The role of task sequencing in L2 development as mediated by working memory capacity

Mayya Levkina

The role of task sequencing in L2 development as mediated by working memory capacity

TESI DOCTORAL
presentada per

Mayya Levkina

com a requeriment per a l'obtenció del títol de

Doctora en Filologia Anglesa

Director: **Dr. Roger Gilabert Guerrero**

Tutor: **Dra. Elsa Tragant Mestres de la Torre**

Departament de Filologia Anglesa i Alemanya

Universitat de Barcelona

Barcelona, desembre de 2013

"Les idées ne sont pas faites pour être pensées mais vécues".
André Malraux

To my daughter

ACKNOWLEDGMENTS

I feel so happy to be writing these lines of gratitude to all the people who accompanied me a long and very tough path. I consider myself a very lucky person to work hand in hand with exceptional professionals in linguistics all over the world. This dissertation is dedicated to each and every one of them.

First of all, my words of infinite gratitude are addressed to my supervisor, Dr. Roger Gilabert Guerrero. From the very beginning he has always been the source of my inspiration and motivation, first in the MA thesis, and then in my PhD dissertation. I would have never finished it without his readiness to help me at every moment, his guidance during these years and his support in difficult moments of my professional and personal life. Dr. Roger Gilabert has always been enthusiastic about my initiation into academic life as a young researcher encouraging me to participate in international conferences and to publish my early work. Finally, Roger is a wonderful, sunny person, with whom it has been a great pleasure to work at the University of Barcelona.

Another word of my deep gratitude is addressed to Dr. Carmen Muñoz who believed in me 5 years ago and made my whole academic life possible. She gave me the opportunity to work with a group of highly trained professionals and to acquire an invaluable experience. Apart from being an exceptional researcher, she has always been a very kind and comprehensive person with me, which made working with her a great pleasure.

I will always be very grateful to Dr. Peter Robinson who accepted my first research stay in Aoyama Gakuin University, Tokyo (Japan) and helped me ever since in countless ways: sending me lots of documents to get my Japanese visa, organizing my whole stay in Tokyo before I came and constantly assisting me with my research during my stay. He contributed enormously to the design of the final experiment of the present dissertation by giving new ideas for how operationalize task complexity, as well as to the literature review, and by sharing many books and articles with me which undoubtedly enriched this work. I am also very thankful to Daisuke Nakamura, for helping me a lot to get around during my first days in

Tokyo and also for sharing his knowledge and skills about cognitive tests with me. I will always remember the nice meeting with Dr. Tomohito Ishikawa and his wonderful family in Soka and the tour they gave to me around the Soka Women's College. This made my stay in Tokyo even more remarkable.

I am also absolutely indebted to Dr. Judit Kormos who organized my second research stay at Eötvos Lórand University, Budapest (Hungary) in 2012, when my entrance visa for the UK to spend a research stay at Lancaster University was rejected. During three months in Budapest, Dr. Judit Kormos devoted a lot of time to my work assisting me with many statistical issues in my dissertation and giving me support with some aspects of the methodological part. Another person who I would like to specially thank is Dr. Andrea Révész from the College of Education at London University who has always been extremely helpful, patient and comprehensive during the whole process of my visa application to Great Britain. Finally, I would like to express my words of gratitude to Dr. Agnés Albert for having organized two seminars on Task Complexity and Individual Differences with her students at Eötvos Lórand University, where I could share my research. I am also grateful to the dean of the Department of English Applied Linguistics, Dr. Edit Kontra, for supporting me with the bureaucratic part.

I am endlessly thankful to all members of the GRAL group (Dr. María Luz Celaya, Dr. Joan Carles Mora, Dr. Teresa Navés, Dr. Raquel Serrano, Dr. Natalia Fullana, Dr. Mar Suárez, Dr. Laura Sánchez, Dr.Maria Àngels Llanes) for helping me to develop my academic and personal skills. You've always been a source of inspiration and motivation. I am especially grateful to Dr. Elsa Tragant for her help with Factor Analysis, and to Dr. Imma Miralpeix for her support and advice on elaborating one of the tests of the present dissertation.

Another word of my infinite gratitude is devoted to Jessica Mackey who helped me on numerous occasions with data collections for my pilots and my main experiment. Without her help, this dissertation would have never happened.

I will never forget all the former research assistants of the GRAL group, Júlia Barón, Anna Marsol, Mireia Ortega, Cristina Aliaga, Helena Torres and Leo

McPortland who have been always ready to offer not only their professional encouragement, but also their absolutely invaluable friendship.

I will specially remember my present colleagues from the McDermot office, Veronica Rodríguez, our literature soul, Marisa Camuñas, our source of never-ending energy, Melissa Cokely, our best Boston girl ever, and Ferran Gesà, the unique in every sense guy of our office. I'm so thankful to all of you for all the hugs, kisses, cafes, lunches, smiles and laughter you shared with me this year! Without you, guys, I would never be able to finish the dissertation.

This dissertation would not be possible without the assistance of my friend Elena Safronova who helped me countless times with collecting data, and has always cheered me up in difficult moments of my life.

I am eternally appreciative to all the students from the University of Barcelona and the School of Modern Languages who consented to voluntarily participating in my multiple pilots and in the main experiment. Without their patience and their consistency, this dissertation would never have been completed.

My life in Barcelona would not be the same without Laura López, my life guide, who supported me a million times from the very initial stages of my academic life. I won't be able to count all the coffees we had together, but I will always remember our Irish pink car. My working days at the University would not be the same without my artistic friend, Mercedes, who has always encouraged me professionally and personally.

Finally, I dedicate this dissertation to my mother and my grandmother who throughout my whole life have helped me to become what I am right now. Without their constant support and motivation, I would not have arrived to the final stage of this hard journey.

Por último, dedico esta tesis a mi pequeña familia, quienes han sido mi fuerza propulsora que me motivó enormemente y me ayudó en el último tramo de la carrera para poder llegar hasta el final.

ABSTRACT

The present dissertation was inspired by the increasing interest in task sequencing in L2 syllabus design, on the one hand, and the lack of empirical studies, on the other hand, where task sequencing and its effects on L2 development would be theoretically well-grounded and empirically researched. Moreover, no research has been carried out to investigate the mediating role of working memory capacity (WMC) in L2 development along task sequencing. These considerations became fundamental in the design of the present research. The study has been created based on Robinson's Cognition Hypothesis (2001a, 2001b, 2003, 2005, 2007b), the SAARC model (2010) and its predictions regarding the effects of task sequencing on L2 acquisition (Robinson, 2005, 2007a, 2010). As for the construct of working memory capacity, Baddeley's model was adopted for the present work (Baddeley & Hitch, 1974; Baddeley, 2000, 2003).

Three pilots preceded the main study, where the operationalization of task complexity was empirically determined, along with the list of target items for L2 students employed in the main part. The experimental design consisted of a pretest, an immediate posttest and a delayed posttest which included a descriptive task and a two-part vocabulary test as two control tasks, as well as a treatment session consisting of input and a series of three tasks manipulated along the +/- spatial reasoning demands. Out of 91 participants who originally participated in the data collection 61 were finally selected to be included in the statistical analyses of the data. All the participants were learners of English as an L2. They represented 4 groups of students: two laboratory groups (English Philology, University of Barcelona) and two classroom groups (School of Modern Languages, University of Barcelona). Participants from each context (laboratory and classroom) underwent two different treatments: a) a task sequence from cognitively simple to cognitively complex along spatial reasoning dimension involved in the task and b) a randomized sequence. Working memory capacity was measured by means of the automated operation span test (Unsworth et al., 2005). The control tests were administered before treatment, immediately after treatment and two weeks after treatment.

The results showed that task sequencing (from simple to complex) played a beneficial role in input retention in the case of both the laboratory group and the classroom group. Furthermore, it was found that WMC moderated L2 development independently from task sequencing, but most interestingly WMC was very influential in the results obtained by the target laboratory group that performed treatment with the tasks sequenced from simple to complex.

The results are discussed in light of the Robinson's Cognition Hypothesis (2001a, 2001b, 2003, 2005, 2007b) and the SAARC model of task sequencing (2005, 2007a, 2010, as well as with reference to models of working memory capacity (Baddeley & Hitch, 1974; Baddeley, 2000, 2003) and to some previous studies in applied linguistics on L2 development and WMC and in cognitive psychology on the acquisition of spatial relations.

RESUM

Aquesta dissertació ha estat motivada per l'interès creixent en la seqüenciació de tasques en el disseny curricular de la L2, per una banda, i la manca d'estudis empírics, d'altra banda, on la seqüenciació de tasques i els seus efectes en el desenvolupament de la L2 seria basat en la teoria i investigat empíricament. D'altra banda, no hi ha hagut investigació dut a terme per veure quina és la funció mediadora de capacitat de memòria operativa en el desenvolupament de la L2 a través de la seqüenciació de tasques. Tot això va ser fonamental en el disseny d'aquesta tesi. L'estudi s'ha creat sobre la base de la Hipòtesi Cognitiva de Robinson (2001a, 2001b, 2003, 2005, 2007b), el model SAARC (2010) i les seves prediccions sobre els efectes de la seqüenciació de tasques en l'adquisició de la L2 (Robinson, 2005, 2007a, 2010). Pel que fa a la construcció de capacitat de memòria operativa, model de Baddeley va ser utilitzat per aquest treball (Baddeley & Hitch, 1974; Baddeley, 2000, 2003).

Els tres pilots van precedir l'estudi principal, on la operacionalització de complexitat de tasques va ser empíricament determinada, juntament amb la llista d'expressions d'espai utilitzats en l'experiment principal. El disseny experimental va constar de pretest, de posttest immediat i de posttest aplaçat que incloïa dues tasques de control: una tasca descriptiva i una tasca de vocabulari de dues parts. La sessió de tractament consistia en input i una sèrie de tres tasques manipulades a través de + /- raonaments espacials. Dels 91 participants, qui originalment va participar en la recollida de dades, 61 van ser finalment seleccionats per ser inclosos en l'anàlisi estadística de les dades. Tots els participants eren estudiants d'anglès com a L2. Representaven 4 grups d'alumnes: dos grups de laboratori (Universitat de Barcelona, Filologia Anglesa) i dos grups d'aula (Universitat de Barcelona, Escola d'Idiomes Moderns). Participants de cada context (aula i laboratori) van tenir dos tractaments diferents: a) una seqüència de tasques de cognitivament simple a cognitivament complexa a través de +/- raonaments espacials implicats en la tasca i b) una seqüència barrejada. La capacitat de memòria operativa es va mesurar mitjançant la versió automatitzada de *reading span* (Unsworth et al., 2005). Les proves de control es van administrar abans de tractament, immediatament després del tractament i dues setmanes després del tractament.

Els resultats van mostrar que la seqüenciació de tasques (de cognitivament simple a cognitivament complexa) va jugar un paper beneficiós en la retenció d'input en cas del grup de laboratori com a en cas del grup d'aula. A més a més, es va trobar que la memòria operativa modera desenvolupament de la L2 independent de la seqüenciació de tasques, però més curiosament la memòria operativa ha sigut molt present en els resultats obtinguts pel grup de laboratori, que va realitzar el tractament amb les tasques seqüenciades de simple a complexa.

Els resultats són posats en relació a la Hipòtesi Cognitiva de Robinson (2001a, 2001b, 2003, 2005, 2007b) i el model de SAARC de seqüenciació de tasques (2005, 2007a, 2010), així com pel que fa a models de capacitat de memòria operativa (Baddeley & Hitch, 1974; Baddeley, 2000, 2003;) i a alguns estudis previs en lingüística aplicada sobre el desenvolupament de la L2 i WMC i en psicologia cognitiva sobre l'adquisició de les relacions espacials.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Task, task complexity and task sequencing

In the last two decades there has been a growing interest in the use of tasks as a tool for language learning, principally because of the potential of pedagogic tasks to approximate L2 performance in real conditions (Breen, 1989; Bygate, Skehan & Swain, 2001; Crookes, 1986; R. Ellis, 2003; Long, 1985; Nunan, 2004; Prahbu, 1987; Richards, Platt & Weber, 1985; Robinson, 1995a; Skehan, 1996). Through the use of tasks, learners gain an opportunity to learn the L2 gradually by preparing themselves for real life needs as users of a second or foreign language. In this context, the need to give a clear definition of what is meant by a pedagogic task emerged, and it turned out to be a challenging endeavor.

