups in Task 2. Again, it is important to highlight that the *control group* performed Task 2 under a *no planning* condition whereas the *experimental group* performed Task 2 under a *planning* condition.

Table 15

Descriptive Statistics – Fluency Measures

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Speech rate unpruned	Control	25	28.07	107.69	72.10	20.85
	Experimental	25	41.25	125.40	79.30	21.48
Speech rate pruned	Control	25	27.52	107.69	70.07	20.88
	Experimental	25	40.50	125.40	76.89	21.45
Number of pauses per c-unit	Control	25	.20	3.50	1.44	.9724
	Experimental	25	.30	12.10	1.81	2.25
Total % of pausing time	Control	25	.03	.76	.3164	.2161
	Experimental	25	.07	.51	.2645	.1273

Table 16

Descriptive Statistics - Accuracy Measures

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Number of errors/100 words	Control	25	1.49	12.00	6.36	3.007
	Experimental	25	00	9.60	4.96	2.700
% of error free clauses	Control	25	.33	.90	.6140	.1565
	Experimental	25	.47	1	.7148	.1475

Table 17

Descriptive Statistics – Complexity Measure

Measures	group	N	Minimum	Maximum	Mean	Std. deviation
Number of clauses/c-unit	Control	25	1.10	2.0	1.39	.2458
	Experimental	25	1.20	2.10	1.55	.2567

Table18

ANOVA - Performance on the Second Narrative Task (Control and Experimental groups)

Performance measures		Sum of squares	df	Mean square	F	Sig.
SRU	Between groups	647.424	1	647.424	1.444	.235
	Within groups	21519.148	48	448.316		
	Total	22166.572	49			
SRP	Between groups	582.019	1	582.019	1.298	.260
	Within groups	21516.311	48	448.256		
	Total	22098,330	49			
PCU	Between groups	1.696	1	1.696	.561	.458
	Within groups	145.174	48	3.024		
	Total	146.871	49			
TPT	Between groups	3.370E-02	1	3.370E-02	1.071	.306
	Within groups	1.510	48	3.145E-02		
	Total	1.543	49			
ACCW	Between groups	24.654	1	24.654	3.468	.079
	Within groups	392.127	48	8.169		
	Total	416.782	49			
ACCC	Between groups	.127	1	.127	**5.492***	**.023**
	Within groups	1.110	48	2.313E-02		
	Total	1.237	49			
COMP	Between groups	.320	1	.320	**5.067***	**.029**
	Within groups	3.031	48	6.315E-02		
	Total	3.351	49			

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses /c-unit; TPT = total percentage of pausing time; ACCW = number of errors/100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

* p<. 05

As can be seen in Table 18, there were significant differences in the performance of the experimental group when compared to the performance of the control group in the performance of the second narrative task. There were significant differences in accuracy as measured by the percentage of error free clauses (f = 5.492*, p= 0. 023), with a medium effect size (d=.66), and in complexity as measured by the number of clauses per c-unit (f = 5.067*, p= 0. 29), with a medium effect size (d=.65). Bearing in mind that there were no significant differences between the control and experimental groups in the performance of the first narrative task, it can be argued that that pre-task planning led to significant differences in accuracy and complexity. In general, these results corroborate previous studies in the literature, in which planning also leads to gains in performance (Foster & Skehan, 1996; Mehnert, 1998; Ortega,

1999, among others). However, the aspects of performance for which planning leads to gains in most previous studies are *fluency* and complexity. These results will be discussed in the next chapter of this thesis.

In brief, results concerning the impact of planning on L2 speech performance show that:

1. There are no significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of fluency.
2. There are significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of accuracy as measured by the percentage of error free clauses.
3. There are significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of complexity as measured by the number of clauses per c-unit.

Reflecting upon these results, I can return now to some of the hypotheses proposed earlier in this dissertation. Hypothesis 4, which predicted that under planning condition there would be greater fluency for the experimental group when compared to the control group under no planning condition, is not supported.

Hypothesis 5, which predicted that under planning condition there would be greater accuracy for the experimental group when compared to the control group under no planning condition, is partially supported. There were significant differences in favor

of the experimental group in terms of percentage of error free clauses but not in terms of number of errors per a hundred words.

Hypothesis 6, which predicted that under planning condition there would be greater complexity for the experimental group when compared to the control group under no planning condition, is supported. These results will be discussed in the next chapter of this thesis. Now I turn to the results of the ANOVAs comparing differences in the speech performance of higher and lower spans under planning conditions.

4.5 Differences between the performance of lower and higher spans

As previously explained in Chapter 3-Method, I adopted an extreme-group design in the attempt to scrutinize how lower and higher spans differ when they perform a task under a planning condition. Table 19 displays the descriptive statistics in order to give an overall view of the performance of lower and higher span participants in Task 2, which was carried out under a planning condition. Table 20 displays the results of the ANOVA computed to compare the performance of lower and higher span participants in Task 2.

Table 19

Descriptive Statistics - Speech Performance Measures Task 2 (lower and higher spans-Experimental group)

Performance measures	group	N	Minimum	Maximum	Mean	Std. deviation
SRU	Experimental low spans	8	41.25	85.82	63.3250	17.9846
	Experimental high spans	8	48.29	125.40	88.5425	24.3321
SRP	Experimental low spans	8	40.50	82.02	60.8662	16.8081
	Experimental high spans	8	47.65	125.40	86.8788	24.5519
PCU	Experimental low spans	8	.50	3.20	1.5375	.8123
	Experimental high spans	8	.30	12.10	2.4613	3.9556
TPT	Experimental low spans	8	.12	.49	.3150	.1311
	Experimental high spans	8	.09	.41	.2259	.1317
ACCW	Experimental low spans	8	1.70	9.60	6.2625	2.6645
	Experimental high spans	8	.00	7.80	3.6650	2.7706
ACCC	Experimental low spans	8	.53	.94	.6838	.1351
	Experimental high spans	8	.54	1.00	.7850	.1450
COMP	Experimental low spans	8	1.20	1.80	1.4000	.2330
	Experimental high spans	8	1.40	2.10	1.6625	.2875

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses per / c-unit; TPT = total percentage of pausing time; ACCW = number of errors per/ 100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit

Table 20

ANOVA - Speech Performance in Task 2(lower and higher spans-Experimental group)

Performance measures		Sum of Squares	df	Mean Square	F	Sig.
SRU	Between Groups	3132.641	1	3132.641	**8.676***	**.011**
	Within Groups	5055.004	14	361.072		
	Total	8187.645	15			
SRP	Between Groups	3310.564	1	3310.564	**9.473***	**.008**
	Within Groups	4892.570	14	349.469		
	Total	8203.133	15			
PCU	Between Groups	3.422	1	3.422	.420	.528
	Within Groups	114.138	14	8.153		
	Total	117.560	15			
TPT	Between Groups	5.210E-02	1	5.210E-02	3.521	.082
	Within Groups	.207	14	1.480E-02		
	Total	.259	15			
ACCW	Between Groups	30.140	1	30.140	3.903	.068
	Within Groups	108.107	14	7.722		
	Total	138.247	15			
ACCC	Between Groups	6.002E-02	1	6.002E-02	2.663	.125
	Within Groups	.316	14	2.254E-02		
	Total	.376	15			
COMP	Between Groups	.490	1	.490	**6.725***	**.021**
	Within Groups	1.020	14	7.286E-02		
	Total	1.510	15			

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses per /c-unit; TPT = total percentage of pausing time; ACCW = number of errors per/ 100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit p<0. 05.

As can be seen from the results displayed in Table 20, there were significant differences between lower and higher spans when they perform a task under a planning condition. There were significant differences in terms of fluency as measured by speech rate unpruned (f = 8.676, p = 0. 011) and pruned (f = 9.473, p = 0. 008); and there were also significant differences in terms of complexity as measured by number of clauses per c-unit (f = 6.725, p = 0.021). Although differences in accuracy as measured by number of errors/100 words did not achieve significance, they approached significance (3.093, p = .068).

The fact that higher spans outperformed lower spans in fluency and complexity under the planning condition may be due to the fact that higher and lower spans were different a priori, that is to say, in the performance of Task 1 under a no planning

condition. In other words, it may be that higher spans outperformed lower spans because of individual differences in working memory capacity *only,* regardless of planning.

In order to check whether planning may have had any sort of effect on the significant differences which emerged between the performance of lower and higher spans under planning condition, it is necessary to examine how they behaved under no planning condition as well. Table 21 displays the results of the descriptive statistics in order to give an overall view of the performance of lower and higher span participants in Task 1, and Table 22 displays the results of the ANOVA computed to compare the performance of lower and higher span participants in Task 1.

Table 21

Descriptive Statistics - Speech Performance Measures Task 1 (lower and higher spans-Experimental group)

Performance Measures	group	N	Minimum	Maximum	Mean	Std. deviation
SRU	Experimental low spans	8	39.09	88.86	64.5063	14.8751
	Experimental high spans	8	48.75	115.47	82.6825	21.7166
SRP	Experimental low spans	8	35.40	79.81	61.1388	14.0329
	Experimental high spans	8	47.34	114.33	80.8400	21.4712
PCU	Experimental low spans	8	1.20	3.20	1.8125	.7434
	Experimental high spans	8	.36	2.09	1.1138	.5223
TPT	Experimental low spans	8	.23	.56	.3288	.1141
	Experimental high spans	8	.09	.56	.2681	.1495
ACCW	Experimental low spans	8	3.80	15.70	8.2337	4.2096
	Experimental high spans	8	1.30	13.90	5.2288	3.9189
ACCC	Experimental low spans	8	.25	.81	.5675	.2029
	Experimental high spans	8	.21	.88	.6413	.2511
COMP	Experimental low spans	8	1.10	1.50	1.3375	.1506
	Experimental high spans	8	1.10	2.10	1.4750	.3105

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses per /c-unit; TPT = total percentage of pausing time; ACCW = number of errors per/ 100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit.

Table 22

ANOVA - Speech Performance in Task 1 (lower and higher spans - Experimental group)

Performance measures		Sum of Squares	df	Mean Square	F	Sig.
SRU	Between Groups	865.684	1	865.684	2.284	.153
	Within Groups	5306.481	14	379.034		
	Total	6172.164	15			
SRP	Between Groups	966.588	1	966.588	2.655	.125
	Within Groups	5096.194	14	364.014		
	Total	6062.782	15			
PCU	Between Groups	.718	1	.718	1.536	.236
	Within Groups	6.549	14	.468		
	Total	7.267	15			
TPT	Between Groups	5.820E-02	1	5.820E-02	1.771	.204
	Within Groups	.460	14	3.286E-02		
	Total	.518	15			
ACCW	Between Groups	24.354	1	24.354	1.534	.236
	Within Groups	222.219	14	15.873		
	Total	246.573	15			
ACCC	Between Groups	7.225E-03	1	7.225E-03	.162	.694
	Within Groups	.626	14	4.468E-02		
	Total	.633	15			
COMP	Between Groups	9.000E-02	1	9.000E-02	1.556	.233
	Within Groups	.810	14	5.786E-02		
	Total	.900	15			

Note. SRU =speech rate unpruned; SRP= speech rate pruned; PCU = number of pauses per /c-unit; TPT = total percentage of pausing time; ACCW = number of errors per/ 100 words; ACCC = percentage of error free clauses; COMP = number of clauses /c-unit.

As can be seen in Table 22, the mean differences in all measures of speech performance favor higher spans when compared to lower spans. However, as shown in Table 23, none of these differences achieved statistical significance. There were no significant differences between lower and higher spans in the performance of the first narrative task. These results suggest that significant differences in the performance of lower and higher spans in Task 2, previously reported in Tables 19 and 20, may have taken place *not only* because of individual differences in working memory capacity, *but also* due to the opportunity of pre-task planning.

In brief, results concerning whether higher spans significantly outperform lower spans in L2 speech performance under planning conditions show that:

1. Higher spans significantly outperformed lower spans in terms of fluency as measured by speech rate unpruned and pruned.
2. Higher spans outperformed lower spans in accuracy as measured by number of errors/100 words, however, not significantly; the difference in favor of the higher spans only approached significance.
3. Higher spans significantly outperformed lower spans in terms of complexity as measured by number of clauses per c-unit.

Reflecting upon these results, I can return now to some of the hypotheses proposed earlier in this dissertation. Hypothesis 10 predicted that, within the experimental group under planning conditions, higher working memory spans would significantly outperform lower working memory spans in terms of fluency. There were significant differences in fluency as measured by speech rare unpruned and pruned, thus, Hypothesis 10 is partially supported.

Hypothesis 11 predicted that, within the experimental group under planning conditions, higher working memory spans would significantly outperform lower working memory spans in terms of accuracy. This hypothesis was not supported; differences in accuracy as measured by errors/100 words only approached significance.

Hypothesis 12 predicted that, within the experimental group under planning condition, higher working memory spans would significantly outperform lower working memory spans in terms of complexity. Higher spans significantly outperformed lower spans in complexity of speech performance, thus, Hypothesis 12 is supported. These results will be discussed in the next chapter.

In this section, I have reported results concerning the effects of planning on performance and the differences between the performance of higher and lower spans under planning condition. Now I turn to the results of the analysis of the verbal protocols carried out in order to examine the processes learners engage in when they plan performance of an oral task.

4.6 Analysis of the protocols

In this section, I will report the results of the analysis of the protocols conducted in the attempt to scrutinize the processes learners engage in during planning. First, I will present the strategies reported by participants and provide examples of these strategies with excerpts from the protocols. Second, I will present the raw and percentage of learners reporting each strategy. Third, I will report the results of the descriptive statistics in order to give an overall view of strategy types reported by speakers. Finally, I will present the results of the *t*-tests computed in order to compare strategies used by lower and higher spans based on the extreme-group design adopted in this study. Table 23 provides examples of each strategy taken from learners' protocols, and Table 24 displays the raw number and percentage of learners reporting the strategy types documented in the online protocols during pre-task planning and in the interviews after task performance.