Some researchers have provided all-inclusive definitions of what a task is, such as the one proposed by Long (1985:89): "A task is a piece of work undertaken for oneself or for another, freely or for some reward […] by "task" is meant the hundred and one things people do in everyday life, at work, at play, and in between" (Long, 1985:89). However, this definition may be applied to any kind of task, including those which are independent from language, such as painting a fence; therefore, a narrower definition of a task was needed, where the framework for its use was reduced (Crookes, 1986:1): a task is "a piece of work or an activity, usually with a specified objective, undertaken as part of an educational course, at work, or used to elicit data for research"

The next important step in defining a task was made by adding communicative characteristics, use in the classroom and a focus-on-meaning rather than focus-on-form goal. All of these points were reflected in the definition given by Nunan (2004:10): a communicative task is "a piece of classroom work which involves learners in comprehending, manipulating, producing, or interacting in the target language while their attention is principally focused on meaning rather than on form. The task should also have a sense of completeness, being able to stand alone as a communicative act in its own right".

Finally, R. Ellis (2003:16) defined a task as "a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed". To sum up, a task represents a pedagogic unit which should be designed according to some real-life conditions a learner will possibly need to deal with in the future; it is characterized by a primary focus on meaning, though focus on form might be also present in the process of completing the task[1].

In order to use the task for learning purposes in classroom or laboratory settings it has been widely acknowledged (Candlin, 1987; Long, 1985; Long & Crookes, 1992; Nunan, 2004; Prahbu, 1987; Robinson, 1995a; Skehan, 1996, 1998) that there exists a need to organize tasks in a well-reasoned, logical way which would help learners in the process of language learning. In the field of applied linguistics there is not yet general agreement as to what the criteria designers and teachers

---

[1] These are not all the definitions of "task" but rather a selection of definitions, which are well-accepted in the field.

should use to organize and sequence pedagogic tasks. While some scholars suggest that difficulty of input or content should be the main criteria for sequencing, other scholars have created models of task sequencing that organize pedagogic tasks from cognitively simple to cognitively complex.

For the present study, Robinson's (2005, 2007b, 2010) model of task sequencing has been chosen because it is widely considered a theoretically well-grounded and systematically organized proposal of task sequencing and is based on many previous studies on tasks, L2 acquisition and an array of empirical and theoretical studies in psychology. In his proposal Robinson (2005, 2007b, 2010) claims that tasks should be organized and sequenced by means of manipulating their cognitive complexity, that is, the conceptual, attentional, memory and reasoning demands that the structure of tasks imposes on the learner's processing.

The construct of cognitive task complexity was developed by Robinson into the Cognition Hypothesis (2001b, 2003, 2005, 2007b), where tasks could be cognitively manipulated along two dimensions: resource-directing variables, which aim to direct learners' attention to some of the specific linguistic features for completing the task (e.g. the amount of causal reasoning demands) and resource-dispersing variables that disperse learners' attention among different linguistic features of the language (e.g. the amount of pre-task planning time). Both variables are equally important for L2 learning since tasks designed on the basis of resource-directing variables are predicted to be good for practicing new linguistic information, whereas tasks based on resource-dispersing variables are thought to automatize the learners' pre-existing linguistic repertoire.

Although many scholars have stressed the need to investigate task sequencing, since it is an integral part of tasks' implementation into syllabus design, it is still a new trend in Second Language Acquisition (SLA) and in Task-Based Language Teaching (TBLT) and it is empirically under-researched. Therefore, the first main objective of the present dissertation is to test Robinson's (2010) SSARC[2] model in order to assess whether sequencing tasks from simple to complex on the basis of their cognitive complexity is more beneficial for L2 development than other alternative ways of task sequencing, such as presenting tasks in a randomized order. In this study, the focus is on the learning of language related to spatial relations, where two sequences of tasks (a sequence from simple to complex versus a randomized sequence) are tested to find out whether target task sequencing as proposed by Robinson (2005, 2007b, 2010) has any significant benefits for the learning and retention of spatial expressions. The second goal of the study is to analyze the role that individual differences may play in the development of spatial expressions. To my knowledge, no studies have looked at the effects of task sequencing on L2 development and working memory capacity simultaneously. In this sense, the present dissertation aims to contribute new data to our understanding of how tasks in a sequence may impact L2 development.

1.2 Working memory capacity in L2 development

The construct of working memory capacity was created by Baddeley & Hitch (1974) and completed by Baddeley (2000). Working memory represents a process of simultaneous processing and retention of information and consists of the phonological loop, the audio-visual sketchpad, executive control and the episodic

[2] SSARC: SS – "simple/stabilizing interlanguage"; A – "automatizing access to interlanguage"; RC – "restructuring and complexifying of the learner's current level of interlanguage ». The detailed explanation of the model will be provided in Chapter IV.

buffer. Several spans were created to measure this construct; the most popular is the reading span test (Daneman & Carpenter, 1980). Initially, this construct was empirically tested in psychological studies, however, in recent years, several scholars in SLA have also focused their research on WMC as a mediating factor in L2 production, development and acquisition.

Applied linguistics and psychologists previously supported this interest by highlighting the close relationship between WMC and SLA (N. Ellis, 1996; N. Ellis & Schmidt, 1997; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Robinson, 2003; Skehan, 1996). Robinson (1995b:283) has argued that "individual differences in memory and attentional capacity both affect the extent of noticing, thereby directly influencing SLA". As it was seen that WMC is highly compatible with information-processing models (Harrington, 1992), such as the Multidimensional Model (Pienemann & Johnston, 1987; Pienemann, Johnston & Brindley, 1988), the Three-stage Model of second language acquisition (DeKeyser, Salaberry, Robinson, & Harrington, 2002; VanPatten, 1996; VanPatten & Cadierno, 1993a, 1993b), and the Cognition Hypothesis (Robinson, 2001a, 2001b, 2003, 2005, 2007a), several empirical studies were carried out in its relationship with other concepts, such as L2 performance and WMC (Bergsleithner, 2010; Gilabert & Muñoz, 2010; Kormos & Sáfár, 2008), recasts and WMC (Révész, 2012), output and WMC (Mackey, Adams, Stafford & Winke, 2010) and L2 development & WMC (Mackey, Philp, Egi, Fujii & Tatsumi, 2002; Sagarra, 2007). Many of these studies detected a significant relationship between greater WMC and better L2 performance, a higher number of correct forms retained from recasts and a higher level of L2 development. However, although WMC is directly connected to the concept of cognitive task complexity and task sequencing since it forms part of cognitive processing, no studies up until now have specifically looked at L2 development through task sequencing as mediated by WMC.

In order to fill the gap in the empirical studies on task sequencing with WMC as its mediator, the objective of the present study is two-fold: the first aim is to explore whether task sequencing plays a significant role in L2 development in the use of spatial expressions. The second aim is to examine the mediating role that working memory capacity may play on L2 development of the use of spatial expressions through task sequencing.

1.3 The present study

The dissertation is organized into eight chapters. Chapter II explores the construct of task complexity where, first, the concept related to fields other than SLA is defined to offer a linguistically independent view of how task complexity may be understood. Furthermore, a more specific definition of task complexity applicable to L2 performance and development as forwarded by Robinson (2001a, 2001b, 2003, 2005, 2007a) is provided. The definition of task complexity is completed by the presentation of two competing theories of task complexity: Robinson's Cognition Hypothesis and Skehan's Trade-off Hypothesis. The chapter ends with the justification for choosing Robinson's Cognition Hypothesis in the present dissertation and with a very brief overview of the Cognition Hypothesis in relation to task sequencing and individual differences.

In Chapter III, the construct of spatial reasoning demands chosen for the cognitive manipulation in this study is defined on the basis of a body of psychological literature as well as a selection of empirical studies in cognitive psychology and applied linguistics. Finally, a series of tools most extensively used in applied linguistics to independently measure task complexity are reviewed (an affective variable questionnaire and a time estimation task, among others).

Chapter IV explains the concept of task sequencing, first in fields other than applied linguistics such as mathematics or computer modeling, where task sequencing is already a well-established construct. Further, a detailed description of Robinson's SSARC model (2010) is laid out, which, up to date, is the most complete model of task sequencing for L2 learning, and offers a proposal for how to organize pedagogic tasks within a syllabus in order to optimize learners' gains in their learning process through tasks. The final part of the chapter is devoted to a few studies carried out recently which are related to task sequencing in various ways.

The construct of WMC is presented in detail in Chapter V. First, several models of WMC are discussed, with Baddeley's model of WMC (Baddeley, 1986; Baddeley, 2000; Baddeley & Hitch, 1974) as the selected model for the present study. The theoretical framework of the construct is followed by the explanation of different measures of WMC (reading span, counting span, and arithmetic span, among others), where a special section, 5.3.2, is given to the reading span and its automated version as it is the measure selected for the present study due to its generalizability and high correlation with other WM spans. Finally, some empirical studies on WMC and L2 development are discussed.

The design of the present dissertation and its methodology are explained in Chapter VI. First, task complexity and task sequencing operationalization issues are tackled, followed by the justification for the selection of target items through three pilots also briefly described in the chapter. Second, the design of input and treatment materials is justified and their elaboration is shown through a series of pilots. Third, the detailed description of the design of the study (participants, final materials for input, treatment tasks and control tasks, and procedure) is provided.

In Chapter VII, the results of the tests performed by the participants are presented. They are divided into two sections, one for classroom groups and one for laboratory groups. The two tests (a descriptive task and a two-part vocabulary test) used for the analysis of gains during treatment are reported separately. Results of WMC for laboratory groups are given at the end of the chapter.

The results reported in Chapter VII are discussed in Chapter VIII. First, the results for the descriptive task of the two contexts (two laboratory groups and two classroom groups) in two times (posttest and delayed posttest) are presented, followed by the results for the two-part vocabulary test in the two settings (laboratory and classroom) obtained in two times (posttest and delayed posttest). The discussion concludes with the results of WMC in relation to L2 development of the two laboratory groups immediately after treatment and over time. Finally, at the end of Chapter VIII a conclusion for the entire study is given with the summary of the main findings of the dissertation, the limitations of the present study and some directions for future research. At the end a list of bibliographical references followed by a list of appendices close this dissertation.

CHAPTER II

TASK COMPLEXITY

2.1 Introduction

As stated in the introduction, organizing tasks into a sequence in syllabus design should be based on theoretically justified and empirically testable criteria. For the present study, the decision was made to sequence tasks according to their cognitive complexity as it is considered to be a reasonable criterion for task organization. For example, while we cannot predict individual differences or proficiency levels before a program starts, we can decide on the internal features of tasks, since they are invariant for all learners (in other words, a complex task will be more demanding than a simple task for everyone, regardless of what they bring to the task). For that reason, this chapter focuses on the construct of task complexity and its general definition, subsequently completed by a more specific definition of task complexity as elaborated by Robinson (2001a, 2001b, 2003, 2005, 2007a, 2011). The latter definition is specifically applicable to L2 performance and development. In what follows, first, a brief overview of research issues that preceded task complexity is provided, along with the definition of task complexity and its distinction from task difficulty. Next, two competing theories of task complexity are presented: Robinson's Cognition Hypothesis and Skehan's Trade-off Hypothesis and the justification for choosing Robinson's Cognition Hypothesis for the present study is outlined. Finally, the most recent model of task complexity, the SSARC model (Robinson, 2005, 2007b, 2010) and the role of individual differences in the context of task complexity are laid out (Robinson, 2001c).

2.2 History behind task complexity

The construct of task complexity is currently widely accepted and used in SLA research (Cadierno & Robinson, 2009; R. Ellis, 2005; Ishikawa, 2007; Gilabert, 2005; Gilabert, 2007a, 2007b; Kim, 2009; Kuiken, Mos & Vedder, 2005; Kuiken & Vedder, 2007a, 2008; Michel, Kuiken & Vedder, 2007; Rahimpour, 2007; Robinson, 1995a, 2001a, 2005, 2007a, 2011; Robinson & Gilabert, 2007; Yuan & Ellis, 2003) and it is also at the core of the present study. However, the origins of this construct preceded its implementation into SLA. For this reason, the decision has been to introduce the construct of task complexity, first, independently from the L2 theories. Below, some of the crucial articles from cognitive psychology related to task complexity with a special focus on human performance are reviewed.

The first key article tackling the definition of task complexity was written by Wood (1986:60). The main goal of the article was "to describe one important characteristic of tasks, complexity". He criticized some previous studies (Hackman & Lawler, 1971; McCormick, 1976; Peterson & Bowners, 1982) which had searched the defining characteristics of task complexity by resorting to their empirical findings, which led to confusion between pure task characteristics (such as formal task structure) and non-task characteristics (e.g. group organization requirements). Consequently, the characteristics themselves and their numbers were different from study to study, which made it impossible to come up with a valid classification of tasks for their use in research. An alternative approach was needed to use a theoretical framework where task complexity was defined *a priori* and then implemented into research. Among several theoretical frameworks, Wood (1986) leaned towards the use of "behavior requirements" and "task qua task" framework (Hackman, 1969), which was likely to be more easily operationalized and at the same time more suitable for developing a task complexity construct. Assuming that a task contained products, specific acts and information cues, task complexity was considered to be "an important determinant of human performance through the demands it places on the knowledge, skills, and resources of individual task performers" (1986:66). Next, Wood defined three types of task complexity:

1) *component complexity* – "a direct function of the number of distinct acts that need to be executed in the performance of the task and the number of distinct information cues that must be processed in the performance of those acts" (1986:66).

2) *coordinative complexity* – "the nature of relationships between task inputs and task products" (1986:68).

3) *dynamic complexity* – "is due to changes in the states in the world which have an effect on the relationships between task inputs and products" (1986:71).

The sum of these three types determined the complexity of the task and also the knowledge that would result from it and from the individual skills of the task performers. Furthermore, for each of the types of task complexity, Wood (1986) provided a specific formula, followed by a combined formula for total task complexity, which could be calculated with precision. Such an analytical framework was seen as being beneficial for the generalizability of task effects in different cognitive areas of human performance (from piloting a plane to making a medical decision). Wood's model of task complexity was widely used in empirical studies (e.g. Johnson & Kanfer, 1992; Simnett, 1996; Topi, Valacich & Hoffer, 2005).

Campbell (1988) elaborated another well-known model of task complexity. It identified four task qualities, which made possible the distinction among tasks with different levels of complexity: (1) *multiple paths*, many ways to arrive to a desired outcome, which increase task complexity; (2) *multiple outcomes*, if the number of possible outcomes increases, the complexity increases as well; however, if the outcomes are positive (i.e. interchangeable), it makes the task less complex; (3) *conflicting interdependence among paths*, if two outcomes are desirable the conflict may appear in achieving both paths at the same time, which increases task complexity; (4) *uncertain or probabilistic linkages*, which could also affect information load and make a task more complex. Based on these task attributes, Campbell offered a hypothetical typology of complex tasks with a total of 16 variations (see Table 1).

Table 1. *Complete typology of complex tasks*

	Sources of Complexity			
Task Type	**Source 1**	**Source 2**	**Source 3**	**Source 4**
1.	-	-	-	-
2.	X	-	-	-
3.	-	X	-	-
4.	-	-	X	-
5.	-	-	-	X
6.	-	-	X	X
7.	-	X	-	X
8.	X	-	X	-
9.	-	X	X	-
10.	X	-	-	X
11.	X	X	-	-
12.	-	X	X	X
13.	X	-	X	X
14.	X	X	-	X
15.	X	X	X	-
16.	X	X	X	X

Source 1 = presence of multiple paths to a desired end-state
Source 2 = presence of multiple desired end-states
Source 3 = presence of conflicting interdependence
Source 4 = presence of uncertainty or probabilistic linkages

According to this typology tasks could be just simple tasks, with no traces of task complexity; decision tasks (e.g. choosing a flat from a number of options); judgment tasks, which involve situation analysis (e.g. stock market analysis); problem tasks (e.g. personnel scheduling); fuzzy tasks (e.g. business ventures). All these types of tasks are linked to "objective complexity" (*a priori* complexity, where a task is intrinsically complex for any task doer), as opposed to "subjective complexity"[3] (*a posteriori* complexity, where the perception of its difficulty may vary from doer to doer).

Further research into the construct of task complexity (Bonner, 1994) used similar paths and replications of Wood and Campbell's task complexity models. A very recent study by Liu and Li (2012) also offered a definition and a model of task aiming at finding an objective and universal construct definition. According to Liu and Li (2012:559) task complexity can be defined as "the aggregation of any intrinsic task characteristic that influences the performance of a task". The general task model forwarded by the two authors consisted of six components: (1) *goal* - a completion of the task or an attainment of a proficiency level; (2) *input* – as seen as its clarity, quantity, diversity, accuracy, rate of change, redundancy, mismatch etc.; (3) *process*, such as clarity, quantity of paths, repetitiveness among others; (4) *output*; (5) *time* – concurrency and pressure, and (6) *presentation* – format, heterogeneity, and compatibility. Finally, ten dimensions of task complexity were distinguished: size (number of task components), variety, diversity in terms of the number of task components, ambiguity, variability (changes and unstable characteristics of task components), unreliability, novelty, incongruity, action complexity (cognitive and human actions during the performance of a task), and temporal demand.

[3]This issue will be thoroughly explained in the next section.

When analysing different models, several problems were identified. The first problem appears when separating objective task complexity from subjective task complexity or task complexity from task difficulty. The second conceptual problem arises when defining task complexity, since each study had its own operational definition. So, a different conceptualization of task complexity was observed not only in different fields, but also even within the same domain. These general task complexity definitions constitute the basis of what Robinson (2001a) defined as task complexity in the context of L2 performance and development.

2.3 Theories on task complexity in TBLT

As an alternative to the linguistic unit in synthetic syllabi, the concept of task was developed in analytical approaches to language teaching (Long & Robinson, 1998; R. White, 1988, Widdowson, 1979; Willis, 1990) and the need to organize tasks in a logical way in order to transfer them to the classroom context was raised. A number of researchers who take an analytical approach to syllabus design have argued that tasks should be delivered not on the basis of the linguistic demands, but by approximating them to real-world target tasks (Crookes, 1986; Long, 1985; Nunan, 2004; Prahbu, 1987; Skehan, 1996, 1998).

Though it was generally assumed that in order to take advantage of task-based language learning tasks should be organized and gradually sequenced from simple to complex (Long, 1985; Nunan, 2004; Skehan, 1996, 1998, 2009) to optimize conditions for language learning and task performance, to date, there is no consensus as to how to organize the tasks, or on which basis to distinguish and sequence them. In the context of SLA two models have been developed, the Cognition Hypothesis by Robinson (2001a, 2001b) and the Trade-off Hypothesis (also known as the Limited Attentional Capacity Model) by Skehan (1996, 1998, 2009). For the present research, Robinson's Cognition Hypothesis (2001a, 2001b, 2003) has been chosen as it contains a framework for task sequencing criteria which is feasible and testable for syllabus and task design in the context of L2 learning. As seen below, this framework also advances a research agenda to investigate tasks characteristics and their effects on L2 performance and development. Below, the two above mentioned models are explained and the justification for choosing Robinson's Cognition Hypothesis is given.

2.3.1 Theory behind Robinson's Cognition Hypothesis

Robinson's Cognition Hypothesis makes predictions on two issues: (1) it predicts the effects task complexity may have on L2 performance and comprehension, and (2) it predicts the effects task complexity may have on L2 development through tasks graded according to their cognitive demands. Robinson makes a distinction of two dimensions of task complexity, resource-directing dimensions, which direct learner's attention to particular linguistic features of a task, and resource-dispersing dimensions, which deplete learner's attention over the different elements of the tasks. An example of a resource-directing dimension is reasoning demands, where tasks do not demand reasoning from learners, but just represent a simple transmission of information, require less conceptual and therefore linguistic effort and resources than a task with some reasoning demands, where at least cause-consequence subordination (e.g. *because*, *therefore*) is needed. Similarly, in the case of a direction-giving map task, dealing with a well-organized map with all the landmarks in place is easier and does not require a great mental effort from a learner as compared to dealing with a confusing, fuzzy map with no clear landmarks with multiple perspectives and interpretations (Becker & Carroll, 1997). In Robinson's view, increasing cognitive complexity along this type of dimensions will affect fluency negatively, but will promote accuracy and complexity[4].

An example of a resource-depleting variable is access to planning time during task performance, as giving no planning time increases the complexity of a task by simply dispersing attentional resources over the different aspects of the task. However, this dimension is also seen as important for syllabus design, as it prepares learners for real-life conditions, so "practice along them could be argued to facilitate real-time access to an already established and also to a developing repertoire of language"(Robinson, 2003:59).

As far as the predictions of the Cognition Hypothesis on L2 language production in the case resource-directing variables are concerned, Robinson (2001b, 2003, 2005, 2007b), resorting to the claims by Givon (1985, 1995, 2002), Klein and Perdue (1992, 1997), Perdue (1993a, 1993b) and Rohdenburg (1996), argued that task complexity negatively affects fluency, but promotes accuracy and complexity. Manipulating certain characteristics (e.g. the number of elements, the amount of reasoning) may direct attentional and memory resources to task completion and therefore generate more accurate and more complex speech; at the same time, fluency would be negatively affected. Moreover, in an interactive task, as opposed to

[4] Complexity, accuracy, and fluency (CAF) are dimensions of task performance that are largely used in TBLT literature. Although there is some reconsideration as to how to precisely measure these three dimensions (Norris & Ortega, 2009; Pallotti, 2009; Skehan, 2009), by means of general or specific measures, there is a consensus regarding the fact that these dimensions are valid for the analysis of oral and written task performance.

a monologic task, task complexity should also result in an increased use of comprehension checks and clarification requests and other interactive moves.

Regarding resource-dispersing dimensions, the prediction is made toward a negative effect of an increase in task complexity on all aspects of L2 production, however it will enhance interaction (see Table 2).

Table 2. *Effects of task complexity on CAF*

Dimension	Type of task	Measure of L2 production			Interaction
		Fluency	Complexity	Accuracy	
Resource-directing	Simple	+	+	+	-
	Complex	-	+	+	+
Resource-dispersing	Simple	+	+	+	-
	Complex	-	-	-	+

To sum up, the Cognition Hypothesis, proposed by Robinson, offers a model of task design and organization according to increasing cognitive demands along two kinds of dimensions, resource-directing and resource-depleting. Although these dimensions affect L2 performance in two completely different ways, both are essential for L2 development and training, since they bring a learner closer to real-life target task conditions, such as no planning time or multiple conditions put together.

Importantly, Robinson (2003) stated that the Cognition Hypothesis is only applicable to adult language development. First, Robinson believes that adults have limited access to innate knowledge, unlike children learning their L1 who possess this capability (Carey & Spelke, 1994; L. White, 2003), which according to some researchers gradually disappears after the Critical Period (Hyltenstam &

Abrahamsson, 2003; Long, 1990). Second, adults have developed their metacognitive and metalinguistic skills to a greater extent than children, which are essential for classroom L2 learning (Bialystok, 1991; DeKeyser, 2000; Harley & Hart, 2002). However, when grading tasks according to their cognitive demands, Robinson claims that it is important to follow a "natural order" for sequencing that is the order of cognitive and linguistic development. For example, in the case of reasoning demands, a sequence may start with tasks that require spatial reasoning based on simple topological relations and finish with tasks that include axis-based relations (Taylor & Tversky, 1996), which is the way children acquire this relationship (Becker & Carroll, 1997). And the example is the case of +/- causal reasoning, in which sequencing may be organized from tasks with no casual relations to tasks that establish those relations (Bergman & Slobin, 1994; Niwa, 2000) also following human cognitive development.