Table 23

Strategies Reported by Participants

STRATEGY	Examples
METACOGNITIVE STRATEGIES	
Organizational planning	"I was thinking of names to the characters and putting the story together" (p16)
Problem identification	"I was thinking what I do if I don't remember a word" (p02)
monitoring	"I wrote here -*didn't should*- and should is a modal and I don't need to use the -*didn't*" (p07)
evaluation	"I'm trying to check if the mains ideas were organized in my story" (p18)
Rehearsal	"if I read it again I will imagine the story in my head to remember when I tell" (p.25)
COGNITIVE STRATEGIES	
Writing/outlining/summarizing	"I wrote something like a skeleton" (p17)
Elaboration	"I am improving my sentences., for example, I said -*they started to talk*- and now I said -*it seems that they stated to talk about…"* (p01)
Imagery	"I was just remembering all the gifts that he bought to her and drawing the gifts" (p16)
Lexical search	"I was thinking about the presents the man gives to the woman the name of the presents I'm not sure if *anel* is ring " (p22)
Avoidance	"I tried to remember the verb -*ter coragem*- but I will use a different idea" (p04)
Lexical compensation	"I am changing the word –*oprimido*- for another like the man is shy and quiet and tiny" (p18)
Translating	"I thought of a verb, no an expression –*finally*- I thought in Portuguese then in English" (p06)
Cross language analysis	"I don't know how to say he's trying to *let* him crazy but I don't know if that's the word *let*, in Portuguese we say -*deixar louco*- (p12)
SOCIAL/AFFECTIVE STRATEGIES	
Appeal for help	"Please, what do I do if I don't remember a word?"(p03)
Lowering anxiety	"I was thinking that if I start to worry too much about grammar I will be too nervous, I can't, I try not worry too much" (p25)

Table 24

Strategies Reported by Participants

Strategy types	Total sample (N = 25)		Lower spans (N = 8)		Intermediate spans (N = 9)		Higher spans (N = 8)	
	Raw	Percent	Raw	Percent	Raw	Percent	Raw	Percent
METACOGNITIVE STRATEGIES								
Organizational planning	16	64%	4	50%	6	66.6%	6	75%
Problem identification	7	28%	2	25%	2	22.2%	3	37.5%
monitoring	15	60%	4	50%	4	44.4%	7	87.5%
evaluation	7	28%	3	37.5%	2	22.2%	3	7.5%
Rehearsal	11	44%	2	25%	5	55%	4	50%
COGNITIVE STRATEGIES								
Writing/outlining/ summarizing	21	84%	6	75%	7	77.7%	8	100%
Elaboration	10	40%	2	25%	4	44.4%	4	50%
Imagery	5	20%	1	12.5%	2	22.2%	2	25%
Lexical search	24	96%	7	87.5%	9	100%	8	100%
Avoidance	3	12%	1	12.5%	1	11.1%	1	12.5%
Lexical compensation	7	28%	1	12.5%	3	33.3%	3	37.5%
translating	2	8%	1	12.5%	1	11.1%	0	0 %
Cross language analysis	2	8%	1	12.5%	0	0%	1	12. 5%
SOCIAL/AFFECTIVE STRATEGIES								
Appeal for help	1	4%	1	12%	0	0%	0	0%
Lowering anxiety	1	4%	0	0%	0	0%	1	12.5%

In general, the strategies most reported by participants were lexical search (96%), writing/summarizing, outlining (84%), organizational planning (64%), monitoring (60%), rehearsal (44%), and elaboration (40%). In addition, it can be seen that more learners in the high span memory group reported using these most frequent strategies than did learners in the lower span memory group. Out of the learners who reported using organizational planning, six were higher spans and four were lower spans. This pattern was even more evident in monitoring with seven higher spans and only four lower spans reporting this strategy. As for writing/summarizing/outlining, it can be seen that this strategy was frequently reported by learners in all span groups, but again the number of higher spans, eight, was greater than the number of lower spans,

six. The differences were identical for elaboration and rehearsal, with four higher spans and two lower spans reporting the use of these strategies.

Overall, the most frequent strategies (lexical search, organizational planning, writing/summarizing/outlining, monitoring, rehearsal and elaborating) were reported more frequently by higher spans than by lower spans. In order to examine whether there are statistically significant differences between higher and lower spans as regards the strategies reported, an independent *t*-test was performed.

Since the use of social/affective strategies was extremely low with only *one* lower span learner reporting appeal for help and only *one* higher span learner reporting a lowering anxiety strategy, these strategies were not included in the independent t-test. The focus was on examining differences between the number of metacognitive and cognitive strategies as well as differences in the total number of strategies (cognitive and metacognitive all together) utilized by learners. Table 26 displays the descriptive statistics of strategies reported by lower and higher spans, and Table 27 displays the results of the independent *t*-test.

Table 25

Descriptive Statistics – Strategy Types Reported by Learners

	All strategies	Metacognitive	Cognitive
Lower spans (N=8)			
Mean	3.5	1.6	1.75
SD	1.06	.51	.88
Minimum	2	1	1
Maximum	5	2	3
Higher spans (N=8)			
Mean	5.25	2.75	2.37
SD	1.48	.88	1.18
Minimum	2	1	1
Maximum	8	4	4

Table 26

Independent t-test - strategy types reported by lower and higher spans

	Group	N	t	df	Sig. (2-tailed)
All strategies	higher spans	8	**2.701***	12.706	**.018**
	lower spans	8			
metacognitive	higher spans	8	**3.100***	11.276	**.010**
strategies	lower spans	8			
cognitive strategies	higher spans	8	1.193	12.951	.254
	lower spans	8			

p<0. 05

As can be seen in Table 25, the means of strategies reported (all strategies, metacognitive and cognitive) all favor higher spans when compared to lower spans. As shown in Table 26, some of these differences achieved statistical significance. There were statistically significant differences in the number of all strategies reported by lower and higher span learners (t = 2.701, p = .018); and there were also statistically significant differences in the number of metacognitive strategies reported (t =3.1, p =.010). Being the fact that differences in the number of cognitive strategies reported was not significant (t = 1.193, p = .254), it seems that it was the difference in the number of metacognitive strategies which accounted more for the differences in the total number of strategies.

In brief, results concerning the processes learners engage in show that:

1. Learners engaged mainly in writing/outlining, summarizing, lexical search, organizational planning, monitoring, rehearsal, and elaboration during pre-task planning.
2. Higher span learners employed significantly more metacognitive strategies than lower spans during pre-task planning.

Reflecting upon these results, I can return now to the last hypotheses proposed in this dissertation. Hypothesis 13 predicted that learners would engage mainly in organization of ideas, lexical-grammatical search, task rehearsal, and monitoring. Results reported showed that learners engaged in in these processes and also in writing/outlining/summarizing and elaboration. Consequently, Hypothesis 13 is confirmed.

Hypothesis 14 predicted that lower and higher spans would differ in the processes they engage in during pre-task planning. Results showed that lower and higher spans differed in the number of metacognitive strategies employed; thus, Hypothesis 14 is confirmed.

Having reported the results of the study, I will present a summary of all results. Table 27 presents a summary of all hypotheses of the study, their predictions, and whether such predictions were supported or not by the results of this study.

Table 27

Summary of Results

Hypotheses	Prediction	Result
Hypothesis 1	Participants' working memory capacity scores will significantly correlate with fluency measures of L2 speech performance under no planning condition.	**Partially supported**. Working memory capacity scores significantly correlated with fluency only as measured by total pausing time for the experimental group under no planning condition.
Hypothesis 2	Participants' working memory capacity scores will significantly correlate with accuracy measures of L2 speech performance under no planning condition.	**Partially supported**. Working memory capacity scores correlated with accuracy as measured by number of errors per a hundred words and percentage of error free clauses for the control group only under no planning condition.
Hypothesis 3	Participants' working memory capacity scores will significantly correlate with complexity measures of L2 speech performance under no planning condition.	**Not supported.** Working memory capacity scores did not correlate with complexity under no planning condition.
Hypothesis 4	Under planning condition there will be greater fluency for the experimental group when compared to the control group under no planning condition.	**Not supported.** There were no significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of fluency.
Hypothesis 5	Under planning condition there will be greater accuracy for the experimental group when compared to the control group under no planning condition.	**Partially supported.** There are significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of accuracy as measured by the percentage of error free clauses.

Table 27

Continued

Hypothesis 6	Under planning condition there will be greater complexity for the experimental group when compared to the control group under no planning condition	**Supported.** There are significant differences in L2 speech performance in favor of the experimental group under planning conditions when compared to the control group under no planning conditions in terms of complexity.
Hypothesis 7	Participants' working memory capacity scores will significantly correlate with fluency measures of L2 speech performance under planning condition.	**Partially supported.** There were significant correlations between working memory capacity and fluency as measured by speech rate unpruned and pruned in the performance of task 2 (planning).
Hypothesis 8	Participants' working memory capacity scores will significantly correlate with accuracy measures of L2 speech performance under planning condition.	**Not supported.** There were no significant correlations between working memory capacity scores and accuracy in the performance of task 2 (planning).
Hypothesis 9	Participants' working memory capacity scores will significantly correlate with complexity measures of L2 speech performance under planning condition.	**Supported.** There were significant correlations between working memory capacity scores and complexity in the performance of task 2 (planning).
Hypothesis 10	Within the experimental group, under planning condition, higher working memory spans will significantly outperform lower working memory spans as regards fluency of L2 speech performance.	**Partially Supported.** Higher spans significantly outperformed lower spans in terms of fluency as measured by speech rate unpruned and pruned.
Hypothesis 11	Within the experimental group, under planning condition, higher working memory spans will significantly outperform lower working memory spans as regards accuracy of L2 speech performance.	**Not supported.** Higher spans outperformed lower spans in accuracy as measured by number of errors/100 words, however, **not significantly**; the difference in favor of the higher spans only approached significance
Hypothesis 12	Within the experimental group, under planning condition, higher working memory spans will significantly outperform lower working memory spans as regards complexity of L2 speech performance.	**Supported.** Higher spans significantly outperformed lower spans in terms of complexity
Hypothesis 13	When planning an oral task, learners will engage mainly in the processes of (1) organization of ideas, (2) lexical-grammatical search, (3) task rehearsal, and (4) monitoring.	**Partially Supported.** Learners engaged mainly in lexical search, organizational planning, monitoring, rehearsal, and also writing/outlining/summarizing and elaboration during pre-task planning

Table 28

Continued

Hypothesis 14	Higher and lower span individuals will differ in terms of the mental processes they engage when they plan.	**Supported.** Higher span learners employed significantly more metacognitive strategies than lower spans during pre-task planning

In the next chapter, I will discuss the results reported in the present section by addressing the research questions and hypotheses which guided the study by drawing on the literature reviewed in Chapter II.

CHAPTER V

DISCUSSION

5.1 Introduction

In this chapter, the results will be discussed by addressing the research questions and hypotheses in light of the literature in the fields of task-based planning, working memory, and speech production reviewed in Chapter II. The chapter is organized as follows. First, I will discuss results concerning the relationship between working memory capacity and L2 speech performance in no planning and planning conditions. Second, I will discuss results of the effects of planning on L2 speech performance. Third, I will discuss results concerning the differences between higher and lower spans on L2 speech performance under the planning condition. Finally, I will discuss the findings regarding the mental processes learners engage in when they plan by addressing: (a) strategies employed by learners during planning, (b) differences between higher and lower spans concerning the strategies employed. In all sections of the discussion in each one of these issues, I will start by summarizing the respective results.

5.2 The relationship between working memory and L2 speech performance in no planning and planning conditions

In this section, I will carry out some reflections upon the results concerning the correlations between working memory capacity and L2 speech performance in planning and no planning conditions. The first research question of the present study asked whether measures of working memory capacity would significantly correlate with

measures of L2 speech performance under no planning conditions. The third research question whether there would be significant correlations between measures of working memory capacity and L2 speech performance under the planning condition.

To reiterate, there were three instances of performance under no planning conditions in the present study. The first and second narrative tasks of the control group and the first narrative task of the experimental group; and there was only one instance of performance under planning condition, that is, the second narrative task of the experimental group.

In brief, results of the correlations between working memory capacity and L2 speech performance under no planning and planning conditions showed that:

1. There was a significant negative correlation between working memory capacity and fluency, as measured by total pausing time, in the performance of the experimental group under no planning condition.
2. There was a significant correlation between measures of working memory capacity and accuracy of L2 speech performance, as measured by the number of errors per a hundred words and percentage of error free clauses, in the performance of the control group under no planning conditions in both Task 1 and Task 2.
3. There were no significant correlations between working memory capacity and complexity in the performance of control and experimental groups under no planning conditions.
4. There were significant correlations between working memory capacity and fluency, as measured by speech rate unpruned and pruned, in the performance of the experimental under planning condition.

5. There were no significant correlations between working memory capacity measures and accuracy in the performance of the experimental group under planning condition; the correlations between working memory and accuracy only approached significance.
6. There were significant correlations between working memory capacity and complexity in the performance of the experimental group under planning condition.

I will address these results as follows. First, I will discuss the correlations between working memory capacity and L2 speech performance under no planning condition. Second, I will provide tentative explanations as to why there were significant correlations between working memory and *only one* aspect of speech performance under no planning conditions, *fluency* for the experimental group and *accuracy* for the control group. In other words, I will attempt to explain why correlations between working memory and the other aspects of L2 speech performance failed to reach significance. Third, I will also provide a tentative explanation as to why there were differences in group orientation, that is, why working memory correlated with *fluency* for the experimental group and with *accuracy* for the control group under no planning condition. Finally, I will discuss results of the correlations between working memory and L2 speech performance for the experimental group under the planning condition.

As previously stated, there was a significant negative correlation between working memory capacity and fluency, as measured by total pausing time, in the performance of the experimental group under no planning condition. In other words, learners with higher working memory spans produced a lower percentage of total pausing time. These significant correlations between working memory and fluency

corroborate previous results in the literature (Fortkamp, 1999, 2003; Mizera, 2006; Xhafaj, 2006).

Fortkamp (1999) found correlations between working memory capacity and fluency as measured by speech rate, and Fortkamp (2003) found correlations between working memory and fluency as measured by speech rate, mean length of run, and number of silent pauses per minute. Mizera (2006) found correlations between working memory and fluency as measured by speech rate. Xhafaj (2006) found correlations between working memory and fluency as measured by frequency of within boundary pauses and mean length of run.

According to Engle et al. (1999), in order to provide an understanding on the relationship between working memory capacity and performance in other cognitive tasks, it is crucial to specify what processes are tapped by the working memory span test that are also tapped by the other cognitive task at hand – L2 speech performance, as in the case of the present study.

Bearing in mind that, in the present study, fluency was conceptualized as the ability to cope with real time communication (Skehan, 1996, 1998), and it was operationalized in terms of temporal measures, it is crucial to explain what processes are involved in fluent L2 speech that are also involved in the Speaking Span Test.

According to Daneman (1991), fluent speech requires a skillful coordination of the storage and processing components of speech. Fortkamp (1999) postulates that working memory coordinates the execution of processes – establishment of communicative intention, conceptualization of the message, formulation of the message – as well as the storage of the products of these processes – preverbal message, surface structure, and the phonological plan.

A skillful coordination of storage and processing components is also required in the performance of the Speaking Span Test in which participants must keep words highly activated in an accessible state and then recall these words in order to produce sentences which are grammatically and semantically accepted containing each recalled word. According to Kane, Conway, Hambrick, and Engle (2007), working memory span tests tackle the ability to control attention by requiring subjects to maintain or recover access to task relevant information while that access or recovery is confronted by the shifting of attention between the storage and processing components of the tasks.

According to Kane et al. (2007), working memory capacity is important in controlled a processing activity, that is to say, in tasks in which attentional control is required. Levelt (1989) postulates that conceptualizing, message construction, and monitoring involve controlled processing whereas formulation and articulation are highly automatic. As for conceptualizing and message construction, speakers do not have a fixed slot of intentions to convey; thus, communicative intentions can fluctuate in countless ways. However, Levelt (1989) also claims that; in adult speakers; not all conceptualization and message construction is under executive attentional control since the adult speaker's experience is so vast that it allows whole messages to be available in long-term memory, which will be retrievable without much effort.

It is important to highlight that Levelt's (1989) model accounts for mature *L1* speakers. If one takes an *L2* speaker into account, conceptualizing and message construction may be even more harshly under attentional control since an L2 speaker's experience is not so vast when compared to an L1 speaker. In the case of L2 speakers at the intermediate level as it is the case of the present study, even formulation is not as highly automatic as it is for L1 speakers. In this sense, Fortkamp (2003) proposes that

the formulation stage of *L2* speech production, particularly at the level of grammatical encoding, is a controlled processing activity.

Bearing that in mind, some lines of thought can be put together at this point. First, conceptualizing, message construction, and monitoring are the components of speech which draw more heavily on working memory because they require attentional control (Levelt, 1989), and, according to Fortkamp (2003), L2 formulation also requires attentional control. Second, fluent speech involves continuous speech in real time communication, which implies a high degree of automaticity. Moreover, fluent speech involves effective coordination of *all* the stages of speech, some of which (e.g. message construction, formulation, and monitoring) require attentional control. In order to speak fluently, one must establish communicative intentions, construct messages, formulate, and articulate.

In the present study, fluency was operationalized in terms of temporal variables, that is, speaking in real time communication. Thus, I am inclined to believe that what seems to account for the correlations between *working memory capacity* and *fluency* for the experimental group is not concerned *mainly* with articulation or monitoring, but *particularly* the ability to control attention during conceptualization, message construction, and formulation so as to allow continuous speech, in real time communication, to take place.

The correlations between working memory capacity and fluency indicate that learners with higher working memory capacity have more attentional resources available to allocate towards the storage and processing components tackled by the Speaking Span Test and L2 speech production tasks. These results suggest that higher spans tend to be more able to cope with the cognitive loads of L2 speech production

and, thus, tend to be more able to sustain continuous performance in real time communication (Fortkamp, 2003).

What is intriguing about the findings of the present study concerning the relationship between working memory and fluency is that significant correlations were revealed *only* for fluency as measured by means of total pausing time. At first glance, I would expect that participants who produced a statistically significant lower percentage of total pausing time would also produce statistically significant fewer pauses and, thus, attain statistically significant faster speech rate. Therefore, I would expect that correlations between total pausing time and working memory would also reflect correlations between speech rate and working memory as well as number of pauses per c-unit and working memory.

However, when checking intercorrelations among fluency measures, results revealed that speech rate unpruned and pruned correlated significantly with number of pauses per c-unit: (r = -748**, p = 000) and (r = -778**, p = 000), respectively. However, the percentage of total pausing time did not correlate significantly with speech rate unpruned (r = -335, p = .102), speech rate pruned (r = -323, p = 115), or with number of pauses per c-unit (r =182, p = 383).

Bearing in mind that total pausing time did not correlate with speech rate or with number of pauses per c-unit, a possible explanation for correlations between working memory and total pausing time is that total pausing time was significantly lower for higher spans not because they produced statistically significant fewer pauses per c-unit, but because they produced pauses of shorter length. Moreover, the fact that correlations between working memory and total pausing time did not reflect correlations between working memory and speech rate may be due to a possible increase in the use of hesitations (e.g., filled pauses).

Fortkamp (2003) reported a trade-off between silent pauses and hesitations. She found significant negative correlations between silent pauses and working memory, but significant positive correlations between hesitations and working memory. In other words, Fortkamp (2003) found that higher spans produced fewer silent pauses but relied extensively on the use of hesitations. Possibly, in the present study, higher spans of the experimental group produced silent pauses with shorter length but made use of more hesitations which may have impacted negatively on speech rate and thus there were no significant correlations between working memory capacity and speech rate.

The present study did not assess fluency in terms of hesitations but the inclusion of such measure would have provided a more comprehensive picture of the relationship between working memory capacity and fluency for the participants of the experimental group under no planning condition.

Still related to performance under no planning condition, besides the significant correlations between working memory and *fluency* as measured by total pausing time in the *experimental group*, there were also significant correlations between working memory and *accuracy* as measured by number of errors per one hundred words and percentage of error free clauses in the performance of the *control group* in Tasks 1 and 2. In other words, within the control group, higher spans made fewer errors per one hundred words and produced more error free clauses. These results corroborate previous results found in the literature (Bergsleithner, 2007; Fortkamp, 2003; Mizera, 2006;[20]).

If on the one hand fluent speech performance implies automaticity, which encompasses managing all stages of speech production in an effective way so as to allow continuous speech in real time communication to take place, accurate speech

[20] Although Mizera (2006) did not investigate accuracy, one of his tasks to assess fluency, The Imitation Grammaticality Task , actually involved an element of grammatical accuracy since participants were supposed to imitate and correct errors they detected in samples of exchanges in Spanish.

performance, on the other hand, implies monitoring, which is considered a cognitively demanding process (Level, 1989). Monitoring demands attentional control, in which the speaker attends to his/her own internal and overt speech (Levelt, 1989). Therefore, what seems to account for the correlations between working memory capacity and accuracy is *not mainly* the ability to control attention during conceptualization, message construction, and formulation (as in the case of fluency), *but particularly* the ability to control attention during formulation and monitoring.

Along the same lines, Rosen and Engle (1997) provided evidence that individuals with higher working memory capacity tend to be more able to engage in self-monitoring, which may explain why participants with higher working memory capacity were the ones producing fewer errors and more error free clauses.

One striking issue about the findings of the relationship between working memory and L2 speech performance under no planning conditions is the fact that, for both the control and the experimental group, working memory correlated with *only one* aspect of L2 speech performance: only fluency in the experimental group and only accuracy in the control group. The questions deserved to be asked seem to be: Why did working memory correlate significantly with fluency in the performance of the experimental group but failed to correlate significantly with all the other aspects of performance? And why did working memory correlate significantly with accuracy in the performance of the control group but failed to correlate significantly with the all other aspects of performance?

In the realm of memory research, working memory capacity is important in tasks which require attentional control (Engle et al., 1999; Hambrick & Engle, 2003; Heitz, Unsworth, & Engle, 2005; Kane et al., 2007). In other words, working memory capacity refers to attentional processes in charge of maintaining relevant information in an active

and easily accessible state under conditions of interference, distraction, conflict, or competition (Kane, et al.).

In the area of task-based planning research, fluency, accuracy, and complexity are claimed to compete for learners' limited attentional resources, which leads to trade-off among these aspects of performance (Skehan, 1996, 1998; Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003). In other words, because attentional resources are limited, it is unlikely that learners will sustain simultaneous high levels of performance in terms of fluency, accuracy, and complexity.

Bearing that in mind, it seems reasonable to argue that Skehan's proposal of attentional trade-offs among fluency, accuracy, and complexity in the context of learners' limited attentional resources is compatible with Engle's attentional view of working memory capacity. Most studies on planning have shown trade-off effects among the goals of fluency, accuracy, and complexity. The body of research results tends to show that fluency and complexity tend to improve at the expense of accuracy (Mehnert, 1998). Results of the present study show that, under no planning conditions, there are significant correlations between working memory and fluency in the performance of the experimental group. Possibly, greater fluency was achieved by higher spans at the expense of accurate and complex speech production.

Results of the present study also show, that under no planning condition (in both Task 1 and 2), there are significant correlations between working memory and accuracy in the performance of the control group. Possibly, greater accuracy was achieved by higher spans at the expense of fluent and complex speech production. As previously stated, most planning studies provide evidence that under *planning* conditions, fluent and complex speech are achieved at the expense of accurate speech. However, performing under a *no* planning condition tends to be more difficult for most learners

and, thus, it is possible that learners will focus on *one* aspect of L2 speech at the expense of others.

Possibly, performing the tasks under no planning condition was too difficult, and may have led learners to some degree of discomfort or nervousness. Participants of the present study reported that performing tasks under the no planning condition was difficult. The following questionnaire excerpts illustrate learners' voices reporting the difficulties they faced:

Excerpts

"It was very difficult for me because I didn't know what to say" (p29)

"It was difficult…I couldn't elaborate a good story" (p6)

"It was difficult to create a story as you tell it at the same time" (p27)

In addition to that, learners also reported being nervous when performing the first narrative task, as the following excerpts illustrate:

Excerpts

"It was very difficult to me to tell a story immediately after looking at the because I'm not confident in my English, in fact, I know I still have a bad English" (p02)

"I guess it was Ok, the big problem was that I got too scared and it didn't get the way I really wanted, but that's okay" (p25)

"I was a little nervous and at this point I forgot vocabulary, simple vocabulary, deu branco" (p09)

In the realm of affective/emotional variables, research has shown that anxiety may affect performance when a task is hard or when performance is under evaluation (Lee, 1999). The performance of the tasks under no planning condition may have led to some degree of anxiety from the part of the learners since they reported being nervous or worried about task performance. Research has shown that anxiety may lead learners to engage in negative internal dialogues or worrisome thoughts about themselves or

about their performance, and these thoughts may actually interfere with working memory performance because some portion of capacity is directed at such thoughts (Eysenck, 1992). Possibly, working memory capacity correlated with only *one* aspect of performance under no planning condition because worrisome thoughts may have been at play.

Klein and Boals (2001) claim that stressful and worrisome thoughts work as distracters that need to be inhibited so that attention can be maintained on the task being performed. Likewise, Unsworth, Heitz and Engle (2005) claim that individuals who differ in working memory capacity will also differ in the capacity to inhibit thoughts called to mind by stress and task manipulations. In other words, working memory capacity is related to the ability to inhibit unwanted thoughts. Possibly, higher spans were better able to inhibit such unwanted thoughts during performance; thus, correlations between working memory capacity and L2 performance could emerge, at least for one aspect of performance.

Mizera (2006) also reported lack of correlations between working memory and some aspects of L2 speech performance. In his view, the complexities involved in L2 speech performance may involve factors other than working memory capacity. He claims that personal and affective factors may also play a role in L2 speech.

Interestingly, participants of the pilot study (Guará-Tavares, 2006) also reported some discomfort and difficulties when performing the narrative tasks under no planning conditions. In Guará-Tavares (2006), there were no correlations *at all* between working memory and L2 performance in task under no planning conditions. Thus, one question which merits to be addressed is: Why task difficulty prevented the emergence of individual differences in working memory under no planning conditions for *all*

performance aspects in Guará-Tavares (2006) but still yielded individual differences in working memory for at least *one* aspect of performance in the present study?

In the attempt to answer the question just posed, I think it is important to bring the distinction between task complexity and task difficulty into play. Although, task difficulty and task complexity mean the same in Cognitive Psychology, they are slightly different in SLA. According to Robinson (2001), the factors contributing to task complexity are related to design features, such as 'here-and-now' or 'there-and-then', and planning or no planning. Robinson (2001) emphasizes that "these factors need to be distinguished from the learner factors contributing to task difficulty" (p. 295). Task difficulty is related to learners' perceptions of the task and may be determined by affective factors such as anxiety and motivation, and also by ability factors such as aptitude and proficiency (Robinson, 2001).

In this sense, it is possible to manipulate task complexity, as I have attempted to do in the present study and in the pilot study by using 'there-and-then' tasks so that individual differences in working memory capacity would be likely to emerge. However, "affective variables contributing to task difficulty are hard, or impossible to diagnose in advance" (Robinson, 2001, p. 295), as it is the learner who asserts it.

Bearing the distinctions between task complexity and task difficulty in mind, it seems plausible to state that the 'there-and-then' narrative tasks may have been *extremely difficult* for participants of the pilot study, thus, individual differences in working memory capacity did not emerge because learners may have performed the tasks beyond the limits of their cognitive resources. On the other hand, the same 'there-and-then' tasks may have been *difficult* for participants of the present study but not to the same degree as for learners of the pilot study; thus, individual differences in working memory capacity could emerge, at least concerning one aspect of L2

performance, that is, accuracy for the control group and fluency for the experimental group.

For individual differences in working memory capacity to emerge, the task under performance has to be difficult (Fortkamp, 2000; Conway et al., 2005; Just & Carpenter, 1992; Tomitch, 1996). Tasks which are either too easy or too difficult do not seem to reveal individual differences in working memory capacity.

One question to be addressed concerning conflicting results of the pilot and the preset study is: Why were tasks more difficult for participants of the pilot study? A tentative explanation may be level of proficiency. Although participants of the pilot study were also considered intermediate learners according to the criteria of the rating scale proposed by D'Ely and Weissheimer (2005), the means of participants' performance in the pilot study was 2.5 whereas the means of participants' performance in the present study was 2.95. It is important to highlight that, in both the pilot study and the present one, the raters in the selection of participants were the same. Moreover, participants of the pilot study had an even shorter length of time to look at the pictures before performance, only 40 seconds.

Finally, the fact that, under no planning conditions, working memory capacity correlated significantly with *accuracy* for participants of the control group, but correlated with *fluency* for participants of the experimental group is another striking finding which merits a reasonable speculation. What would be a possible explanation for the correlations between working memory and *different* aspects of performance for these two groups (control and experimental)? Why did learners in the control group tend to prioritize *accuracy* under no planning condition? Why did learners in the experimental group tend to prioritize *fluency* under no planning condition?

Tentative explanations for this difference in what learners seem to have prioritized in the performance under no planning condition may be found both in the areas of working memory and SLA. According to Ellis (2003), it is the learner who decides what kind of 'activity' to engage in during performance, and such choices determine what to prioritize. The first question to be addressed seems to be: Is the 'choice' learners make towards what aspects to prioritize a deliberately conscious choice or is it triggered automatically?

According to Feldman-Barrett et al. (2004), "although attentional control can sometimes occur with a feeling of conscious deliberation and choice, it need not" (p. 555). These authors claim that controlled attention may be at play even in early perceptual stages affecting how information is selected and processed before subjective experience (deliberate conscious choice) takes place. In other words, Feldman-Barrett et al. claim that a stimulus (e.g. a task) may capture attention in a reflexlike fashion. However, these automatic forms of attention are reliant on more controlled forms of attention. The reflexive allocation of attention tends to take place more easily when individuals attend to features of a stimulus.

In order to exemplify such claims on automatic and controlled forms of attention, the authors bring evidence from priming[21] studies. Priming procedures activate knowledge representations without participants' awareness but for priming to activate a representation, it is necessary that individuals attend to words on a computer screen. How does these automatic and controlled forms of attention relate to learners' 'choices' on which aspects of performance to prioritize?

[21] One of the tasks used in priming studies require participants to read several lists of words on a computer screen and state which words have been presented a priori and which ones have not.