Associated with the Cognition Hypothesis Robinson has advanced the triadic componential framework (2001b, 2007b), which distinguishes among task complexity factors, task condition factors, and task difficulty factors. In what follows, a schematic perspective of the Cognition Hypothesis, and its elements organized into categories within the triadic componential framework (its first version and an updated version) is presented (Robinson, 2001a, 2001b; Robinson & Gilabert, 2007).

2.3.1.1 Triadic componential framework

The triadic componential framework represents a model of reference which guides the decision on how to grade tasks during syllabus design on the basis of their task complexity (1) and helps in making online decisions according to task conditions (2) and task difficulty (3). Task complexity features are related to the internal structure of tasks, whereas task conditions have to do with the information distribution and online implementation, and task difficulty to the individual differences learners may bring into the task. As seen in Section 2.3, it is important to note that even though task complexity and task difficulty tightly interact with each other they should be distinguished from each other. Additionally, these two elements should also be distinguished from task conditions.

As described above, task complexity contains two dimensions: (a) resource-directing and (b) resource-depleting. In the first version of the triadic componential framework, Robinson (2001b) distinguished among three resource-directing variables (i.e. +/- few elements; +/- here-and-now; +/- no reasoning demands) and three resource-depleting variables (i.e. +/- planning time; +/- single task; +/- prior knowledge). All these variables represented the recommendation for sequencing criteria, as well as the guidance for future decisions on task design as they can all be manipulated before syllabus implementation. For instance, a task can be complexified by adding more elements or by forcing displayed past tense reference. It can also be complexified by giving no pre-task planning time to learners or by giving them unfamiliar content.

Task difficulty considers learner characteristics from two different perspectives: (a) unstable and unfixed affective variables that can be changed in a relatively short period of time (e.g. motivation, anxiety and confidence), and (b) relatively stable ability variables, which could be even measured beforehand (e.g. aptitude, proficiency and intelligence). These are the features learners bring to task performance and typically little can be done about them before syllabus implementation.

Finally, task conditions are seen as an interaction between learner and task factors, and classified in terms of participation (e.g. open vs. closed task; one-way vs. two-way task; convergent vs. divergent task) and participant factors (e.g. gender, familiarity, power vs. solidarity). They determine the best way to pair up learners in order to maximize learners' interactive conditions. Again, those are difficult to decide on before syllabus implementation and decisions about these conditions are taken online.

Though assuming that task complexity may be controlled and learner abilities measured in advance, some interactions between the three parts of the model can still occur. In this way, task complexity could interfere with learners' perception of task difficulty, or else task condition, for instance, whether information flows one-way or two-way, may affect learners' perception of the difficulty of the task. Below is Figure 1 from Robinson (2001b), which illustrates the first version of the triadic componential framework.

Figure 1. *Task complexity, condition and difficulty (Robinson, 2001b)*

Task Complexity	Task conditions	Task difficulty
(cognitive factors)	(interactional factors)	(learner factors)
(a) resource-directing	(a) participation variables	(a) affective variables
e.g., +/- few elements	e.g. open/closed	e.g., motivation
+/-here-and-now	one-way/two-way	anxiety
+/-no reasoning demands	convergent/divergent	confidence
(b) resource-depleting	(b) participant variables	(b) ability variables
e.g. +/- planning	e.g. gender	e.g. aptitude
+/- single task	familiarity	proficiency
+/- prior knowledge	power/solidarity	intelligence

Later, Robinson and Gilabert (2007) updated the existing version of the triadic componential framework by adding some new elements to it. One of the objectives of the article was to create an operational taxonomy for task type classification suitable for task designers, which could be further adapted for the classroom context. Therefore, the items selected to be included in the new framework were meant, first, to be adequate for pedagogic design and the use in classroom setting, and second, to be theoretically justified. The ultimate aim was to promote, by means of task variables, the retention and automatization of new features and to create quick access to already existing L2 knowledge. As in the previous version, Robinson and Gilabert (2007) made a distinction between task complexity, task difficulty, and task conditions, and within each category a subdivision was done. As we saw previously, each category contains two sub-groups; *task complexity* is divided into resource-directing and resource-depleting variables, *task conditions* are distinguished between participation and participant factors, and *task difficulty* contains affective and ability factors, listed in Figure 2.

Here again, Robinson reconfirms his predictions regarding L2 production and L2 development[5] affected by task complexity. As in previous studies (2001a, 2001b, 2003, 2005) he claimed that increased task complexity along resource-directing dimensions will promote complexity and accuracy, but will negatively affect fluency. In contrast, manipulating task complexity along a resource-dispersing variable will negatively affect all three dimensions of L2 performance. Further, Robinson made a prediction about synergetic effects of task complexity on L2 production, where when two types of variables (resource-directing and resource-

[5]The issue of L2 development and task complexity will be discussed in detail in Chapter 4.

dispersing) are combined a weakened or even neglected effect of resource-directing variables caused by resource-depleting dimensions can be observed.

Figure 2. *The Triadic componential framework for task classification – categories, criteria, analytic procedures, and design characteristics (Robinson & Gilabert, 2007)*

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

A different view regarding the effects of task complexity on L2 production was offered by Skehan with the alternative Trade-Off Hypothesis (1998, 2009), whose tenets are presented in the next section.

2.3.2 Skehan's Trade-off Hypothesis

The Trade-off Hypothesis is based on a series of assumptions regarding individual differences of the learner in relation to performance of tasks with different levels of cognitive complexity. Since WMC and attentional resources are limited, they directly affect and, in a way, impose some limits on L2 performance. As already mentioned, L2 production is usually analyzed in terms of fluency, complexity and accuracy. So, if we assume that performance in each of these areas requires attention and working memory (which are, in turn, limited), they compete among themselves for attentional resources. Form and fluency are in competition, but most importantly, complexity and accuracy compete by default with one another in situations of increased cognitive complexity of tasks. This is where the most significant difference with the Cognition Hypothesis lies, since the latter postulates simultaneous improvements in complexity and accuracy under higher cognitive load of tasks.

Sometimes, as explained by Skehan (2009), the design or the conditions of the tasks could soften the trade-off effects by showing joint advantageous results from task complexity in the case of complexity and accuracy. This was shown, for instance, in tasks with clear structure and background and foreground information provided before completing the task, which led to more accuracy and complexity simultaneously. Regarding task conditions, planning has been shown to jointly raise complexity and accuracy. These particular cases suggest that "the generalization that accuracy and complexity rarely go together, and that this reflects the consequence of limited attention, may not always apply" (Skehan, 2009:14).

As Skehan's account puts research before theory, a general limited WMC of an adult learner is assumed; however, empirical studies are required to shed light on the specific nature of each of the task characteristics and conditions. As seen from a few studies conducted by Skehan (see 2009 for an overview), each task condition or characteristic may have its particular effect on learners' production and also some combinations of these dimensions of a task could result in a different way of performing the task in terms of complexity, accuracy and fluency.

2.3.3 Justification for choosing the Cognition Hypothesis

As observed above, Robinson's proposal of task sequencing is theory-driven and it is based on the Cognition Hypothesis. Such a framework draws on an extensive literature going beyond applied linguistics studies which includes research from psychology, psycholinguistics, sociology and even ergonomics, among others. Additionally, a well-structured triadic componential framework updated on the basis of theoretical as well as empirical studies sets out the agenda for conducting empirical studies. While Skehan provides an interesting reflection on what may affect performance, Robinson comes up with a specific agenda for researching tasks, heavily supported by theory. This agenda allows researchers to have a clear representation of the framework within which to design an experiment related to task complexity, and also gives a theoretical basis for the justification of the findings. Finally, the model provides a wide range of task features logically organized into three categories, which are also clearly separated (task complexity, task conditions, and task difficulty). This helps researchers to avoid possible confusion when putting them into practice, especially in the case of "task complexity" in contrast to "task difficulty". So it makes the model easily treatable both by researchers and syllabus designers. For all these reasons the decision was made to base the experimental design of the study on the Cognition Hypothesis.

Before going into the description of task complexity in relation to task sequencing and individual differences, it is important to make the distinction between the two key concepts of the study: task complexity and task difficulty. The following section will be devoted to the presentation of these concepts and the definition of their main characteristics that differentiate one concept from the other.

2.4 Definition of task complexity: task complexity versus task difficulty

Task complexity and task difficulty are often used interchangeably because of the apparent similarity of the constructs they refer to, which leads to a widespread confusion of the terms. The distinction is still not unanimous and a vivid debate exists around the two concepts. In order to summarize such a debate, four viewpoints are given below.

As task complexity and task difficulty have an apparently similar basis, some researchers consider task complexity as a part or sub-element of task difficulty, since a complex task tends to be difficult, but a difficult task is not necessarily always complex one (Byström, 1999). Others, on the contrary, assume that task difficulty is a part or a sub-concept of task complexity along with task structure (Bonner, 1994). This distinction is opposed to a more radical one, where task complexity is interchangeable with task difficulty (Hendy, Liao & Milgram, 1997) or, on the contrary, task complexity and task difficulty represent two absolutely different components (Robinson, 2001a).

Robinson (2001a) distinguished task complexity from task difficulty by the fact that task complexity is defined in terms of the internal structure of the task while task difficulty is defined in terms of the perception by the learners. Therefore, for Robinson (2001a:29) task complexity is "the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner. These differences in information processing demands, resulting from design characteristics, are relatively fixed and invariant". As task complexity is a relatively stable component, in the sense that some tasks are intrinsically more or less complex for everyone, it may serve as a point of reference to task design and sequencing before their implementation into a syllabus. Task difficulty, in contrast, is directly dependent on the learners, since different learners could consider the same task more or less difficult. This is explained, on the one hand, in terms of the differences in cognitive abilities among learners, such as attentional control and memory, and to what extent the learners have attentional and memory and other cognitive resources available to them and, on the other hand, by affective factors, such as confidence, motivation, or anxiety, which can also influence the perception of difficulty of a particular task.

Recently most researchers have adopted the position of Robinson (2001a, 2003, 2005, 2007a) regarding the need to differentiate the two concepts (Aula, Khan & Guan, 2010; Liu, Gwizdka, Liu & Belkin, 2010), pointing out the objectivity of task complexity as compared to the subjective perception of task difficulty by a particular individual. In line with this idea, in the present study the two concepts are also differentiated and treated separately, following the tenets of Robinson's Cognition Hypothesis, described in Section 2.4.1.1. After having separated the construct of task complexity from task difficulty, the next section will deal with different theories of task complexity related to second language learning tasks.

2.5 Task complexity and the SSARC model

Robinson (2005:3) argues that language development based on tasks can occur only when "both cognitive processing and interactive consequences of task sequencing" are equally involved so that, first, gradual processing and retention of new input happens, which subsequently facilitates higher quality output. The justification of the theoretical basis of the Cognition Hypothesis, where the main components are cognitive and interactive demands, lies on a series of research areas, deeply examined by Robinson, such as L1 and L2 development and cognition (Becker & Carroll, 1997; Schmidt, 2001; Slobin, 1993; Tomasello, 2003), cognitive linguistics and cognitive psychology (Carlson, 1997; Givon, 1985).

The main prediction of the Cognition Hypothesis is that increasing cognitive demands along some of the dimensions (1) promotes learners' complexity and accuracy in L2 performance; (2) generates interaction processes through which retrieval and retention of new target forms from the input occurs; (3) affects L2 development when mediated by individual differences, whether cognitive (e.g. working memory) or affective (e.g. motivation). As described above, these dimensions are divided into two groups: resource-directing variables and resource-dispersing variables, which are equally important in the process of L2 development and acquisition. On the basis of this distinction, Robinson elaborated two principles of task sequencing:

1. “Only the cognitive demands of tasks contributing to their intrinsic conceptual and cognitive processing complexity are sequenced” (Robinson 2010:247). That means that when putting tasks into a sequence, those that require few cognitive demands from the learners go first in a sequence followed by those that require higher cognitive demands from them. Therefore, the sequence is built with an increase in the cognitive demands that the structure of the task imposes on learners.