If we take the narrative task as a stimulus, the 'choice' on which aspects should be prioritized during the performance of the task may have been triggered in a reflexlike fashion. However, for this 'choice' to be triggered, learners had to attend to the pictures of the tasks, make sense of the tasks, and engage in the oral performance of the tasks. Then, one question to be pursued is: What causes 'choices' on what to prioritize to take place in a reflexlike fashion without deliberate effort? According to Feldman-Barrett et al. (2004), "properties of the external world can influence properties of the internal world (e.g., goals and motivations), which, in turn, proceed to influence processing and guide behavior in a reflexive way" (p. 555). Bearing that in mind, it can be argued that properties of the external world – e.g. the environment in which the task is being carried out – may predispose learners to act according to their previous experience and background. The narrative tasks of the present study were carried out in a language laboratory in the language school where participants attended classes. Possibly, learners' previous experiences and background may have evoked reflex like choices. According to Batstone (2005), learners' background may predispose them towards prioritizing fluency and/or accuracy.

Although participants of the present study attended L2 classes in the same learning context at the time of data collection, it is important to remark that the Extracurricular Language Courses have students and teachers from all over the country, which makes it likely that this L2 learning context encompasses some degree of variety of educational backgrounds in terms of teaching and learning styles, orientation, and attitudes.

Possibly, learners in the control group come mainly from backgrounds in which emphasis on form is prominent. Perhaps, in their previous L2 learning experiences, attention to formal aspects of the language and error free performance were pervasive

due to the styles of their teachers, the course books which were used, the means of assessment which were commonly conducted, and the like. On the other hand, learners in the experimental group possibly come mainly from backgrounds in which attention to meaning was pervasive. Perhaps, in the course of their L2 experiences, they had teachers who emphasized communication, getting the message across *over* correctness and conservative error free performance.

In the present study, there was no attempt to tackle learners' background and previous experiences in their L2 learning process. There was no attempt to examine their perceptions on oral performance, what effective performance means to them, nor was there any attempt to scrutinize the style of their present and previous teachers and/or course books. Efforts in these directions may provide a better understanding of learners' predispositions on what aspects of performance to prioritize. Ortega (2005) claims that some learners seem to be oriented towards form whereas others tend to be oriented towards meaning. The control and experimental groups clearly presented different orientation, which may have been determined, at least in part, by their learning backgrounds.

Based on what has been said, it is feasible to argue that learners' 'choices' on what to prioritize may be triggered in a reflexlike fashion, without deliberate conscious effort when they attend to the performance of the task at hand. However, it is the capacity to control attention among the various components of L2 speech that will sustain these reflexlike 'choices' during ongoing performance.

Up to this point, I have discussed results of the correlations between working memory capacity and L2 speech performance under no planning condition. Now I turn to the discussion on the relationship between working memory and L2 speech performance under planning condition.

Under planning condition, results revealed significant correlations between working memory capacity and fluency as well as significant correlations between working memory capacity and complexity. Correlations between working memory capacity and accuracy only approached significance. Taken together, these correlations show that under planning conditions higher span individuals are the one whose speech performance is significantly more fluent and complex.

Interestingly, under planning condition individual differences in working memory were related to more aspects of L2 performance when compared to the no planning condition. Recall that under no planning condition, working memory correlated *only* with fluency for the experimental group and correlate *only* with accuracy for the control group. It could be argued that planning made the task more manageable, that is to say, performing the task under planning condition was not as difficult performing it under no planning condition. It seems that the task was difficult enough for individual differences in working memory capacity to emerge more fully.

Recall that fluent speech involves continuous speech in real time communication, which implies some degree of automaticity and involves effective coordination of *all* the stages of speech- (e.g., conceptualization, message construction, formulation, monitoring, and articulation). Fluent speech was operationalized in terms of temporal measures, real time communication. Thus, I am inclined to believe that what seems to account for the relationship between working memory and fluency is not concerned *mainly* with monitoring and articulation, but *particularly* the ability to control attention during conceptualization, message construction, and formulation effectively, so as to allow continuous speech, in real time communication, to take place.

Moreover, the benefits of planning may also rely on the ability to implement what was planned into performance (Ortega, 2005). In other words, the benefits of

planning may also draw upon the ability to retrieve what was planned and implement it into online performance. Individual differences in working memory capacity reflect differences in the ability to retrieve information from long term memory (Rosen & Engle, 1997; Unsworth & Engle, 2007). Therefore, it may be that higher memory spans were more able to retrieve what was planned into real time performance and, thus, achieved higher fluency. This issue will be further discussed when I address the differences between the performance of lower and higher spans under planning condition.

Under planning condition, besides the significant correlations between working memory capacity and fluency, there were also significant correlations between working memory capacity and complexity. According to Skehan (1996), complexity is related to restructuring and regards "the process by which the interlanguage system becomes more complex, elaborated and structured" (p.47). Complexity implies risk taking performance in the attempt to produce more elaborated, cutting edge language.

As previously explained in the Review of the Literature, Skehan (1998) postulates that, in L2 learning and use, learners draw upon a dual-mode processing system, which encompasses the rule-based and the exemplar-based systems. Complex language production implies drawing upon the rule-based system, which prioritizes analyzability, leads to a form-oriented organization that regards development in terms of change and complexity and, according to which, interlanguage development is the outcome of restructuring.

Recall that Feldman-Barrett et al. (2004) also acknowledge the coexistence of two modes of processing. The associative (exemplar) mode functions on the bases previous existing representations in which information is processed automatically. Thus, associative processing is not under the constraints of limitations in working memory

capacity. Rule-based processing, on the other hand, involves symbolic representations, concerns incorporating new or inconsistent information into preexisting representations and, thus, is more harshly under the constraints of attentional control.

Pre-task planning releases the processing load and allows learners to access the upper limits of their interlanguage in the attempt to produce more complex and elaborate language (Crookes, 1989). Since complex speech involves drawing upon the rule-based system and since rule-based processing is under the constraints of attentional control, this may explain why higher spans were the ones who achieved more complex speech under planning condition.

Based on what has been said, what accounts for the correlations between working memory capacity and complex L2 speech? The ability to control attention in the Speaking Span Test, which requires learners to activate words and maintain these words activated and accessible for recall while processing sentences containing the words recalled, parallels the ability to control attention in rule-based processing necessary for complex language production. In complex language production, learners need to activate preexisting representations and maintain them activated and accessible while processing inconsistent representations (e.g. cutting edge language the learner is not sure about) and incorporate this edging information into preexisting representations. According to Feldman-Barrett et al. (2004), rule-based processing is under the constraints of working memory capacity limitations, which may explain why higher spans produced more complex language.

It is important to highlight that these correlations between working memory and performance under planning condition indicate that the higher the memory, the higher the fluency and the complexity. Nevertheless, these correlations do not reveal whether the differences between the performance of lower and higher spans were significant.

This issue will be addressed later in this chapter when I discuss results based on the extreme-group design.

In this section, I have addressed the relationship between working memory and L2 speech performance under planning and no planning conditions. Now I turn to the discussion on the impact of planning on L2 speech performance.

5.3 The impact of planning on L2 speech performance

This section deals with the impact of planning on performance of the experimental group as a whole, *regardless* of individual differences in working memory capacity. To reiterate, the second research question of the present study asked whether pre-task planning would lead to significant increase in fluency, accuracy, and complexity in the performance of the experimental group when compared to the control group. As shown in the previous section of this chapter, all means of L2 speech performance measures in the second narrative task favor the experimental group when compared to the control group.

However, only differences in two of these measures achieved statistical significance: accuracy as measured by the percentage of error free clauses and complexity as measured by the number of clauses per c-unit. Differences in accuracy as measured by the number of errors per a hundred words and differences in fluency as measured by speech rate unpruned, speech rate pruned, number of pauses per c-unit, and total percentage of silent pausing time all failed to achieve significance.

In most studies on task based planning, results have shown a stronger impact for fluency (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999) and complexity (Crookes, 1989; Foster & Skehan, 1996; Ortega, 1999; Yuan & Ellis, 2003). Results

have been more mixed for accuracy. According to Ellis (2005), planning leads to gains in accuracy according to the grammatical features being used (Ellis, 1987; Ortega, 1999), different task types (Foster & Skehan, 1996), and different planning conditions (Mehnert, 1998). Most studies show that gains in fluency and complexity may be achieved at the expense of accuracy (Mehnert, 1998). In this sense, the results of the present study do not corroborate previous findings since the impact of planning was stronger for accuracy and complexity.

This stronger impact for accuracy and complexity is intriguing since, according to Crookes (1989), "it is unlikely that learners who produce more complex speech than they are normally capable of will at the same time maintain a given level of accuracy" (p.379). In other words, as learners take risks in the attempt to produce more complex language, chances are that they will be less prone to avoid errors (Crookes, 1989).

Foster and Skehan (2001) suggest some possibilities towards understanding this stronger effect for accuracy and complexity. According to them, the activities that take place during pre-task planning and the mental processes in which learners engage in are crucial for understanding the impact of planning on performance.

These researchers claim that efforts allocated towards different mental activities entail distinct benefits to performance. In the case of rehearsal, it tends to be mostly language oriented and is likely to affect accuracy (Foster & Skehan, 2001). As for efforts allocated towards retrieval operations, they lead to benefits in complexity by making available a wider language repertoire, allowing learners to access the upper limits of their interlanguage (Crookes, 1989; Foster & Skehan, 2001).

As shown in Table 24, the most frequent strategies employed by learners during pre-task planning were lexical search (96%), writing/summarizing/outlining (84%), organizational planning (64%), monitoring (60%), rehearsal (44%), and elaboration

(40%). These results concerning strategies employed by learners seem to corroborate most of the results reported by Ortega (2005).

According to Ortega (2005), these strategies point at the emphasis on retrieval and rehearsal operations during pre-task planning. Since rehearsal leads to benefits on accuracy (Foster & Skehan, 2001), and retrieval leads to benefits on complexity (Crookes, 1989; Foster & Skehan, 2001), the frequent use of these strategies during pre-task planning provides a path for understanding why there was a stronger effect of planning on accuracy and complexity. The following excerpts illustrate some instances of retrieval and rehearsal operations:

Excerpts

> "I'm improving my sentences for example, I said 'they started to talk' and now I'm saying 'it seems that they started to talk about bla bla bla' I said that 'the wife was saying something' and included 'she was saying horrible things'…" (p01)
>
> "There is a thing I'm not sure, I'm thinking…if the term 'even' can be used substituting the negation not in a negative statement like for example, she doesn't give importance to his presents and even to him or not even to him or she appeared in a car and she even looked at him or not even looked at him…" (p23)
>
> "I'm reading, if I read it again I will imagine the story in my head so I can remember when I tell" (p24)
>
> "I think in Portuguese so the position was wrong, I wrote therapy couple but it is couple therapy" (p01)

As can be seen in the first excerpt, the learner is trying to improve sentences during task planning. The learner is focusing on lexical retrieval and is trying to add some adjectives to her story. The first sentences produced were correct "*the wife was saying something*" and *"they started to talk*". Nevertheless, it seems that the learner is trying to go beyond correctness in the attempt to produce more elaborated sentences such as *"she was saying horrible things"* and *"it seems that they started to talk about…"*

In the second excerpt, the learner actually verbalizes uncertainty about the language being used "*there is a thing here I'm not sure*", which suggests that the

learners is trying to use cutting edge language. First, she uses the negative "*she doesn't give importance to his presents*", which suggests that the learner is able to use the negative properly. She could have simply used the negative again and produced something such as 'she doesn't give importance to his presents and she doesn't give importance to him'.

Nevertheless, she preferred to venture using language she was not sure about, which suggests that the learner was trying to assess the upper limits of her interlanguage, pushing output in the attempt to produce more elaborate language by using the word 'even' in her sentences. It seems that she is actually taking the risk of using this word so as to embellish, elaborate the narrative.

In the third excerpt, the learner is engaged in reading and mental rehearsal *"I'm reading if I read again I will imagine the story in my head*", and it is not the first time she is rehearsing as she actually verbalizes *"…if I read it again…"* The learner also verbalizes that reading will help her remember the story during performance.

In the last excerpt, the learner is focusing on form by monitoring word order. She engages in cross language analysis by comparing word order in L1 and L2, and is able to correct a mistake *"I wrote therapy couple but it is couple therapy"*.

As can be seen from the excerpts aforementioned, learners engaged in retrieval and rehearsal operations during pre-task planning, which might explain why there was an effect for accuracy and complexity under planning condition. These results corroborate Ortega (2005) in which learners also engaged preponderantly in retrieval and rehearsal strategies during pre-task planning.

According to Ortega (2005), the connection between retrieval and complexity seems to be corroborated in her two studies, Ortega (1995) and Ortega (1999). As for the link between rehearsal and accuracy, results are not as evident since there was no

effect of planning on accuracy in Ortega (1995) or in Ortega (1999); results were mixed concerning the effects of planning on accuracy. Learners could produce more accurate of noun- modifier agreement in planned narratives, but there was no difference in accuracy in article use. Results of the present study lend support to the link between retrieval and complexity as well as the link between rehearsal and accuracy. It is important to highlight, however, that these results are suggestive, not conclusive.

It is only possible to make stronger claims about such links between retrieval and complexity and between rehearsal and accuracy if studies are carried out to investigate specific connections between these variables, and if correlation analyses are conducted to examine whether individuals who engage in more retrieval operations are the ones producing greater complexity, and also to examine whether individuals who engage in more rehearsal operations are the ones who produce more accurate speech. Examining these specific connections is beyond the scope of this study.

The emphasis on retrieval and rehearsal also suggests why there were no significant effects on fluency. As pointed by Crookes (1989), as learners take more risks they tend to produce more errors. Since learners were able to take risks and still sustain accurate speech, effects on fluency failed to achieve significance.

Skehan (1998) claims that fluency, accuracy, and complexity compete for learners' attentional resources, and thus trade-off effects take place among these aspects of performance. Possibly, learners attained significantly more complex and accurate speech at the expense of producing significantly more fluent speech. Previous studies also give evidence for trade-off effects (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003) but in a different direction. The research results tend to show that planning impacts predominantly fluency and complexity at the expense of accuracy (Mehnert, 1998). In face of these conflicting results on the impact of planning on fluency reported

in previous studies and the results reported in this study, one question remains without an answer: Why did planning have no statistically significant effect on fluency in the present study?

Recall that, under no planning condition, there was a correlation between working memory capacity and fluency, which implies that participants of the experimental group *as a whole* tended to focus on fluency, and, thus, individual differences in working memory capacity emerged for this aspect of performance. Bear in mind also that, in the attempt to provide an explanation as for why the experimental group focused on fluency under no planning condition, I raised the possibility that the 'choice' of what aspects of performance to prioritize may be a reflexlike behavior that does not take place consciously. It may be a reflexlike behavior due to previous experiences in the course of language learning.

When learners of the experimental group had the opportunity to plan, it was not fluency, but accuracy and complexity which were prioritized. So, are learners inconsistent in what aspects to prioritize since they prioritized fluency under a no planning condition, but prioritized accuracy and complexity under a planning condition? I believe that because performing a task under no planning condition is more difficult and learners were under pressure to start performing right after having looked at the set of pictures, they prioritized fluency in a reflexlike fashion motivated by their previous experiences.

However, when planning opportunity was allowed, there was no longer the time pressure to start performing right after having looked at the pictures, and learners then could attend to aspects of language which could not be attended to under a no planning condition, in which their 'choices' were more automatic, taking place in a reflex like fashion. Therefore, I am inclined to believe that learners are not inconsistent on what

they prioritize. Rather, performance conditions (e.g. planning) seem to influence what they prioritize. Several learners verbalized that they viewed planning as a situation in which they were required to perform better, as the following excerpts illustrate:

Excerpts:

"When you plan, you are forced to do something better" (p.2)

"When I planned I felt more responsibility for doing something very good" (p.6)

"Planning helps but planning also… I need to perform better cause I have no excuse" (p.14)

It seems that planning triggered learners to search for more efficiency in performance, which was possibly reflected in more accurate and complex speech. I believe learners' 'choices' on what aspects to prioritize are not deliberate conscious choices. Rather, they reveal reflexlike behavior based on their learning backgrounds and on performance conditions. These 'choices' start in a reflexlike fashion, but it is attentional control that will be necessary to sustain such 'choices' (Feldman-Barrett et al., 2004), that is, learners will 'choose' what to prioritize as they attend to the tasks, make sense of them, and start performing them.

In brief, a tentative explanation for the lack of planning effects on fluency may be that learners of the experimental group tend to prioritize different aspects of performance vis-à-vis task conditions. When performing a task under no planning condition, learners of the experimental group *as a whole* seemed to prioritize fluency. However, when performing a task under a planning condition, the protocols revealed that they focused on the stages of conceptualization, formulation, and monitoring. They attended extensively to formal aspects of the language, aimed at using more elaborated language, and made more use of monitoring.

According to Ellis (2005), pre-task planning tends to impact mainly on conceptualization and formulation whereas online planning tends to impact mainly on formulation and monitoring. In the present study, however, learners focused on conceptualization, formulation *and* monitoring during pre-task planning; thus, fluency was penalized. These results are in line with those reported by D'Ely (2006), in which she claims that monitoring can be counter productive to fluency.

In addition, planning may have impacted more on fluency in the performance of higher spans; thus, it may not have impacted fluency in the performance of the experimental group *as a whole*. This will be further discussed in the next section.

5.4 Differences between L2 speech performance of lower and higher spans in planning condition

As previously explained in Chapter III, correlations between working memory capacity and L2 speech performance reveal that the individuals with higher working memory are the ones obtaining higher levels of performance in some aspects of L2 speech. To put it in simple words, correlations indicate that the higher the memory, the higher the performance. However, correlations do not reveal whether differences between higher and lower spans are significant. In the attempt to scrutinize differences in the performance of lower and higher spans in this study, an extreme-group design was adopted, and ANOVAs were computed to compare the performance of these two working memory groups.

In brief, results concerning whether higher spans outperform lower spans in L2 speech performance under planning conditions show that:

1. Higher spans significantly outperformed lower spans in terms of fluency as

measured by speech rate unpruned and pruned.

2. Higher spans significantly outperformed lower spans in terms of complexity as measured by number of clauses per c-unit.

These significant differences between the performance of higher and lower spans could be attributed to working memory only, regardless of planning. However, results displayed in Table 22 showed that there were *no* significant differences between higher and lower spans in the performance of the first narrative task under no planning conditions. Therefore, it seems reasonable to argue, again, that, once the task was made more manageable due to the opportunity to plan, individual differences could more fully emerge; thus, higher spans significantly outperformed lower spans in terms of fluency and complexity.

Interestingly, fluency was the dimension which yielded greater differences between higher and lower spans (f =8.676**, p = .011 and f= 9.473**, p =.008 for speech rate unpruned and pruned respectively), (f=6.725*, p= .021 for complexity). In other words, it was the ability to produce significantly more fluent speech while still maintaining significantly more complex, and marginally significantly more accurate speech that yielded greatest differences between higher and lower spans under planning conditions.

Now it seems reasonable to bring the discussion on the impact of planning on fluency back into the present scenario. It is important to highlight that there were no significant differences between lower and higher spans under no a planning condition; but fluency was the dimension which yielded the greatest differences between higher and lower spans under a planning condition, which suggests that there was *some* impact of planning on fluency. However, it seems that higher spans were more susceptible to

the impact of planning on fluency; thus, the *overall* impact of planning on fluency for the experimental group *as a whole* was reduced.

These results not only lend support to the issue of trade-off effects among the goals of fluency, accuracy, and complexity (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003) but also suggest that trade-off effects seem to be acute for learners with lower working memory capacity since higher spans significantly outperformed lower spans in terms of fluency when planning opportunity was provided.

Again, following Fortkamp (2000), it can be argued that under planning conditions, individuals with more working memory capacity have more attentional resources available to allocate towards the processes involved in L2 speech production as a controlled process activity, which may explain the results obtained: (a) correlations showed that higher spans produced significantly more fluent and complex speech, and (b) ANOVAs showed that differences between lower and higher spans were significant in terms of fluency and complexity.

Based on the findings that, under planning condition, higher spans outperformed lower spans in terms of fluency and complexity and that these differences can not be attributed to working memory only, but also to planning, it seems that higher spans were more able to benefit from the opportunity to plan performance of an oral task. Hence, one question mustn't remain unanswered: What is it that planning requires that higher span individuals are better able to cope with and, as a result, they are more able to benefit from planning?

To reiterate, planning is a *problem solving activity,* and according to Hambrick and Engle (2003), a problem is a goal which is not instantaneously achievable and whose most prominent feature is that although the initial state and the target are clear, how to convert the initial state into the target state is uncertain. In planning, the initial

state – start preparing oral performance of a story based on pictures – is clear, how to convert this initial state into the target state – accomplish the preparation of oral performance – is uncertain. To put it in simple words, learners know that they are supposed to start setting up their performance, but they do not know from the start *how* they are going to prepare their performance, what events will happen in the story, what sequence the story will have, what words will be used and so on.

According to Hambrick and Engle (2003), problem solving activities require "the ability to maintain goals, action plans, and other task-relevant information in a highly activated and accessible state, and when necessary, to inhibit activation of irrelevant or distracting information" (p.179). When planning an oral task, learners need to activate task-relevant information, maintain them activated and accessible until this information can be integrated to subsequent information in a coherent way; learners also need to sustain, maintain, and switch attention from the various components of the task (e.g., from meaning to form and vice-versa), suppress irrelevant L2 and L1 information, and monitor. It is the ability to control attention among the various components of planning that higher spans seem to be better able to cope with, which may explain why higher spans benefited more from planning and, thus, significantly outperformed lower spans in terms of fluency and complexity.

The fact that fluency was the dimension which yielded the greatest differences between lower and higher spans is an interesting finding which merits some reflection. Ortega (1999) claims that the extent to which planning leads to benefits on performance also depends on the ability to execute what was planned into online performance. In other words, it also depends on the ability to retrieve what was planned into real time performance. In this study, fluency was operationalized as the ability to perform in real time communication (Skehan, 1996, 1998), and it was measured by speed (speech rate)

and silence (pauses) measures. It seems reasonable to argue that successful retrieval may have aided implementation of what was planned into real time performance, thus, reflecting greater differences between lower and higher spans in terms of fluency.

Individual differences in working memory capacity reflect differences in the ability to retrieve information from long term memory (Rosen & Engle, 1997; Unsworth & Engle, 2007). Unsworth and Engle (2007) provide evidence that higher spans are more effective at retrieving task-relevant information in the face of interference whereas lower spans are more likely to lose access to task-relevant information since they are more susceptible to have their attention captured by distraction and to activate more irrelevant information.

Based on these findings concerning the role of working memory in retrieval, it seems plausible to argue that higher spans were more able to retrieve what was planned into real time performance. Therefore, fluency as assessed by means of real time communication measures was the dimension of speech which yielded the greatest differences between lower and higher spans when performing a task under planning conditions.

Besides the ability to retrieve what was planned into online performance, I believe the ability to implement new ideas online may also have enhanced the benefits of planning on the performance of higher spans. Several learners verbalized that they implemented new ideas during task performance. The following excerpts illustrate this feature:

Excerpts

> "I remembered but I also created new things too because I forgot something and to not don't say anything I invented something at the moment" (p1)

> "I remembered my plan but I created things because I forgot something and also had different ideas" (p11)

These excerpts provide evidence that learners used what was planned but also implemented new ideas online. In this sense, these excerpts show that, although planning assists performance by allowing learners to focus on aspects of speech performance a priori, it does not prevent spontaneity, which is a hallmark characteristic of speaking (Bygate, 2001a), to take place. According to Feldman-Barrett et al. (2004), changing representations online is achieved by rule-based processing since it requires incorporating new information into existing representations. Rule-based processing is under attentional control and may be affected by individual differences in working memory capacity.

Bearing the preceding discussion in mind, I am inclined to believe that a more comprehensive explanation for the relationship between working memory capacity and L2 speech performance under planning condition is that higher spans are *not only* more able to effectively allocate attentional resources towards the processes involved in L2 speech production during task performance, as argues Fortkamp (2003) *but also* more able to cope with the processes involved in planning as a problem solving activity (Hambrick & Engle, 2003), more able to retrieve what was planned into performance (Rosen & Engle,1997; Unsworth & Engle, 2007b) and more able to implement new ideas online (Feldman-Barrett et al, 2004).

The last question to be pursued in this section is: Why did higher spans significantly outperform lower spans in fluency and complexity but not accuracy? I will put forward two tentative explanations. First, it could be due to trade-off effects. Higher spans were more able to achieve significantly more fluent and complex speech, when compared to lower spans, at the expense of achieving more accurate speech. Second, Feldman-Barrett et al. (2004) propose that, in complex tasks, lower spans may have a range of goals; however, they lack sufficient attentional resources to maintain goal-

relevant processing in complex situations. As a result, they end up devoting attention to efficiency over any other processing goal. Possibly, lower spans view error free performance as efficient performance and pursued a more conservative approach to L2 speech under planning condition. Consequently, the differences between higher and lower spans in terms of accuracy did not reach statistical significance.

In this section, I have discussed results concerning the differences in the performance of lower and higher spans under planning condition. In the next section, I will discuss results vis-à-vis the mental processes learners engage in when they plan performance of an oral task.

5.5 The mental processes learners engage in when they plan

In this section, I will address the issue of how planning assists performance by discussing what processes learners engage in when they plan. First, I will recap and discuss these results within the whole experimental group. Then, I will focus on the differences between higher and lower spans during pre-task planning. In short, results show that:

1. Learners engaged mainly in lexical search, writing/outlining/summarizing, organizational planning, monitoring, rehearsal, and elaboration during pre-task planning;

The fifth research question asked what mental processes learners engage in when planning performance of an oral task. This question was addressed in terms of the strategies employed by learners during planning. As shown in Table 24, the strategies most frequently reported by learners were lexical search (96%), writing/summarizing/

outlining (84%), organizational planning (64%), monitoring (60%), rehearsal (44%), and elaboration (40%). These results corroborate those reported by Ortega (2005) concerning organizational planning, writing/summarizing/outlining, lexical searches, rehearsal, and monitoring. However, in Ortega's study, translation and empathizing with the listener were also frequently reported by learners.

As regards empathizing with the listener, participants of the present study did not have a listener present while they performed their planned narratives. I was present during planning time to carry out the retrospective online documentation of what they were planning. Nevertheless, I left the room after planning was over so that they would be comfortable to tell their narratives. This may explain why there were no instances in which learners verbalized any concern with the listener.

Concerning translation, it was frequently reported in Ortega's study but only two learners of the present study reported this strategy. It is important to highlight that in Ortega's (2005) study, learners were given pictures and also listened to a recorded version of the stories in their L1 before retelling. This may have biased learners to rely more on translation during retelling of their narratives (Ortega, 2005). Learners of the present study were asked to tell stories based on pictures only.

Apart from translation and empathizing with the listeners, strategies most frequently reported by learners of the present study corroborate Ortega's findings and point at the emphasis on retrieval and rehearsal operations during pre-task planning. As mentioned earlier in this chapter, retrieval and rehearsal operations are likely to aid complexity *and* accuracy, respectively (Foster & Skehan, 2001).

The protocols revealed that learners try to have a general organization of ideas before they actually think of the specific formal aspects of the language they are going to use. At the beginning of planning, they often referred to the pictures, focused on what

happened in the stories as if they were tying to decide on the content of their narratives. Such mental operations seem to rely upon the conceptualizer, in which the message content is planned (Levelt, 1989). The following excerpts illustrate these instances.

Excerpts

"I was thinking to organize my stories according to the pictures" (p10)

"I'm thinking of each picture and a general comment about them" (p25)

"I was just thinking that the story is about a couple and about what the husband is thinking" (p01)

As these excerpts show, learners seem to be focusing on the general organization of their stories and trying to set their communicative goals before they actually concentrate on more specific aspects of language. First, learners seem to have an overall organization of ideas by focusing on the content of their stories.

As learners move on to more specific aspects of language, the strategy most frequently reported was lexical search. All learners reported a concern with finding proper lexical items to use in their stories. This ubiquitous focus on words is in line with the claim that speech production is lexically driven, that is, knowing words is the paramount condition for expressing communicative ideas (Levelt, 1989). Such lexical searches in which learners engaged draw upon formulation at the level of grammatical encoding, more specifically in *lexical selection*, which involves the identification of lexical concepts that are suitable for conveying the speaker's meaning (Bock, 1995; Bock & Levelt, 1994; Levelt, 1989).

When searching for words, learners would either remember the words and include them in the planning of their narrative tasks or notice a gap in their interlanguage (Swain, 1985) and, consequently, avoid the unknown words by changing

the intended messages or keeping the messages but substituting words. The following excerpts illustrate these instances:

Excerpts

"How to say pedaço de Madeira in English…palavras do tipo bater, jogar agora eu já lembrei"(p2)

"I'm thinking that I don't remember how to say 'ervilha' in English and I will change it to another word…beans" (p7)

"I tried to remember 'ter coragem' but I will change for he did not get to reply or to give a response to her" (p4)

As the first excerpt shows, the learner was able to retrieve the lexical items being searched, whereas in the subsequent excerpts learners were not able to find the lexical items being searched. Participant 07 substituted the word *ervilha* for beans, participant 04 substituted a whole sentence.

After setting the general content of the stories and focusing on some formal aspects of the language in order to convey their communicative ideas, learners often reported being concerned with rehearsing their stories and monitoring overall content and form.