2. “Increase resource-dispersing dimensions of complexity first, and then increase resource-directing dimensions” (Robinson, 2010:247). The two principles represent the core of the SSARC model of task sequencing. According to Robinson (2010), four steps must be taken in order to ensure benefit for L2 development by sequencing of tasks. Tasks with no cognitive demands should go, firstly, “**simple and stable**” (SS); secondly, tasks are gradually increased in complexity along resource-dispersing variables, which facilitates their “**automatization**” (A) of the current repertoire already available to the learners. Finally, tasks are increased in complexity on both dimensions (resource-directing and resource-dispersing), which helps “**restructuring**” (R) the already acquired forms, and “**complexifying**” (C) speech by means of the new target forms (see Figure 3).

Figure 3. *Robinson's SSARC model*

	Name	Description
Step 1	SS: simple and stable	no cognitive demands
Step 2	A: automatization	resource-dispersing variables
Step 3	R: restructuration	resource-dispersing + resource-directing variables
Step 4	C: complexity	maximum complexity at all levels

These four steps constitute the SSARC model of task sequencing based on the Cognition Hypothesis elaborated by Robinson (2010). When thinking of the triadic componential framework described previously (section 2.4.1.2), it is important to distinguish between the functions of two kinds of dimensions (resource-directing variables and resource-dispersing variables) in relation to task sequencing. The resource-directing variables help to train newly acquired forms of the L2 in the following ways:

1) This type of variables directs learners' attention to some of the specific features of a target language, and so it helps to retain new grammatical or lexical forms. When manipulated from simple to complex, it leads learners to a gradual uptake and retention of some new target lexical and / or grammatical forms needed to complete a particular task. As an example, when a task is not manipulated along causal reasoning demands, learners are not led to use any subordination to transmit event relationships; instead a simple narration of facts is required. However, when it comes to tasks where causal reasoning is required to order a sequence of events or to potentially explain somebody's actions, causal conjunctions are hardly avoidable in these cases (*because, despite, so as to, since* etc.).

2) Sequencing from simple to complex along resource-directing variables represents a natural order of acquiring new L1 concepts in childhood and, therefore, corresponds to a natural way of sequencing L2 tasks for adult learners. For example, sequencing tasks along Here-and-Now / There-and-Then completely correspond to the acquisition of the references of temporality by children, who, first acquire a reference system corresponding to the present and later on to the past. The fact is well documented in linguistic and psychological studies (Bronchart & Sinclair, 1973; Shirai & Andersen, 1995; Weist, 1989). The parallelism between L1 and L2

development could be justified by some previous theories claiming that adult learners acquire a new language on the basis of the already established scale of conceptual complexity in their minds. Therefore, sequencing along resource-directing variables corresponds to learners' internal initial expectations.

In contrast, resource-dispersing variables do not promote any attention to features of new target forms, but they help to create a real-time access to the already existing repertoire in the L2. In this way, they are necessary and helpful in L2 performance and in L2 development, as they allow automatic access to already acquired forms of a target language. So, having unfamiliar elements to deal with in a task disperses learners' attention and memory resources to many features and so promotes learners' automatization of lexical and grammatical items to be able to complete this kind of tasks.

To sum up, the two dimensions which are discussed above are equally important in L2 development, but they have a distinct role in task sequencing, where resource-directing variables promote L2 development of new lexical and/or grammatical features, while resource-dispersing variables are responsible for the automatization of new L2 features[6].

2.6 Task complexity and individual differences

There has been an intense debate in the field of applied linguistics about whether acquisition is related to unconscious processes and, thus, to implicit learning unrelated to cognitive abilities or, on the contrary, the level of L2 development achieved by adults L2 learners is more substantially related to L2 learners' individual

[6] Task sequencing will be treated in depth in Chapter IV.

differences. Krashen (1981, 1985), one of the supporters of L2 acquisition unrelated to individual differences, claimed that "adults have two distinctive ways of developing competences in second languages […] acquisition, that is by using language for real communication [...] learning […] knowing about language" (Krashen & Terrell, 1983:26). Solely the information acquired unconsciously is that which is subsequently used in real life conditions, such that no significant role is attributable to differences in cognitive resources of L2 learners. In contrast, Robinson (1997, 2001a, 2003, 2005) argues that L2 development is to be distinguished from L1 development in the sense that individual differences play an important role in the development of learners' L2 repertoire. "Individual differences in the rate and level of ultimate attainment achieved during L2 development are clearly more apparent than they are in L1 development" (Robinson 2003: 52). According to Robinson's Cognition Hypothesis (2001a:56) "individual differences (IDs) in cognitive abilities (e.g., working memory) […] will increasingly affect task-based performance and learning as tasks increase in complexity". That means that (1) individual differences are related to L2 development and (2) the degree of their impact will depend on the complexity of the task. Therefore, individual differences in cognitive abilities are linked to task complexity and should be theoretically and pedagogically investigated in terms of L2 production and more importantly L2 development. A few studies have already given us some insight into the effects some of those individual differences may have on L2 production (Niwa, 2000) or L2 development (Baralt, 2010). Although the scale of these studies was not similar (an MA thesis - Niwa; a PhD thesis - Baralt), both studies looked at the effects WMC may have on L2 performance and L2 development, and in both cases it was found that working memory capacity significantly affects L2 production and development and is directly linked to the degree of complexity a particular task requires. The interface between cognitive abilities and task requirements in L2 production and development needs to be further explored, where each dimension, resource-directing and resource-dispersing, will possibly correspond to a particular cognitive ability of the learner (e.g. attention-switching ability will be more adequate to a +/- single task dimension; WMC will probably correspond to +/- reasoning demands or +/- few elements)[7].

2.7 Summary of the chapter

The aim of this chapter was first, to introduce the construct of task complexity by giving its definition and providing its distinction from task difficulty and second, to present two opposing theories on task complexity: Skehan's Trade-Off Hypothesis versus Robinson's Cognition Hypothesis. Finally, a justification in the choice of Robinson's Cognition Hypothesis was presented. In the next chapter, the construct of +/- spatial reasoning demands (a resource-directing variable), which was used for the present study is defined theoretically and empirically with examples from previous studies in cognitive psychology and applied linguistics.

[7] The issue of working memory capacity will be dealt with in Chapter V.

CHAPTER III

DEFINING THE CONSTRUCT OF SPATIAL REASONING DEMANDS

3.1 Introduction

In the previous chapter the Cognition Hypothesis, which represents one of the main pillars of the present dissertation, was introduced and its associated triadic componential framework was thoroughly explained. Chapter III is fully devoted to the definition and exploration of one of the resource-directing variables selected for manipulation in the experiment of the entire study: the amount of spatial reasoning demands (+/-spatial reasoning).

Before operationalizing the selected dimension of +/- spatial reasoning in the present study, the decision was made to clearly define the level of cognitive complexity of tasks on the basis of previous theoretical work, empirical studies on the issue and also a series of our own empirical studies – three pilots of the present study (the latter ones will be described in Chapter VI). In a number of previous studies the manipulation of cognitive task complexity, though theoretically based, was not always empirically confirmed *a priori*, but justified by post-hoc measures of task complexity (Gilabert, 2005; Gilabert, Barón & Llanes, 2010; Michel, Kuiken & Vedder, 2007; Robinson, 2001a). Yet, this was considered necessary in the studies on task complexity to ensure the right operationalization of the dimension in the experiment (Norris & Ortega, 2009). Therefore, some variables, such as the one of spatial reasoning demands selected for the present study, requires a thorough definition at the theoretical level and also most importantly at the empirical level. In what follows, some theoretical approaches to spatial reasoning demands will be given, together with a series of empirical studies in psychology and in applied linguistics relevant to the present research. Finally, the issue of measuring task complexity will be advanced and discussed. This information was crucial for the operationalization of +/- reasoning demands in the experimental design and later for the interpretation of the obtained results.

3.2 Theoretical approaches to spatial reasoning demands

Spatial reasoning as related to language has a long history of research, from cognitive perspective (Talmy, 1983) or from Gestalt psychology perspective (Levinson, 2003), and analysis in the area of L1 acquisition as well as L2 acquisition. A common question is: why do we succeed in transmitting some kind of information related to space in one situation but then fail in other situations? The explanation may be found in the way we structure the space surrounding us by means of language, which is exactly what the work of Talmy (1983) is about. His publication titled "How language structures space" (1983) is a crucial reference for researchers in this area. According to Talmy (1983), the conceptualization of space is made through two subsystems: the first one comprises static ("region" and "location") and dynamic ("path" and "placement") concepts, which represent a point or a line in any type of space, and the second consists of an "object" or a "mass", which can occupy any volume. When talking about one object in relation to another, people normally attribute a role to each of them, one becoming a "primary object" and the other - a "secondary object". Talmy (1983) provided an example of "the bike near the house". Here it is clearly seen that "the bike" is primary and "the house" is secondary. We could obviously say "the house near the bike ", but in this case it would sound odd, since we are used to conceiving a house as an object with a more or less permanent location, and it is therefore normally considered a point of reference, or "Reference Object", the term used by Talmy (1983). Our linguistic representations of space are related to some geometrical notions, so when talking about static relationships, prepositions that are linguistically responsible for transmitting the idea of a location or a region could be progressively organized according to their "partiteness" (e.g. near – a single point, between – two points; among – a few points, amongst – many points). Relationship between the objects in motion are also represented geometrically where, for instance, prepositions could refer to different geometrical forms (examples are taken from Talmy, 1983:192).

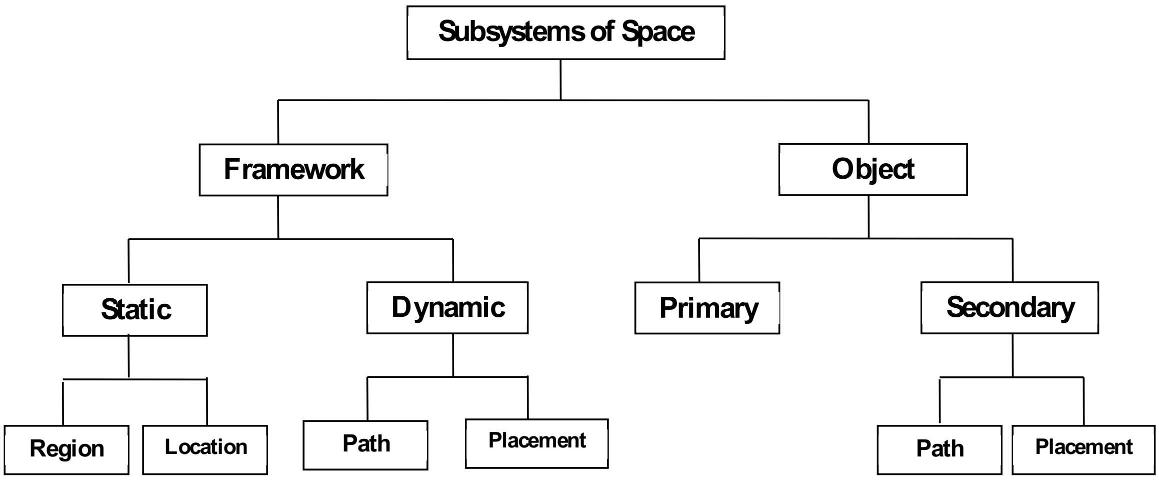
"The bike sped across **the field** – a plane surface".

"The bike sped through **the tunnel** – a cylindrical object".

"The bike sped into **the sports hall** – an enclosed volume".