Excerpts

"I'm reading, if I read it again I will imagine the story in my head so I can remember when I tell" (p24)

"I checked the plural of the words and corrected a mistake" (p22)

" I was reading and I decided something different for the end" (p 20)

As these excerpts show, learners also attempt to rehearse their stories during planning time. Moreover, they monitor for improving overall content as participant 17 verbalizes *"I was reading and I decided something different for the end";* and also monitor for improving grammar *"I checked the plural of the words and corrected a mistake"*.

I think it is plausible to conclude that, in general, planning assists performance by allowing learners to engage in organizational, retrieval, rehearsal, and monitoring operations. More specifically, the strong emphasis on lexical searches, organizational of ideas, and monitoring implies that learners seem to anticipate problems on the stages of conceptualization, formulation, and monitoring.

Finally, I would like to address the discussion of focus on meaning and form during planning wisely put forward by Ortega (2005). In planning, Ortega (2005) argues, "learners engage in solving form-in-meaning problems" (p. 106). In this sense, she advocates the need to challenge the dichotomization of form and meaning. Ortega (2005) distinguishes two positions towards the dichotomization of form and meaning. According to her, Skehan and Foster (2001) and VanPatten (2002) emphasize the dichotomization between form and meaning by drawing on *limited* capacity theories of attention. On the other hand, she states that Dekeyser et al. (2002), drawing on *unlimited* capacity theories of attention, claim that the dissociation between meaning and form is impossible, and attention to both is clearly possible.

Throughout the protocols of the present study, a focus on form on the part of learners was clearly stated. However, these instances of focus on form did not take place in a vacuum; they emerged as learners attempted to convey meaning. The following excerpts illustrate these instances of focus on form in the attempt to convey meaning.

Excerpts

> "I'm still thinking in the things that the man thought, I was trying to remember the pictures…I was thinking in the correct word to use…I think in using *make* but I think *do* is better (p7).

> "I'm thinking about the relationship between Ciao the guy and Ana the girl …I'm thinking of using the word *jealous* in the story and that I've been Ciao once." (p16)

From these excerpts, it can be seen that in the attempt to convey the general meaning of their stories, learners focused on form. As in the first excerpt which shows that the learner is working on content "*I'm still thinking in the things that the man thought, I was trying to remember the pictures"*. The pictures of the narrative being planned by this learner display a series of thoughts of a man in relation to things he would like to do to his wife. In the attempt to express the man's thoughts, the learner focuses on what verb is suitable "*I think in using 'make' but I think 'do' is better*…"

In the second excerpt, the learner also seems to focus on content *"I'm thinking of the relationship of the guy Ciao and Ana the girl*". Then, he focuses on a specific lexical item which seems necessary to express ideas about the relationship of the couple *"I'm thinking of the word 'jealous' in the story"*.

Taking these instances of focus on form in the attempt to convey meaning into account, I believe it is plausible to conclude that learners shift attention from meaning to form and vice-versa. However, I believe the possibility of focusing on meaning and form fits into *limited* capacity theories of attention.

If one takes Engle's et al. (1999) perspective on working memory, individuals differ in knowledge and ability to manipulate knowledge as well as in the capacity for sustaining, maintaining, and *shifting* attention. Therefore, attention to meaning and form may be possible not because attentional resources are unlimited, but because learners shift attention from meaning to form and vice-versa throughout planning time.

During pre-task planning, learners activate information from long-term memory necessary to convey meaning, which may be information containing knowledge about the world, about the L2 (semantic memory), and also information acquired through personal events (episodic memory). Learners need to activate information necessary to convey meaning and maintain this information activated and easily accessible, while

processing formal aspects of the language (e.g. lexical and grammatical problems), which will be subsequently integrated into the information necessary to convey meaning.

I believe simultaneous attention to form and meaning during planning is clearly possible. The *extent* to which meaning and form are activated, that is to say, the ability to control and shift attention from meaning to form and vice-versa is what seems to differ. In this way, by activating meaning information from long-term memory, maintaining it activated and accessible while processing formal aspects of the language, learners seem to address their 'form-in-meaning problems' during planning (using Ortega's terminology).

Having discussed the learners' processes when planning and contributed to the discussion about focus of form and meaning during planning, now I turn to the results on the differences in the processes lower and higher spans engage in when planning.

5.6 Differences in the processes lower and higher spans engage in when planning

The sixth research question asked whether higher and lower span individuals would differ in terms of the processes they engage during planning. In brief, results showed that:

1. Higher spans significantly outperformed lower spans in the number of met cognitive strategies employed.
2. Higher spans significantly outperformed lower spans in the total number of strategies used.
3. Higher spans did not significantly outperform lower spans in the number of cognitive strategies, which suggests that the number of metacognitive

strategies is what seems to account more for the significant differences in the total number of strategies.

In addition to the statistically significant differences that were verified by means of a *t*-test, it is also possible to see some other quantitative differences, but they are better to be seen only as possible tendencies. These differences are related to the use of elaboration and writing/outlining/summarizing strategies, which were also frequently reported. Elaboration was also more frequently reported by higher spans (50%) than lower spans (25%); and also writing/outlining/summarizing was more frequently reported by higher spans (100%) when compared to lower spans (75%).

Ortega (2005) found evidence that individual differences in terms of language expertise reflect in the processes learners engage in during pre-task planning. Her results suggest that advanced learners engage more fully in self-monitoring and are able to allocate efforts towards retrieval and rehearsal operations in a more balanced fashion than low-intermediate learners.

Results of the present study suggest that in a homogeneous group in terms of language expertise, individual differences in working memory capacity may reflect differences in the ways learners approach planning. Results showed that higher spans used significantly more metacognitive strategies. They also tended to use planning time to elaborate and write/outline/summarize more frequently than lower spans.

The greater use of metacognitive strategies by higher spans encompass differences in the use of strategies such as organizational planning, problem identification, monitoring, and rehearsal by higher spans since these were the metacognitive strategies reported throughout the protocols. As can be seen in Table 25, the greater differences between lower and higher spans were in terms of rehearsal,

organizational planning, and monitoring. Rehearsal was reported by 25% of the lower spans and by 50% of the higher spans; organizational planning was reported by 50% of the lower span and by 75% of the higher spans respectively; and monitoring was reported by 50% of the lower spans and by 87.5% of the higher spans respectively.

It seems fair to say that the general tendency was that higher spans were more able to carry out some sort of organization before engaging in the task itself by organizing pictures in a sequence, deciding on general content, and setting communicative goals. Then, they searched for lexical items, engaged in solving lexical grammatical problems, and, finally, still used some of their planning time to monitor, elaborate, and embellish their stories as well as to rehearse their plan for the upcoming performance. Lower spans, on the other hand, did not seem to engage in organizational planning, monitoring, and rehearsal as much as higher spans. Most of them seemed to embark straight in searching for lexical items and solving grammatical problems without a general organization a priori. Moreover, they did not engage in monitoring, rehearsing, and elaboration as much as higher spans after lexical items were searched, grammatical problems were solved, and a general sketch of the story was accomplished.

As previously explained, there were only two instances of social/affective strategies throughout the protocols due to the nature of the monologic task used in the study. One of these strategies was used by a lower span learner and one by a higher span learner. Interestingly, qualitative differences also emerged in this minimal use of social/affective strategies, as can be seen in the following excerpts.

Excerpts

"I was thinking that if I start to worry too much about grammar I will be too nervous, I can't, I try not worry too much try not worry too much" (p24)

"Please, what do I do if I don't remember a word?"(p3)

As these excerpts show, participant 24, who was classified a higher span, was able to detect by herself one element of her behavior which could be detrimental for her performance, and she tried to suppress this element of nervousness; on the contrary, participant 3, classified as a lower span, was not able to overcome a lexical problem on his own and asked for help from the part of the present researcher. It is important to highlight that most learners were able to substitute words and overcome lexical problems; it was a frequently used strategy throughout the protocols. However, overcoming a lexical problem by this learner seems to have been a burden which he could not cope with by himself. Obviously, there was not enough use of social/affective strategies in order to make any strong claims about differences between lower and higher spans. Nevertheless, even this small instance of strategy use points in the same direction, so as to lend support to the finding that there are differences in the ways lower and higher spans approach planning.

So far, results have revealed that higher spans significantly outperformed lower spans in the number of metacognitive strategies used. Within these metacognitive strategies, the differences between higher and lower spans seem to lie mainly on rehearsal, organizational planning, and monitoring. Levelt (1989) claims that *conceptualizing a message* and *monitoring* are the two components of L1 speech production that draw more heavily on learners' attentional resources.

As previously explained, in conceptualizing and message construction, speakers do not have a fixed slot of intentions to convey, and communicative intentions can vary in countless ways. As for monitoring, it demands attentional control in the sense that the speaker attends to his own internal and overt speech (Levelt, 1989). To reiterate, Levelt's (1989) model accounts for L1 speech production, and in the case of L2 speech,

conceptualization and monitoring may be even more severely under attentional control. Moreover, Fortkamp (2003) proposes that L2 formulation, specifically at L2 grammatical encoding, is a controlled processing activity. Since higher spans more frequently deal with communicative goal setting and monitoring during planning, it could be argued that they tend to focus more frequently on the aspects of L2 speech which are more demanding on attentional resources when compared to lower spans.

As previously discussed, the significant differences between lower and higher spans in terms of L2 speech performance were in fluency and complexity, with the greater differences being in terms of fluency, which was tentatively explained by a greater ability to control attention among the various processes involved in speaking *during* task performance (Fortkamp, 2003), a greater ability to control attention among the processes involved in planning as a problem solving activity (Hambrick & Engle, 2003), a greater ability to retrieve what was planned into performance (Rosen & Engle, 1997; Unsworth Engle, 2007b), and a greater ability to implement new ideas online (Feldman-Barrett, et al., 2004) from the part of higher spans.

Since higher spans more frequently tended to deal with conceptualization and monitoring during planning, the cognitive pressure of these two aspects may have been reduced during online performance and, thus, more attentional resources were freed up to be focused on formulation, retrieval of planned information, and implementation of new ideas online.

In brief, results suggest that learners tend to use planning time to anticipate problems in conceptualization of the message, formulation, and monitoring. Taking individual differences in working memory into account, higher spans seem to focus on conceptualizing and monitoring more frequently than lower spans.

The finding that higher and lower spans differed in terms of the processes they engage in during planning is an interesting one in itself which merits reasonable speculation. The last question to be pursued in this discussion of results is: Why do higher spans tend to employ strategies more effectively during pre-task planning when compared to lower spans?

I believe that the greater ability to control attention among the various requirements of planning as a problem solving activity (Hambrick & Engle, 2003) seems to allow higher spans to sustain, maintain, and shift attention among the different strategies employed during planning – organizing ideas, searching lexical items, monitoring, rehearsing, and elaborating – in a more balanced fashion when compared to lower spans. In other words, higher spans have more ability to control and allocate attention towards different strategies during planning.

In this chapter, I have discussed results of the present study. In the next chapter, I will present a summary of the main findings of the study, draw some pedagogical implications, point out limitations of the study, and provide suggestions for future research.

CHAPTER VI

CONCLUSION

6.1 Final Remarks

This study aimed at investigating the relationship among individual differences in pre-task planning, working memory capacity, and L2 speech performance. It was assumed that individual differences in working memory capacity would emerge in no planning and planning conditions. It was hypothesized that: 1) participants' working memory capacity scores would significantly correlate with fluency measures of L2 speech performance under no planning condition, 2) participants' working memory capacity scores would significantly correlate with accuracy measures of L2 speech performance under no planning condition, 3) participants' working memory capacity scores would significantly correlate with complexity measures of L2 speech performance under no planning condition, 4) under pre-task planning condition, there would be greater fluency when compared to the control group, 5) under pre-task planning condition, there would be greater accuracy when compared to the control group, 6) under pre-task planning condition, there would be greater complexity when compared to the control group, 7) participants' working memory capacity scores would significantly correlate with fluency measures of L2 speech performance under pre-task planning condition, 8) participants' working memory capacity scores would significantly correlate with accuracy measures of L2 speech performance under pre-task planning condition, 9) participants' working memory capacity scores would significantly correlate with complexity measures of L2 speech performance under pre-task planning condition, 10) within the experimental group, under pre-task planning

condition, higher working memory spans would significantly outperform lower working memory spans as regards fluency of L2 speech production, 11) within the experimental group, under pre-task planning condition, higher working memory spans would significantly outperform lower working memory spans as regards accuracy of L2 speech production, 12) within the experimental group, under pre-task planning condition, higher working memory spans would significantly outperform lower working memory spans as regards complexity of L2 speech production, 13) when planning an oral task, learners would engage in the following processes: (a) organization of ideas, (b) lexical-grammatical search, (b) task rehearsal, and (d) monitoring, and 14) higher and lower span individuals would differ in terms of the mental processes they engage in when they plan.

To test the 14 hypotheses, 50 intermediate learners were submitted to two phases of data collection. For the control group, the first phase consisted of a speech generation task under no planning condition, and the second one consisted of a memory test (The Speaking Span Test), a speech generation task also under a no planning condition, and a semi-guided interview. For the experimental group, data collection procedures were different. The first phase consisted of a speech generation task under a no planning condition, the second consisting of a memory test (The Speaking Span Test), a speech generation task under a planning condition, a retrospective online protocol, and a semi-guided interview. Participants' speaking samples were analyzed in terms of fluency (speech rate pruned and unpruned, number of pauses per c-unit, and total percentage of pausing time), accuracy (number of errors per a hundred words, percentage of error free clauses), and complexity (number of clauses per c-unit).

In general terms, results show that (a) under no planning conditions, working memory capacity is related to accuracy of L2 speech performance for the control group,

and fluency of L2 speech performance for the experimental group; (b) under planning condition, working memory is related to fluency and complexity of L2 speech performance, with higher spans significantly outperforming lower spans in fluency and complexity, but not in accuracy; (c) under planning condition, the greatest differences in the performance of lower and higher spans are in terms of fluency; (d) learners engage mainly in organizational, retrieval, rehearsal, monitoring, and elaboration during planning; and (e) higher spans employ significantly more metacognitive strategies when compared to lower spans.

In order to account for the relationship between working memory and L2 speech performance under no planning condition, it has been argued based on the attention-view of working memory capacity (Engle & Oransky, 1999; Engle et al., 1999; Hambrick & Engle, 2003; Heitz et al., 2005; Kane et al., 2007) that participants with higher working memory capacity tend to have a superior ability to control attention among the various components that L2 speech production encompasses. In the case of the relationship between working memory capacity and fluency under no planning condition for the experimental group, it was suggested that higher spans tend to be more able to control attention among all components of L2 speech, but particularly conceptualizing, message construction, and formulation. As for the relationship between working memory capacity and accuracy under no planning condition for the control group, it was suggested that higher spans are more able to control attention among all the processes of L2 speech, particularly formulation and monitoring.

It was also argued that learners' 'choices' on what aspects of performance to prioritize are not deliberate conscious 'choices'. Rather, such 'choices' take place in a reflexlike fashion (Feldman-Barret et al., 2004), being triggered automatically by the environment vis-à-vis task performance conditions. Based on Batstone (2005), it was

also argued that these reflexlike 'choices' triggered by the environment may reflect learners' backgrounds and experiences in the course of L2 learning.

As for the finding that planning led to significant differences in accuracy and complexity, it was attributed to the extensive use of retrieval and rehearsal operations during planning (Foster & Skehan, 2001). The fact that, in this study, planning did not lead to gains in fluency is at odds with previous studies which reported more consistent effects of planning on fluency and complexity, but not on accuracy (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Crookes, 1989; Foster & Skehan, 1996; Ortega, 1999; Yuan & Ellis, 2003).

In order to account for the finding that planning did not lead to gains in fluency, it was suggested, based on Skehan's (1998) proposal for trade-offs among the goals of L2 speech, that gains in accuracy and complexity took place at the expense of gains in fluency. Moreover, it has been suggested that planning led to *some* increase in fluency predominantly in the performance of higher spans and, thus, the overall impact on fluency in the performance of the experimental group *as a whole* was reduced. What these results suggest is that trade-offs seem to be more acute for lower spans.

As for the finding that higher spans significantly outperformed lower spans in fluency and complexity, but not in accuracy, it has been argued that higher spans were more able to retrieve what was planned into performance (Rosen & Engle, 1997; Unsworth & Engle, 2007); more able to control attention among the processes of L2 speech during task performance (Fortkamp, 2003), and also more able to implement new ideas online (Feldman-Barrett et al., 2004), which reflected significant differences in fluency. As for the differences in complexity, it was argued that higher spans were more able to draw upon rule-based processing (Feldman-Barrett et al.) and, thus, made more use of cutting edge language. As for accuracy, it has been argued that lower spans

may view efficient performance as error free performance and, thus, prioritized efficiency above any other goal (Feldman-Barrett, et al.), which reflected *no* significant differences between lower and higher spans in terms of accuracy.

Concerning the processes learners engage in during planning, results show that learners in the experimental group *as a whole* engage in the processes of organizational of ideas, retrieval, rehearsal, monitoring, and elaboration, corroborating most of Ortega's (2005) results. As for the differences between lower and higher spans in terms of the processes they engage in, results reveal that higher spans employ significantly more metacognitive strategies when compared to lower spans.

In order to account for the finding that higher spans tend to benefit more from planning (in terms of fluency and complexity), it was suggested that the processes tapped by the Speaking Span Test also seem to be tapped in planning as a problem solving activity, which requires learners maintain task-relevant information activated and accessible, and to inhibit irrelevant information (Hambrick & Engle, 2003).

Within the scope of the megtacognitive strategies, results suggest that higher spans tend to focus on rehearsal, organizational planning, and monitoring more frequently than lower spans. According to Levelt (1989), communicative goal setting, message construction, and monitoring are the stages of speech that draw more heavily on attentional resources. Based on that, it was suggested that higher spans are more able to relief the pressure on these stages of speech and, thus, have more resources available to focus on the retrieval of what was planned into online performance, formulation, and implementation of new ideas online.

In order to account for the finding that higher spans tend to use strategies in a more balanced fashion, making use of more metacognitive strategies, it was suggested that higher spans are more able to cope with the requirements of planning as a problem

solving activity, which demands controlled attention (Hambrick & Engle, 2003). This more effective attentional control towards the requirements of planning allows learners to employ strategies in a more balanced fashion during planning. In other words, based on the attention-view of working memory (Engle & Oransky, 1999; Engle et al., 1999; Hambrick & Engle, 2003; Heitz et al., 2005; Kane et al., 2007; Unsworth & Engle, 2007b), higher spans tend to be more able to activate and manipulate knowledge, as well as to sustain, maintain, and shift attention (e.g., from meaning to form and vice-versa) during pre-task planning.

The findings of the present study are relevant since they go beyond the general speculation that the effects of planning are not achieved simultaneously to the same extent for fluency, accuracy, and complexity due to limitation in attentional resources (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003). This study represents a step forward by providing evidence that individual differences in working memory capacity mediate L2 speech performance under no planning and planning conditions. Moreover, the findings of the present study suggest that lower spans tend to be more susceptible to attentional trade-off effects among fluency, accuracy, and complexity. Obviously, findings of the present study are to be seen as suggestive rather than conclusive due to its several limitations.

6.2 Limitations and suggestions for future research

The present study represents a tentative and preliminary attempt to examine the relationship among individual differences in working memory capacity, pre-task planning, and L2 speech performance. Results are to be seen as modest and suggestive rather than conclusive due to the several limitations of the study.

The present study is limited in its sample size; it was conducted with only fifty participants. Due to this reduced sample size, the extreme-group design was conducted based on tertiles, not quartiles. Moreover, the differences between lower and higher spans were based on a more reduced sample of only sixteen participants, 8 classified as lower and eight classified as higher spans. Although, in the L2 field, samples of fifty participants are considered as appropriate for experimental studies, in the area of working memory research, most studies are conducted with far more participants. Therefore, future studies need to consider expanding sample size.

The present study is also limited in the sense that there was only one test to assess working memory capacity. Conway et al. (2005) suggest that at least two measures of working memory should be used whenever possible. However, due to participants' time constraints, it was only possible to use one test in this study. Future studies need to consider including more measures of working memory capacity in order to reach firmer grounds on the relationship between working memory and L2 performance.

One more limitation concerns the fact that only monologic 'there-and-then' narratives were used. Alternatively, future research could make use of 'here-and-now' narratives, or even interactive tasks in order to expand the scope of individual differences within the effects of planning on performance.

The study is also limited in relation to the level of proficiency. Only intermediate students took part. It would be interesting to compare the role of working memory in the performance of learners from different proficiency levels. Future research could address the relationship between pre-task planning, working memory, and L2 speech performance of beginners and/or advanced learners.

One more limitation of the study is related to the measures used to assess L2 speech performance. As for fluency, only speed fluency (speech rate unpruned and pruned), and breakdown fluency (number of pauses per c-unit, total pausing time) were used. Complementary measures of repair fluency such as repetitions, hesitations and self-repairs should be taking into account in order to give a more comprehensive view of fluent L2 speech performance. Moreover, Foster and Skehan (2005) claim that pauses are always treated in the same way, but it is important to distinguish between pauses at the end of clauses, which are more natural, from pauses which take place at the middle of clauses.

As for accuracy, two measures were employed: number of errors per a hundred words and percentage of error free clauses. Some researchers raise the possibility that percentage of error free clauses may mask general achievements in accuracy (Bygate, 2001b; Foster & Skehan, 2005). Therefore, Foster and Skehan (2005) suggest that when dealing with such measures, the length of the clauses also needs to be taken into account.

Bearing the aforementioned limitations in mind, the conclusions of present study concerning the relationship among working memory, pre-task planning, and L2 performance are restricted to the performance of young adult intermediate learners of English when working memory is assessed by means of the Speaking Span Test conducted in the L2. The generalization of these findings to other populations, languages and other working memory tests remains to be empirically shown.

One issue which merits to be highlighted is the relevance of investigating planning through a *process*-product oriented approached. According to Ortega (1999, 2005), most of the research on planning is product oriented focusing on the impact of planning on performance. The present study took a *process*-product oriented approach

and went beyond the scrutiny of learners' processes in the sense that it examined individual differences in terms of the processes triggered by planning. Nevertheless, most of the research on planning still remains focused on its product. More research is needed on the *process*-product oriented paradigm in order to understand learners' perceptions and motivations towards planning and how these perceptions and motivations may impact the act of planning itself.

The need for more *process*-product oriented research points to another limitation of the present study, which concerns the use of retrospective online protocols. Leow and Morgan-Short (2004) gathered evidence for the lack of reactivity effects in the use of online protocols. However, these researchers claim that the issue of reactivity still needs further scrutiny. Future studies on planning from a *process*-product oriented perspective could have two planning groups, one using retrospective online protocols and another group using retrospective interviews, for instance. Efforts in this direction would help us reach firmer grounds on the issue of reactivity, and would, consequently, shed some light on what type of protocols to use. If future efforts to scrutinize learners' processes in task-based planning research are to be made, it seems crucial to gain a better understanding of the instruments to assess these processes.

At this point, I would like to point out a limitation that applies not only to this study but to most studies on task-based planning. According to Batstone (2005), the research paradigm on planning has been *essentially* cognitive and little is known about the role of the social contexts in which planning takes place. Efforts towards examining planning in a more socially embedded perspective may be enlightening since "both the learners' capacity to plan and their ability to act on planning by pushing output are socially rooted" (Batstone, 2005, p. 278). Results of this study showed a difference in group orientation in prioritizing fluency or accuracy in the performance under no

planning conditions, which may be related to learners' backgrounds. However, this remains essentially speculative as there were no attempts to take a closer look at the learning contexts of the participants of the present study. Future research needs to address planning from a more socially grounded perspective. As a consequence of the essentially cognitive oriented approach on planning, most of the research is experimental. Planning, however, is a tool which can be easily implemented in L2 classrooms. In this sense, future research is needed in the attempt to scrutinize how planning takes place in the classroom in more interactive contexts.

It is also important to point out that research on planning so far has focused solely upon the impact of planning on L2 *performance.* I believe the field is ripe to take a further step in the attempt to examine whether planning may have any effects on L2 *acquisition*. During pre-task planning learners notice gaps in their interlanguage as well as undergo metacognitive reflection. According to Swain (1995), noticing of gaps and metalinguistic reflection play a role in acquisition. Therefore, there seems to be enough room to hypothesize that planning may assist acquisition. Future research is needed in this direction. An interesting avenue of inquiry would be to investigate the relationship between individual differences in working memory capacity, pre-task planning, and L2 *acquisition.*

Another interesting avenue of investigation would be to address the relationship between retrieval and working memory capacity in pre-task planning. In the present study, I raised the possibility that higher spans were more effective in retrieving what was planned into online performance. However, this claim was essentially speculative since there were no attempts to scrutinize retrieval of planned ideas into performance. Future research could address this issue by examining learners' planning notes and protocols in relation to their actual performance.

Based on the results of the present study, I am inclined to believe that the *attention-view* of working memory seems promising for SLA research addressing the relationship between L2 learning/performance and working memory, for at least three main reasons. First, it is a consolidated view in the area of working memory research, which has generated extensive research (e.g., Cantor & Engle, 1993; Engle, 1989; Engle, Cantor & Carullo, 1992; Engle & Oransky, 1999; Hambrick & Engle, 2003; Heitz et al., 2004; Kane et al, 2007; Rosen & Engle, 1997; Unsworth & Engle, 2007, just to mention a few). Second, it lays emphasis on the construct of attention, which is a key construct in the field of SLA (Schmidt, 1990, 1993; Tomlin & Villa, 1994). Third, particularly in the field of task based research, the attention-view of working memory is compatible with Skehan's (1996, 1998) proposals of trade-offs among fluency, accuracy, and complexity in the context of learners' *limited attentional* resources.

On having advocated that the attention-view of working memory is promising for SLA, one question deserves to be asked: Would it be true the other way around? In other words, how about the research in SLA, can it be helpful for working memory research? Research on language performance has already been useful for working memory research. Seminal studies in L1 reading comprehension (Daneman & Carpenter, 1980, 1983), and L1 production (Denamen, 1991 ; Daneman & Green, 1986) have proved to be useful windows through which to look at individual differences in working memory capacity and have contributed to the growth of working memory research. Studies on L2 performance are also mounting and shown to be fruitful avenues for research on individual differences in working memory capacity (Bergsleithner, 2007; Fontanini et al., 2005; Fortkamp, 1999, 2003; Mizera, 2006; Torres, 2003; Weissheimer, 2007;)

According to Kintsch, Healey, Hegarty, Pennington and Salthouse (1999), one criticism that can be raised towards working memory research is the use of artificial tasks, such as the Tower and Hanoi[22] tasks, for instance. Although, these authors acknowledge the importance of tasks like the Tower and Hanoi in experimental research, they state that tasks of this sort are believed to be rather artificial when compared to real world cognitive tasks (e.g., comprehension of a text accompanied by diagrams) and, thus, may not accurately reflect performance on complex cognitive tasks encountered in everyday cognition.

In addition to that, Hambrick and Engle (2003) state that research on problem solving is sometimes viewed as a narrow area of investigation since it is limited to tasks such as the Tower and Hanoi. However, they advocate that many cognitive tasks can be viewed as examples of problem solving as long as they involve 'purposeful, goal-directed behavior' (using the terminology of Hambrick and Engle's). In this respect, research on pre-task planning as a problem solving activity may be helpful in the attempt to broaden the scope of research on problem solving.

Interestingly, Feldman-Barrett et al. (2004) provide a list of processing outcomes associated with working memory capacity, and, in this list, cognitive activities such as reading comprehension, listening comprehension, spelling, vocabulary learning, and taking lecture notes are grouped under the label of 'real-world cognitive tasks'. All the tasks in the list provided by Feldman-Barrett et al. (2004) involve aspects of language. In this sense, the SLA field may be promising for working memory research by providing complex tasks which are encountered in everyday cognitive settings. Language per se is already a system, which is inherent to all human beings, at least the

[22] The Tower and Hanoi task requires individuals to move a set of colored balls across different–sized pegs to match a target configuration.

ones without any major impairment. L2 learning/ performance is also at play worldwide, present in over a million of individuals' daily cognition.

In brief, as far research on individual differences in working memory is concerned, I believe L2 learning and performance contexts are promising in the sense that they provide complex cognitive tasks that are more common to everyday cognitive settings such as L2 reading tasks, L2 speech tasks, L2 writing tasks, L2 planning and the like. If one considers the tasks used in the present study – the task of producing an oral narrative under an experimental condition and the task of planning L2 speech – these tasks may not be so common to everyday cognition when compared to reading a newspaper, writing an email message, and so on. However, these tasks are not artificial either in the sense that they are frequent in *L2 learning* contexts, which are common worldwide. Hambrick and Engle (2003) state that "there is still much to be learned about the role of working memory in real-world cognitive functioning" (p. 177). I believe looking at L2 learning and performance may be fruitful in this direction.

6.3 Pedagogical implications

Although planning has been approached as a research construct in the field of task-based research, it is a relatively simple procedure in terms of pedagogy (Batstone, 2005), which can be implemented in the classroom as a tool for fostering the speaking skill. According to Bygate (2001), although the main feature of speaking is spontaneity, it needs to be treated in a systematic fashion in the L2 classroom.