In this example, the preposition changes according to the form of the reference object. Sometimes, there is more than one reference object, which involves the distinction among several reference objects: primary, secondary, tertiary and so on and so forth. Moreover, Talmy made a distinction within the secondary reference objects between "encompassive" and "external" ones. "Encompassive secondary objects" are those that encompass the primary object. When saying, "John is ahead of Mary", the reference is made not only to the position of John, but also to his direction. The opposite type of the secondary reference object is an external one, which already contains the required information without enclosing the primary reference object. In Figure 4 there is a graph which summarizes the structure described above.

Figure 4. *Talmy's space structure*



After the object and the framework are figured out, the schematization of their representation is needed, which could include three processes: idealization, abstraction and selection (Talmy, 1983; Herskovits, 1986, 1997).

1) **Idealization** refers to our simplification of an object from real forms to simple points and lines. Thus, in the phrase "the apple on the table", table will be visualized as only the top of the table and not the whole object. Importantly, when it comes to visualization of objects and figures in motion, there is evidence that it is done by idealizing them to a point. Some of the exceptions to this pattern given by Herskovits (1997) are idiomatic expressions (e.g. roll over) or some cases where the prepositions include all or most parts of an object.

2) **Abstraction** is an essential factor in language representation, since human beings have to abstract the whole scenery from distinguishable characteristics when they put it into words. When saying "Philip is waiting for a taxi", we abstract

away from many details that are present, such as (1) the position of Philip, (2) the position of a taxi; (2) Philip's location's properties – type (road or square etc.; width and so on).

3) **Selection** involves the use of a part of the object to represent the whole object. In the same example given above "the glass on the table", when representing the table, only one part is mentally selected - its surface, not the whole.

Figures, objects and people in motion are fundamental for linguistic representations of space. As mentioned previously, when a dynamic scene has to be verbalized, despite having a three-dimensional nature, in words it is reduced to a two-dimensional picture with objects being points and their trajectory being linear. In other words, the linguistic representation of navigation paths takes place on the basis of cognitive maps. In this case shape or magnitude of the object will be considered irrelevant, as the selection of an adequate preposition will be made on the basis of the relationship between the two objects (primary and secondary).

"The ant crawled across my palm".

"The man walked across the field".

"The bus drove across the country".

Even if "ant", "man" and "bus" have different shapes and "palm", "field" and "country" different magnitude, they all execute the same trajectory, which is transmitted by "across". As for the line, it will be structured by axes (Marr, 1982), so the shape and the position of the line will be seen through its axes. In Table 3 below taken from Herskovits (1997), there is a brief summary of the schematization of space.

Table 3. *Schematization or object geometry selection functions (Herskovits, 1997: 181)*

1. Idealizations to a
    - point
    - line
    - surface
    - plane
    - ribbon
2. Gestalt processes:
    - linear grouping (yields a two- or three-dimensional linear object)
    - two- and three-dimensional grouping (yields an area, or volume)
    - completed enclosure
    - normalized shape
3. Selections of axes and directions:
    - model axis
    - principal reference axis
    - associated frame of reference
    - direction of motion
    - direction of texture
    - direction of maximum slope of surface
4. Projections:
    - projection on layout plane
    - projection on plane of view
5. Part selections (triggered by the high salience of the part):
    - three-dimensional part
    - oriented free top surface
    - base

There are other alternative representations of space, such as Landau and Jackendoff's (1993), which argues that all spatial characteristics (static and dynamic) could be fit in a unique system for language representation. They suggest a difference between the "whats" and the "wheres", where the language encodes these two subsystems separately, establishing the relationship between them with the help of a preposition. However, Herskovits's (1997) main criticisms related to the schematization done by Landau & Jackendoff (1993) are about the fact that they seem to ignore the shape of an object, which, in a way, could influence preposition selection and also about the selection issue in the representation of the object, which cannot always be reduced to one of its parts in a verbal visualization).

3.2.1 Verbalization of space across languages

In a more recent study, Levinson (2003) discusses the account of coordinate systems (frames of reference) and then raises the issue of cross-linguistic diversity in the representation of frames of reference across languages. The notion of "frames of reference" comes from Gestalt psychology, and in reality it refers to a coordinate system within which we place and thereafter identify the objects in terms of location or movement. As it will be seen later on in Table 4, some distinctions of the concept of frames of reference taken from Levinson (2003:26) are presented. In his book Levinson (2003) gives a brief review of all the distinctions which appear in different areas of science. However, below an overview of some of the most relevant distinctions is provided, as sometimes these categories overlap.

According to Levinson (2003) the absolute versus relative distinction is related to whether the frame of reference is invariant or subject to modification depending on the situation and the speaker himself. Thus, using fixed notions such as "west", "east" will be considered an absolute frame of reference, contrary to "on the right of, on the left of " which may vary in their position and function depending on the context.

The egocentric versus allocentric distinction is relevant in neuroscience, where it is claimed that depending on the situation humans rely more on an egocentric or an allocentric frame of reference. It is also relevant for conceptual development, where it is known that children during their first months of life analyse the world from the egocentric point of view, but later on they acquire the ability to see it from an allocentric perspective.

Table 4. *Spatial frames of reference: some distinctions in the literature*

"relative" vs. "absolute"

(philosophy, brain sciences, linguistics)

(a) space as relations between objects vs. abstract void
(b) egocentric vs. allocentric
(c) directions: relations between objects vs. fixed bearings

"egocentric" vs. "allocentric"

(developmental and behavioural psychology, brain sciences)

(a) body-centred vs. environment-centred
(b) subjective (subject-centred) vs. objective

"viewer-centred" vs. "object-centred" or

"2.5 D sketch" vs. "3D models"

(vision theory, imagery debate in psychology)

"orientation-bound" vs. "orientation-free"

(visual perception, imagery debate in psychology)

"deictic" vs. "intrinsic"

(linguistic)

(a) speaker-centric vs. non-speaker-centric
(b) centred on speaker or addressee vs. thing
(c) ternary vs. binary spatial relations

"viewer-centred" vs. "object-centred" vs. "environment-centred"

(psycholinguistics)

(a) gaze-tour vs. body-tour perspectives
(b) survey perspective vs. route perspective

The deictic versus intrinsic frame of reference is used mostly in linguistics and refers to the spatial interpretation in context (centred on the speaker) or independent from it. Finally, Levinson (2003), acknowledging that the frames of reference in known European languages are still confused, distinguished among three basic frames of reference:

(1) intrinsic (object-centred) – the coordinate system is defined with respect to the object to be located (e.g. "The apple is on the table").

(2) relative (viewer-centred or deictic) – the object position is considered from the viewpoint of a person - egocentric perspective (e.g. "The pet shop is at the back of the store").

(3) absolute (environmental-centred) – the position of the object is linked to the properties of the surroundings where it is located (e.g. "The town is located south of London") .

Researchers may be satisfied with different models of structuring "frames of reference". However, it is not as simple as it may seem when it comes to a real language context, where depending on the language a series of problems emerge. As Levinson (2003) argues, frames of reference are not equally distributed even among European languages. As far as other languages are concerned a deep gap still exists in space representation research, but there is clear evidence that all over the world space relationships differ largely. It is also argued that even within the languages from the same linguistic family there is not an absolute coincidence in the use of prepositions of space. Some of the space conceptual linkage may, for example, just be absent from the repertoire of a language.

3.2.2 Summary of the approaches to spatial reasoning

Research into space and its relation to language established a system of space, where a distinction between framework and object was made (Talmy, 1983).

Object included "Primary Object" and "Secondary Object", the latter having a path or placement. Furthermore, the conceptualization of the object went along with three processes: idealization, abstraction and selection, where the role of size and magnitude is minimalized, though not completely ignored. Finally, many different frames of reference were introduced, from which Levinson made a selection to include intrinsic, relative and absolute frames of reference. It is also suggested that the relationships of objects and the selection of frames of reference may vary from one language to another one.

3.3 Empirical studies into spatial reasoning in psychology

In the last twenty years, psychologists have been interested in seeing whether people possess a unitary spatial model[8] (Rock, Wheeler & Tudor, 1989) or, on the contrary, there are several spatial systems that are used according to a series of factors, such as stimulus, addressee, environment, and previous experiences with orientation among others. In this context, many studies have been carried out to test some of the theories and to establish some of the patterns of human mental functioning in terms of space. Furthermore, several studies have researched and analyzed the perspective-taking issue in the last few years. These studies are especially interesting for the present dissertation, since they test which conditions (s) make(s) a spatial task more complex to complete (normally observed through increased reaction

[8]Rock, Wheeler and Tudor (1989) carried out a series of experiments in order to show that people are unable to imagine an object from different viewpoints, if these viewpoints are unknown to them. As an example, students of their experiment had to mentally represent and draw or recognize from several options a wire metal object from a different viewpoint (90^{o} left or right). In all the cases participants had difficulties in completing the task, and during the feedback interview they commented that they could not mentally rotate the object in order to see it in a different place. Rather they tried to find an alternative, more analytical strategy, such as rotating wire metal object segment by segment.

time). Finally, many empirical studies have dealt with the issue of selection of frames of reference. Hereafter, five empirical studies will be presented organized into three themes: unitary spatial models versus multi-spatial models, perspective-taking tasks and frames of reference.

In the studies regarding the unitary spatial model versus multi-spatial model several lines of research have been observed for establishing relations between objects or people (1 & 2) and for people's or objects' transformations (3):

(1) solving problems by using one-model systems versus multi-model systems with one or several countable solutions and without conclusion in a closed task. For instance, organizing space on the table with a countable number of objects and countable number of possible positions would be a multi-model system with several countable solutions (Byrnes & Johnson-Laird, 1989);

(2) solving problems by using one-model systems versus multi-model systems. Do people construct a single preferred model first and then choose among alternative models when it comes to multi-model system tasks? For example, it is arranging a table in a preferred established previously way even if there are several more beneficial alternatives for organizing it (Rauh, Hagen, Knauff, Kuss, Schlieder & Strube, 2005).

(3) completing tasks which require people's or objects' transformation – which system will be used: a unitary transformation system or a multi-transformation system?

Byrne and Johnson-Laird (1989) carried out two experiments to test a theory of spatial interference according to which it is easier to solve problems requiring only one mental model than those which require two or more mental models. The objective of the first experiment was two-fold: to see whether people are capable of making two-dimensional inferences; additionally, the authors aimed to draw a comparison between theory and their own results. For that purpose, they used three kinds of interference: one-model problems with a valid conclusion; multi-model problems with a valid conclusion; multi-model problems without a valid conclusion (see Figure 5).

Figure 5. *Example of a two-model problem without a valid conclusion*

B is on the right of A	C A B	A C B
C is on the left of B	D E	E D
D is in front of C		
E is in front of A		

What is the relation between **D** and **E**?

According to the model theory, the easiest problem would be the one-model problem with a valid conclusion, and the most difficult would be the multi-model problem with no valid conclusion. Subjects were given five, six or seven instances of each of the three types of problems, where each problem contained different lexical instances related to table settings (plates, cups, etc.). The experimenter read a description of an arranged table once and the participants of the experiments were then asked to describe the relation between some of the objects on the table. Participants were informed that in some cases they did not have enough information to establish a relation between the objects. The results showed that the participants could cope with both one- and two- dimensional interferences equally well, however, they were more accurate with one-model problems than with more than one model. Furthermore, problems with a valid conclusion turned out to be easier than the ones without a valid conclusion. The second experiment also examined three models: one-model problems, multiple-model problems with a valid conclusion and multiple-model problems without a valid conclusion. Here the participants were given four orientations and 18 problems. As predicted, one-model problems were easier than multi-model problems and a problem with a valid conclusion was easier than the one without a valid conclusion. So, only a valid versus invalid answer was statistically reliable, not the number of inferential steps. The overall results of both experiments showed that a valid conclusion is statistically important in solving spatial reasoning problems[9].

Rauh et al. (2005) conducted a series of experiments to see whether a person first constructs a single preferred model for a particular task on spatial reasoning and then tries to base other tasks on that preferred mental model by approximating other models to that one[10]. It generally states that human reasoning comprises three stages (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991): (1) *model construction* – mental representation of received spatial information by building up an array (matrix); (2) *model inspection* – verification of some of the easy relations already established between

---

9 These findings were directly related to the manipulation of task complexity, where, on their basis, a more complex task was considered the one with more than one way of placing furniture (the detailed description of the tasks is given in Chapter VI).

10 The study is partially based on the research conducted by Byrne and Johnson-Laird (1989) where they tested one-model problems and two-model problems with relation to spatial-reasoning tasks and they discovered that tasks, which allow for two spatial layouts were more difficult to complete than tasks which determine only one spatial layout. An example of a two-model problem (Byrne & Johnson-Laird, 1989, p. 567):

B is on the right of A
C is on the left of B

D is in front of C
E is in front of B

What is the relation between D and E?

C	**A**
B	
D	
E	

A	**C**
B	
D	
E	

objects on the basis of the inferences; (3) *model variation* – a definition of more complex relations that have not been established yet in the previous stage. Although a body of research substantially supports this statement, little is still known about what happens within each of the stages. The four experiments carried out by Rauh et al. (2005) aimed to fill this gap. In the first experiment, researchers wanted to explore whether people come up with a single mental model of spatial reasoning or whether they are open to creating several mental models at the same time. For this purpose, 24 students from the University of Freiburg were recruited for the experiment. After a definition and learning phase, they were offered 144 inferences to solve, from which 72 were multiple models. The chi-square test conducted for the series with multiple models showed a significant result that supported the hypothesis regarding the existence of a preferred mental model. Once the evidence for the existence of preferred mental models was provided, the researchers continued with the analysis of mental processes in the cases where uniquely alternative mental models are suitable to find a correct solution. In the second experiment, 26 participants from the University of Freiburg took part. They were asked to verify 210 inferences (156 relevant problems, 39 fillers and 15 practice trials). Results showed a significant difference between the solution corresponding to preferred mental models and those with alternative valid solutions. Additionally, verification latencies also showed that students triggered more time to check a problem, which contained only alternative mental models. 24 students without special preparation in logics were engaged in the third experiment. They were offered 74 three-series problems with 3-, 5-, 9-, and 13-model solutions. Their task consisted of working at their pace to analyse a problem which subsequently disappeared from the screen, and the students were asked to offer as many mental models as possible for its solution. The results showed that indeed, students first started with their preferred mental model and then based on that model they searched for alternative mental models in order to solve a task.

Zachs and Tversky (2005) carried out three experiments [11] to find out whether people have a unitary spatial transformation system or it may vary based on the type of transformation to be performed: object-based transformation ("transformation involving the motion of external objects", 2005:272) or perspective transformation ("transformations involving the motion of external objects", 2005:272). Zachs and Tversky (2005) predicted that three factors could influence the choice of a system:

(1) It will depend on the type of spatial judgment required in the particular context; a person's movement into space is more likely to be considered as a perspective transformation and a body's movement is more easily to be seen as an object-based transformation.

(2) A distinction is made between bodies and objects' representations, where bodies could be equally depicted with the use of both systems (object-based transformation and perspective transformation) whereas objects are more likely to be referred to by means of an object-based transformation.

(3) Since, people have control over both systems, they could be instructed to give preference to one system over the other.

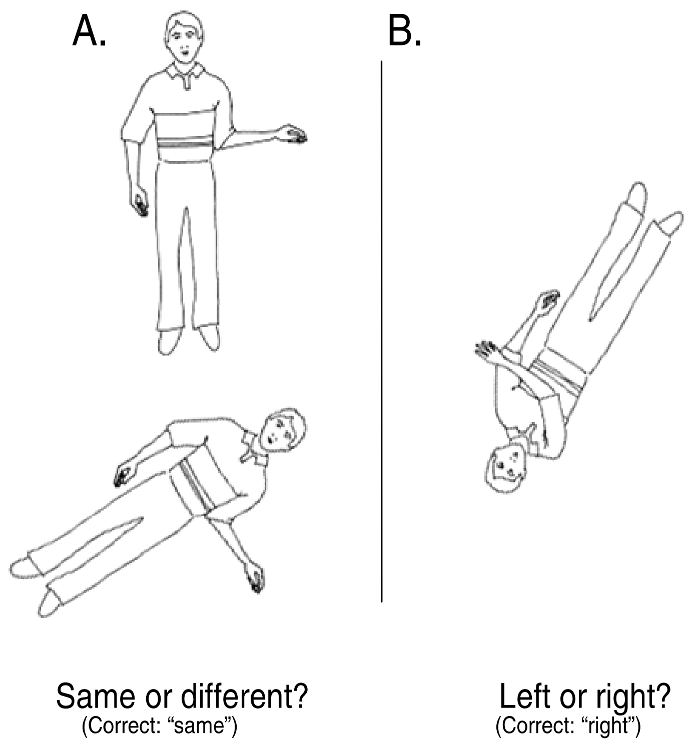
Three experiments were carried out subsequently to determine the existence of two transformation systems, where each one is applicable naturally according to a series of conditions. 42 students were involved in the first experiment. They were given six blocks of trials. Each trial was about identifying (1) "same or different"

[11]As the third experiment was a replication of a part of the second experiment, for the sake of space it will not be reported here. For more information, see Zacks and Tversky (2005).

images of two bodies – one in a normal position and the other one in a rotated position and (2) "left or right" outstretched arm of a rotated body (see Figure 6).

The two first blocks were carried out with no explicit instruction, whereas in the second two blocks participants were instructed to use the perspective condition and in the last two blocks participants were instructed to use the object-based condition.

Figure 6.*Tasks and stimuli for judgments about bodies in Experiment 1-3 (Zacks* & *Tversky, 2005:275)*



As predicted, the results showed that giving instructions with the focus on perspective significantly slowed down the reaction time in the same-different task, and providing instructions with the object-based focus has a slowing down effect on the left-right task.[12] That finding corresponded to the hypothesized natural transformation and its effect on time reaction in case of discrepancy between the instructions and the natural transformation. Furthermore, the experiment showed that both instructions and the type of required judgment affect the choice of spatial transformation operation.

The objective of the second experiment was to find out whether the stimulus itself affects the choice of the transformation system and processing times of information before giving a correct response. For this, along with bodies, small, operable objects were introduced. Additionally, as before, the tasks were manipulated with left-right judgments and same-different judgments. 40 Washington University students were recruited to take part in the experiment. They were tested individually and were given several combinations of task and stimulus with 96 trials for each of them. The trials within each set were always randomized and the elements of the trials such as rotation amount, direction of rotation and others, counterbalanced. The results showed strong correlations between orientation and task and between orientation and stimulus and also a three-way correlation between task, orientation and stimulus. Moreover, in the case of bodies, correlations were only positive in same-different judgments; however, in the case of objects correlations were highly positive for both same-different and left-right judgments.

To sum up the findings regarding mental models of spatial identification, a series of results are highlighted.

1. It is easier to deal with one mental model with a conclusion than with two or more mental models at once with a conclusion; moreover, independently from the

[12] This finding was used for the operationalization of cognitive task complexity, which is thoroughly explained in Chapter VI).

number of mental models to deal with it is much more difficult to identify a problem without conclusion whenever it has one or more mental models.

2. Human beings tend to choose their preferred model for a specific task, on which they base the subsequent task solutions; therefore, a task is more complex if a different model is required.

3. Finally, it was shown that people base their transformation not on a unitary spatial transformation system, but on different transformation systems (object-based transformation, perspective transformation), the choice of which could depend on a series of factors such as stimulus, instructions or type of spatial judgment required for a particular task. If a selected spatial transformation system does not correspond to its natural human selection, the task becomes more complex to complete. This issue is particularly important for the operationalization of +/- spatial reasoning in the task used in the experiment, because the increasing cognitive complexity will be based on this.

Two more studies (Mainwaring, Tversky, Ohgishi & Schiano, 2003; Nori, Iachini & Giuseberti, 2004) regarding perspective-taking and frame of reference issues are briefly presented. Mainwaring et al. (2003) carried out a series of experiments to establish which factor(s) affect(s) the choice of perspective (people, objects or environment). Additionally, they looked at the role that social background played in perspective selection. In the first experiment the participants, 71 undergraduates at Stanford University and 70 undergraduates at Kanazawa University (Japan) organized into pairs were asked to give information as precisely and as concisely as possible to each other, regarding a position of a target item (out of two identical ones). Participants were informed that they had several ways to complete the task and they were free to decide which was the most appropriate one in each case.

Each task contained a small description with a brief key in some cases to understand the diagram and in other cases with the compass indicating direction to the top of the diagram. Finally, two types of scenarios were employed: 20 co-present (participants were part of a situation) and 8 remote (participants controlled the situation from a distance). Each description was analyzed in terms of the perspective of the speakers, the addressees, a landmark (objects or people in the area to make reference to), or a compass (e.g. east, west, north) and some combinations of these. The results showed that when the scenario specified the addressee's location, this was used most of the time (76% US students, 70% Japanese students). Compass and landmark were used less frequently, and speaker information was almost ignored.

In the second experiment 24 US students from Stanford University and 54 Japanese students from Kanazawa University were involved. The conditions of the task were exactly the same as in the first experiment, with the only difference being that here the participants were asked to write a note for themselves and for another person. Similar to the first experiment, participants tended to use the proximity criterion in choosing perspective, so, for instance they used landmark perspective when they had an object as a point of reference close to them and if not, they use the addressee's perspective if he was situated close to the target object. On the contrary, they generally preferred landmark or compass perspectives and where those proved difficult to use, they referred to their own perspective. Sometimes they also used an addressee's perspective, but in this case it was likely to be considered as an additional landmark. Finally, as in Experiment 1, no palpable difference was detected between the US and Japanese students, where even landmark perspective was used similarly.

In the third experiment, participants were asked to obtain the information about a location of a target object from an imaginary partner with a simple question requiring a yes or no answer. 64 Stanford students and 54 Kanazawa students participated in the experiment. They were offered a total of 28 scenarios and the results were analyzed following the same model as in Experiment 1. The results of the third experiment were conditioned by the two-fold task which participants had to fulfill: make the message clear for the addressee and keep the message clear for themselves. For that reason, the use of neutral perspectives (landmark and / or compass) was dominant. Alternatively, participants favored the use of addressee's perspective over their own perspective.

Nori et al. (2004) focused their study on analyzing the selection of egocentric versus allocentric frames of reference depending on the localization (subject-to-object versus object-to-object). They hypothesized that if localization was subject-to-object the description would tend to be egocentric and, on the contrary, when localization was identified as object-to-object the selected frame would be allocentric. 16 students from the University of Bologna took part in the experiment on a volunteer basis. They were asked to first remember the position of eight real objects in a room and then to retrieve the position of each of the objects taking into account the object's localization (subject-to-object or object-to-object). Additionally, participants' position toward an object was modified in terms of the angle (from 0^{o} to 315^{o}, eight different positions in total). The results showed that participants adopted an egocentric frame of reference to all the situations as being the one that required less mental effort and used it throughout all the layouts independently of the object's localization.

To summarize the findings with respect to perspective-taking and frame of reference issues, the studies described above came to the following conclusions:

(1) People tend to use the addressee's perspective more when the addressee is the person to whom they have to give the instructions; on the contrary, they used

more landmark or compass orientation in their explanation when they were self-directed.