Within the treatment of the speaking skill, planning could be used as a pedagogical tool since it not only promotes benefits on subsequent performance but also engage learners in processes which may lead to acquisition such as noticing of gaps and

metalinguistic reflection (Swain, 1995). Planning can be used before task performance and assessment. The issue seems to be *how* planning can be implemented in the classroom. Planning is a means of drawing learners' attention to form, and I believe it can be implemented in the L2 classroom by means of incidental and planned focus on form instruction[23]. In incidental focus on form instruction, teachers can simply give an oral task and allow planning time so that learners' general questions and doubts on form would take place as they attempt to convey meaning. In planned focus on form instruction, teachers can ask learners to plan an oral task which requires specific linguistic forms (e.g., past tense). In this case, learners' questions and doubts in relation to the specific forms required by the task would be likely to take place.

In the implementation of planning as a pedagogical tool for fostering the speaking skill, it is important to highlight that learners may differ in their ability to plan. Results of this study suggest that learners may differ in the extent to which they benefit from planning. Some learners may need assistance in order to become more effective planners. In this sense, a variety of approaches towards planning need to be considered such as individual planning, teacher-guided planning, peer planning, and group planning.

Although, being a strategy that leads to benefits and that can be easily implemented, "planning is just *one* [italics added] of a number of strategies for learning within the philosophy that learners should take a greater responsibility for their own learning" (Batstone, 2005, p.284). Bearing that in mind, one question in need to be answered is: Should teachers *always* allow planning time?

---

[23] Ellis (2002) distinguishes among three types of form-focused instruction: focus on form*s*, incidental focus on form, and planned focus on form.

I take the perspective that learners need to be given not only the opportunity to experience planning but also to perform under improvisation. Learners need to be provided opportunities for planning so that they may engage in the processes of organization of ideas, monitoring, rehearsal, and elaboration, and, thus, will be more likely to push output, and use cutting edge language. On the other hand, they need to be prepared to deal with the pressure of performing under more difficult situations in which it is not possible to plan.

Oxford (1993), Felder and Henriques (1995), Wintergerst, DeCapua and Verna (2003), and Guará-Tavares (2007) advocate the idea that learners should be exposed to balanced teaching styles, that is to say, teachers should include learning tasks which match learners' learning preferences as well as learning tasks which mismatch their styles and, thus, challenge them to become more flexible learners.

Following this train of thought, I believe planning should not be *imposed* to learners *all* the time. Planning opportunities should be systematically provided to learners along with no planning opportunities in the attempt to help them become more strategic as learners and enable them to choose which strategy best fits their educational background, styles and/or learning purposes, and also according to the learning/performance demands they face.

The pedagogical implications provided here are to be seen as suggestive rather than prescriptive and any attempts to implement planning in the classroom may need to undergo adaptation in order to fit the teaching/learning contexts in which it is taking place.

In conclusion, the objective of this doctoral study was to address individual differences in working memory capacity within the effects of planning on L2 speech performance. Research on planning is relevant for current theorizing about L2

acquisition in terms of information processing theory and for its usefulness for language pedagogy (Ellis, 2005). I hope the present study sheds some light on how individual differences in working memory capacity may reflect differences in how learners employ a strategy which has shown to clearly impact on the performance of L2 *speaking*, which is considered a complex cognitive skill (Levelt, 1989). In addition, I hope this study provides a step forward by offering some evidence for the role of working memory within task-based planning.

REFERENCES

Armbruster, B. B. (1983). The role of metacognition in reading to learn: A developmental perspective. *Reading Education Report, 40*, Urbana, IL: Center for the Study of Reading [ED228617]

Ashcraft, M. (1994). *Human memory and cognition*. New York: Harper Collins.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225* (2), 82-90.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Hove, UK: Lawrence Erlbaum.

Baddeley, A. D. (1992). Working memory. *Science, 225*, 556-559.

Baddeley, A. D. (1999). *Essentials of human memory.* East Sussex: Psychology Press.

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

Baddley, A. D., Gathercole, S., & Papagano, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol.8, pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 28-59). Cambridge: Cambridge University Press.

Batstone, R. (2005). Planning as discourse actitivy: A sociocognitive view. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 277-295). Amsterdam and Philadelphia: John Benjamins.

Berardi-Coletta, B., Dominowski, R. L., Buyer, L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology, Learning, Memory and Cognition, 21* (1), 205-223.

Bergsleithner, J. M. (2007). *The relationship among individual differences in working memory capacity, noticing, and L2 speech production*. Unpublished Ph.D. dissertation, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller, & P. D. Eimar (Eds.), *Speech, language and communication* (pp. 181-216). San Diego: Academic Press.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). London: Academic press.

Bygate, M. (2001a). Speaking. In *The Cambridge Guide to Teaching English to Speakers of Other Languages*, (pp. 14-20). Ed. Carter, R. & Nunan, D. Chapter 2. Cambridge: Cambridge University Press.

Bygate, M. (2001b). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning and testing* (pp. 37-53). London: Longman.

Cantor, J., & Engle, R. W. (1993). Working memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19* (5), 1101-1114.

Clark, H., & Clark, G. (1977). Psychology and language: An introduction to psycholinguistics. New York: Harcourt, Brace, Jovanocich.

Cohen, J. (1988). *Statistical power analysis for the behavourial sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112* (1), 155-159.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354-373.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4*, 577-590.

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

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

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*, 445-464.

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

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561-584.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology, 17*, 367-378.

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

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

Dekeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's 'Procesing instruction: An update'. *Language Learning, 52*, 805-823.

D'Ely, R. (2004). *An investigation of learners' metacognitive processes of planning and rehearsal as catalysts of interlanguage development*. Unpublished manuscript.

D'Ely, R. (2006). *A focus on learners' metacognitive processes: The impact of strategic planning, repetition, strategic planning plus repetition, and strategic planning for repetition on L2 oral performance*. Unpublished Ph.D. dissertation, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

D'Ely, R., & Weissheimer, J. (2004). *Scale of L2 speaking proficiency*. Unpublished manuscript.

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287-314). Michigan: The University of Michigan Press.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238). Amsterdam and Philadelphia: John Benjamins.

Ellis, N., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49* (A), 234-250.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*, 12-20.

Ellis, R. (1994). *The study of second language acquisitio*n. Oxford: Oxford University Press.

Ellis, R. (2002). Investigating form-focused instruction. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 1-46). Malden, MA: Blackwell.

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

Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.03-34). Amsterdam and Philadelphia: John Benjamins.

Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition.*. New York: Oxford University Press.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 972-992.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.) *Models of working memory: mechanisms of active maintenance and executive control.* New York: Cambridge University Press, pp.102-134.

Engle, R. W., Laughlin, J. E., Tuholski, S. W, & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Engle, R. W., & Oransky, N. (1999). Multi-store versus dynamic models of temporary storage in memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 515-555). Cambridge, MA: MIT Press.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Eysenck, M. W. (1992). *Anxiety: The cognitive perspective.* London: Lawrence Erlbaum.

Faerch, C., & Kasper, G. (1980). Processes and strategies in foreign language learning and communication. *Interlanguage Studies Bulletin, 4*, 47-118.

Faerch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp.37-65). Harlow: Longman.

Felder, R. M., & Henriques, E. R. (1995). Learning and teaching styles in foreign and second language education. *Foreign Language Annals, 28* (1), 21-31.

Feldman-Barrett, L., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*, 553-573.

Field, A. (2005). *Discovering statistics using SPSS*. London. Sage.

Fontanini, I., Weissheimer, J., Bergsleithner, J., Perucci, M., & D'Ely, R. (2005). Working memory capacity and L2 skill performance. *Revista Brasileira de Lingüística Aplicad*a*, 5* (2), 189-230.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition, 32*, 259-296.

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

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

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

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*, 215-247.

Foster, P., Tonkin, A., & Wigglesworth, G. (2001). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*, 354-375.

Freed, B. (2000). Is fluency, like beauty, in the eyes of (and ears) of the beholder? In H. Haggenbach (Ed.), *Perspectives on fluency*. Ann Arbor, MI: Michigan University Press.

Friedman, N. P., & Miyake, A. (2004). The Reading Span Test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*, 136-158.

Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Erlbaum.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language, 27*, 210-223.

Greene, J. O. (1984). Speech preparation processes and verbal fluency. *Human Comunication Research, 11* (1), 61-84.

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

Guará-Tavares, M. G. (2005). *Planning, working memory capacity and L2 speech performance*. Unpublished manuscript.

Guará-Tavares, M. G. (2006). *Planning, working memory capacity and L2 speech performance*. Unpublished manuscript.

Guará-Tavares, M. G. (2007). A matter of style: Looking at L2 teachers' teaching styles from the perspective of learning styles. *Linguagem e Ensino, 10* (1), 109-140.

Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 176-206). London: Cambridge Press.

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

Hatch, E., & Lazaraton, A. (1993). *The research manaual: Design and statistics for Applied Linguistics*. New York: Newbury House.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press

Iwashita, N., Mcnamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing framework approach to task design. *Language Learning, 51* (3), 410-436.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General,* (130), 169-183.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z, & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Touse (Eds.), *Variation in working memory* (pp. 21-48). New York: Oxford University Press.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin and Review, 9*, 637-671.

Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of a goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47-70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate levels of L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 143-164). Amsterdam and Philadelphia: John Benjamins.

Kintch, W., Healey, A., Hegarty, M., Pennington, B. F., & Salthouse, T. A. (1999). Models of working memory: Eight questions and some general issues. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442-481). Cambridge: Cambridge University Press.

Klein, K., & Boals, A. (2001). The relationship of life events stress and working memory capacity. *Applied Cognitive Psychology, 15*, 565-579.

Lee, J. Y. (1999). Test anxiety and working memory. *The Journal of Experimental Education, 67* (3), 218-223.

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

Leow, R. P. (2002). Models, attention, and awareness in SLA. *Studies in Second Language Acquisition, 24*, 113-119.

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26*, 35-57.

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

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

McLaughlin, B., & Heredia, R. (1996). Information-processing approaches to research on second language acquisition and use. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 213-225). San Diego, CA: Academic Press.

Mcnamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29* (1), 10-17.

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

Mendonça, D. (2003). *Working memory capacity and the retention of L2 vocabulary*. Unpublished master's thesis, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

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

Miyake, A., & Shah, P. (1999). Toward a unified theory of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, (pp. 442-481). Cambridge: Cambridge University Press.

Mizera, G. J. (2006). *Working memory and L2 oral fluency*. Unpublished Ph.D. dissertation, University of Pittsburgh, Pennsylvania, The United States of America.

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

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A Research Synthesis and quantitative meta-alanysis. *Language Learning, 50* (3), 417-528.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. New Yourk: Cambridge University Press.

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

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-110). Amsterdam and Philadelphia: John Benjamins.

Oxford, R. L. (1990). Language learning strategies and beyond: A look at strategies in the context of styles. In S. S. Magnan (Ed.), *Shifting the instructional focus to the learner* (pp. 35-55). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.

Oxford, R. L. (1993). Instructional implications of gender differences in language learning styles and strategies. *Applied Language Learning, 4*, 65-94.

Philp, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children's task-based interactions. *System, 34*, 547-565.

Poulisse, N. (1999). Language production in bilinguals. In A. M. B. de Groot & J. K. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 201-225). Mahwah, NJ: Lawrence Erlbaum.

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

Praxedes Filho, P. H. L. (2007). A corpora-based study of the interlanguage of EFL Brazilian learners' development from simpliication to complexification in the light of Systemic-functional Grammar. Unpublished Ph.D. dissertation, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Quirk, R., & Greenbaum, S. (1973). *A university grammar of English.* London: Longman.

Ritchie, W. C., & Bhatia, T. K. (Eds.) (1997). *Handbook of second language acquisition*. San Diego, CA: Academic Press.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45* (1), 99-140.

Robinson, P. (2001). 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* (pp. 287-318). Cambridge: Cambridge University Press.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology, 126* (3), 211-227.

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*, 41-51.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In H. W. Dechert & M. Haupach (Eds.), *Psycholinguistic models of production* (pp. 45-65). Norwood, NJ: Ablex.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111-142). Amsterdam and Philadelphia: John Benjamins.

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

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

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17* (1), 38-62.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1995). Some type and task processing conditions as influences on foreign language performance. *English Language Teaching, 3*, 139-188.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions on foreign language performance. *Language Teaching Research, 1*, 185-211.

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

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

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193-218). Amsterdam and Philadelphia: John Benjamins.

Swain, M. (1985). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in Applied Linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

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

Tomitch, L. M. B. (1996). Individual differences in the text organization perception and working memory capacity. *Revista da ANPOLL, 2*, 71-93.

Tomitch, L. M. B. (2000a). Individual differences in working memory capacity and the recall of predicted elements in the text. *Lenguas Modernas, 26-27*, 31-51.

Tomitch, L. M. B. (2003). *Reading: Text organization perception and working memory capacity*. Florianópolis: PGI/UFSC.

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

Torres, A. C. G. (2003). *Working memory capacity and readers' performance in main idea construction in L1 and L2.* Unpublished Ph.D. dissertation, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 5*, 193-197.

Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-54.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*, 67-81.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132.

Unsworth, N., & Engle, R. W. (2007b). Individual differences in working memory capacity and retrieval: A Cue-dependent search approach. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger* (pp. 241-258). New York: Psychology Press.

Unsworth, N., Heitz, R. P., & Engle, R. W. (2005). Working memory capacity in hot and cold cognition. In R. W. Engle, G. Sedek, U. Hecker, & D. N. McIntosh (Eds.), *Cognitive limitations in aging and psychopathology* (pp. 19-43). New York: Cambridge University Press.van den Broek, P., Young, M., Tzeng, Y., &

Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B. (1990). Attending to content and form in the input: An experiment on consciousness. *Studies in Second Language Acquisition, 12*, 287-301.

VanPatten, B. (1996). *Input processing and grammar instruction.* Norwood, NJ: Ablex.

Weissheimer, J. (2006). *Working memory capacity and the development of L2 speech production*. Unpublished manuscript.

Weissheimer, J. (2007). *Working memory capacity and the development of L2 speech production*. Unpublished Ph.D. dissertation, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Weissheimer, J., & Fortkamp, M. B. M. (2004). *The effects of practice and strategy use on L2 working memory capacity and fluency.* Unpublished manuscript.

Wendel, J. (1997). Planning and second language narrative production. Unpublished Ph.D. dissertation, Temple University, Japan.

Wenden, A. (1991). *Learner strategies for learner autonomy.* London: Prentice Hall.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14*, 133-145.

Wigglesworth, G. (2001). Influences on performance in task-based oral assessments. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning and testing* (pp. 83-101). City: Longman.

Wintergest, A. C., DeCapua, A., & Verna, M. A. (2003). Conceptualizing learning style modalities for ESL/EFL students. *System, 31*, 85-106.

Xhafaj, D. C. P. (2006). *Pause distribution and working memory capacity in L2 speech production*. Unpublished master's thesis, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Appied Linguistics, 24*, 1-27.