(2) As far as the choice of egocentric or allocentric frames of references, students consistently adopt the egocentric frame of reference as it requires less mental effort to complete the task than the allocentric frame of reference. In Table 5, a summary of the results across the five studies is given. These findings represented some of the guidelines for elaborating the operationalization of the tasks used in the experiment.

study	Materials	Results
t, whether pable of mensional	Three types of problems: one-model problems with a valid conclusion; multi-model problems with a valid conclusion; multi-model problems without a valid conclusion	**One-model problems** were easier than **multi-model problems** and a problem with a valid conclusion was easier than the one without a valid conclusion.
her people t their model other s to that	Similar instances as in Byrnes & Johnson-Laird were used in Rauh et al.'s experiment with half one-model problems and the other half two-model problems. Before the main problem-solving part, they had a learning phase, where they theoretically could construct their preferred model.	The results showed that students first started with their **preferred mental model** and then based on that model they searched for **alternative mental models** in order to solve a task.
her people y spatial stem or it n the type n to be	Three experiments were carried out subsequently to determine the existence of two transformation systems – object-based transformation and perspective transformation. As materials, transformation of bodies first and bodies' versus objects' transformation second were performed. Students were instructed to perform transformations first naturally and then either following object-based transformation or perspective transformation.	Participants based their transformation on **object-based transformation** or **perspective transformation systems**, the choice of which depended on stimulus, instructions or type of spatial judgment required for a particular task. If a selected spatial transformation system did not correspond to its **natural human selection**, the task became more complex to complete.

n factor(s) e of e, objects	Three experiments reported in the present study where participants had to (E1&E2) report a place of an object in the room (either directly to the addressee or by writing a note both for themselves and for the addressee) and (E3) to ask a yes-or-no question regarding the place of a target object.	Participants used the **addressee's perspective** more; in contrast, they used more **landmark or compass orientation** in their explanation when they were self-directed.
ection of of	Participants had to first remember the positions of the objects in the room and second retrieve all the positions.	Students consistently adopted the **egocentric frame of reference** instead of the **allocentric frame of reference**, as it required **less mental effort to complete the task**.

3.4 Empirical studies into spatial reasoning in applied linguistics

In recent years, there has been a growing interest in how adults acquire spatial relations in L2 with their already established system of spatial relations in L1, as it had been shown that in childhood during the pre-language period we already acquire some knowledge about space and then in the process of learning a language adjust our knowledge of space to words which adults use around us. (Bowerman & Choi, 2003; McDonough, Choi & Mandler, 2003)[13]. Therefore, when speaking about spatial relations and expressions of space in L2 acquisition, the L1 should be taken into account. For that reason, below a series of studies on how people acquire spatial relations in their L2 in relation to their L1 are presented.

A large project conducted by Becker and Carroll (1997) was aimed entirely at researching how people express spatial relations, how they do so at any given stage of the acquisition process and also what factors are determinant in this process of acquiring spatial relations in the L2. The design of the project was both cross-sectional and longitudinal. In the cross-sectional research their interest was focused on the combination of a source language and a target language, where both could be either two close languages (from the same family of languages eg. Spanish and French) or two unrelated languages (e.g. Spanish and Swedish) and additionally the selected languages needed to be widely spoken cases in Western Europe, such as Spanish, French, English or Italian. In the end, they came up with 10 pairs. The main goal of the longitudinal study was to identify the factors that motivate L2

[13]In the experiment children from 9 to 14-month old exposed both to English and Korean and adults speaking either English or Korean were involved. Two spatial contrasts were tested: tight vs. loose containment (wordily distinguished only in Korean) and tight containment vs. loose support (wordily distinguished both in English and Korean). Non-verbal, preferential-looking tasks were used with both children and adults. Children could easily distinguish both relations, whereas English speakers had problems in distinguishing the tight containment vs. loose containment, which is not presented in English.

development and the factors leading to fossilization in the language. To do so, 40 participants (4 participants per pair of languages) were periodically recorded over a period of 2 years. The tasks designed for the project were aimed at retrieving spatial expressions of location and motion. The results showed that at early stages the primary role plays the acquisition of the basic variety, which is "characterised by the increasing use of explicit relaters – prepositions /particles, and verbs" (1997:187). It was observed that dynamic relations (e.g. *go up, down, ahead, away, out*) had been acquired before static relations (e.g. *be left, right*) following the order of acquisition of dynamic relations in the L1. Both topological and axis-based relations were concurrently acquired. However, in the case of static relations, topological links were acquired before axis-based and, if L1 and L2 overlap, the order of acquisition was similar to that of the source language, but if this was not the case more confusion and variation took place. As an example, regarding the acquisition of axis-based relations, Italian speakers could easily distinguish and at the same time acquire the three types (vertical axis, lateral axis and sagittal axis) in English, whereas they needed more time and more effort to acquire the same relations in German.

That study showed how L2 acquisition of spatial relations is a long complex process which depends on many factors and is even more difficult if spatial systems of the source language and the target language are not entirely related. One possible way of acquiring spatial relations in the L2 could be by means of tasks, which are thought to approximate real life conditions as closely as possible, which gives an opportunity to a non-native speaker of a given language to understand how a spatial concept is expressed in the target language in its daily use.

Up until now the task, widely used in TBLT, which has been mostly analysed in terms of its effects of +/- spatial reasoning on L2 production, is the map task. In what follows, some of the studies are briefly described to give an idea of how they are designed and subsequently fitted into TBLT research.

Robinson (2001a) carried out a study in which 44 Japanese students from Aoyama Gakuin University were asked to perform two map tasks (cognitively simple versus cognitively complex). The objective was to analyse whether task complexity (simple versus complex), task sequencing (simple-complex versus complex-simple) and role (information-giver versus information-receiver) played a significant role in L2 production. The simple version of the task was a map of the Ayoama Gakuin University area the students were familiar with, which was easy to interpret for a speaker. The complex version of the task was a map of a larger area in Tokyo less likely to be known by students which contained many landmarks that were more difficult to distinguish. The results for task complexity and task production displayed significant differences in terms of lexical complexity, fluency and confirmation checks. As far as sequencing and L2 production are concerned a significant difference was found, where complex to simple sequence positively affects accuracy of both tasks (but not for interaction).

In other, more recent studies, Gilabert (2007) and Gilabert, Barón & Llanes (2009) also employed map tasks, along with two narrative and two decision-making tasks. In both studies the same tasks were used: in the simple version of a task, the map contained a few easily distinguishable landmarks and the students were asked to give the instructions for how to get to a newsstand, to the post office and to a flower shop, where they were supposed to use only the lateral axis (e.g. straight, left, right). In the complex version of a task, the map included many very similar landmarks and directions along lateral, vertical (e.g. up, down) and sagittal axes (e.g. back, front). In the first study, Gilabert (2007a) looked at the number of repairs used under both conditions and hypothesized that a more complex task triggered fewer errors and a greater amount of repairs. That hypothesis was largely supported by the findings. The second study (Gilabert et al., 2009) looked at the amount of interaction that each of the conditions (simple versus complex) could potentially generate. The results showed that a complex version of an instruction-giving task triggered a higher number of clarification requests and comprehension checks.

3.4.1 Summary of the findings

Since spatial representations are not equal across languages and even sometimes differ among people that share the same language, its acquisition becomes a hard task to complete. As shown in Becker and Carroll's study (1997), successful acquisition of spatial forms in the L2 depends on several linguistic (the role of the L1) and cognitive factors (sometime replicating the order of acquisition in the L1). As far as research in TBLT is concerned, this dimension of spatial reasoning is largely under-researched. Even the brief overview of some studies with the use of map tasks proved reasoning demands to be an important cognitive dimension which promotes the use of static and dynamic relations and L2 development in terms of acquisition of spatial expressions. However, up until now this dimension was only researched with regard to L2 production. Therefore, there is a gap in the studies concerning the effects of cognitive task complexity along +/- spatial reasoning which the present work aims to fill.

In previous sections of this chapter, an issue about the features that make a task less complex or more complex has already been mentioned with regard to psychology as well as applied linguistics. Importantly, in TBLT, when it comes to the use of the construct of task complexity (i.e. spatial reasoning, as a resource-directing variable) some objective measures of its complexity are required. In the last two decades, a series of instruments have been selected which are described in the following section.

3.5 Measuring cognitive complexity

After having devoted a large section of the chapter to what the construct of reasoning demands is in psychological and in linguistic terms, it is necessary to give an overview of how cognitive complexity is measured and how it may be operationalized in task-based language learning studies (TBLT) and, even more importantly, to discuss what the controversies and grey areas are regarding this issue. A number of scholars have suggested that it is necessary to theoretically and, more importantly, empirically justify the adequacy of the operationalization of tasks along their cognitive task complexity. Initially, some of the researchers just relied on their intuition when designing a series of tasks with different levels of cognitive load. In the previous decade, the tendency was to use at least one psychological measure to objectively justify task operationalization. Recently a call was made by Norris and Ortega (2003) for the independent measurement of cognitive task complexity, which has led to the use of more sophisticated techniques in the analysis of task complexity. Below, a brief overview of the measures and their uses is given.

3.5.1 Standard measure – affective variable questionnaire

The affective variable questionnaire is the most frequently used measure of task complexity, due to its easy and fast administration during data collection. It consists of a simple questionnaire with items in a Likert-scale (e.g. from 1 to 6, 1 to 10 or 1 to 15), the questions of which may vary depending on the ultimate goals of the research. Normally, it includes statements about task difficulty, participants' motivation, and participants' anxiety among other variables. An example of a pair of statements is:

I found the task easy 1 2 3 4 5 6 I found the task difficult.

Subjective ratings have been seen to be an effective measure, as several studies in psychology have demonstrated that people can clearly rate a task as easy or difficult which can be used to truly discriminate among simple and complex tasks. Therefore, this technique was adopted to capture differences in task complexity (Gilabert, 2005, 2007b; Gilabert et al, 2009; Ishikawa, 2008; Michel et al. 2007; Robinson 2001a). More recently, Révész, Michel and Gilabert (2012) have shown this measure to be likely the most cost-effective one in comparison to many others.

3.5.2 Updated measures of task complexity

Following the call by Norris and Ortega (2003, 2009) for more objective methods of measuring task complexity independently, researchers have started to investigate alternative ways of measuring task complexity. One of the measures quickly adopted by L2 researchers is time estimation. To this date and to my knowledge, the first researcher in applied linguistics to use this technique was Baralt (2010). Although this technique was quite common in psychological research, little was known about it as applied to linguistics. The main idea behind this technique is that while part of our attention goes to completing the task, some part of our attention goes to keeping track of time. When cognitive load increases, typically people devote less attention to keeping track of time. Therefore, when people are asked to guess how much time they needed to complete a task, they are likely to be more precise in guessing time after doing a cognitively simple task than a cognitively complex task. A number of studies in psychology have found that information processing *per se* led to increases in time estimation and also these increases depended on the amount of information to be processed (Hicks, Miller & Kinsbourne, 1976; Ornstein, 1969; Thomas & Weaver, 1975; Underwood & Swain, 1973). This was explained by Block, Hancock and Zakay (2010) in their overview of many studies related to time estimation, where they claimed that people lose track of time when their attention is fixed on task completion. Some studies dealt with language aspects in L1 processing, however before Baralt (2010) nobody had used time estimation in the context of L2 development and acquisition, so it was hard to say whether the results would replicate those in psychological studies. Indeed, time estimation was successfully integrated into her PhD project and the results were consistent with the previous findings in the area of time estimation. After that, several studies started incorporating this technique along with the affective variable questionnaire (Malicka & Levkina, 2012; Révész, Michel & Gilabert, 2012). Some other techniques that have been suggested for measuring task complexity, such as eye-tracking, brain scanning and dual task methodology are very briefly presented below.

Eye-tracking is the process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head. It has been widely used in applied linguistics: lexical access (Traxler & Pickering, 1996) and language processing (Kaushanskaya & Marian, 2007; Tanenhaus, Magnuson, Dahan & Chambers, 2000; Tanenhaus & Spivey-Knowlton, 1996) among other areas.

Brain-imaging techniques mainly use functional neuroimaging to measure an aspect of brain function, often with a view to understanding the relationship between activity in certain brain areas and specific mental functions. In applied linguistics this technique is used to analyse bilinguals' brain processing (Frenck-Mestre, Anton, Roth, Vaid & Viallet, 2005; Kovelman, Baler & Petitto, 2008; Rodriguez-Fornells, van der Lugt, Rotte, Britti, Heinze & Munte, 2005) or language processing in individuals with some mental diseases, such as dyslexia (Dalby, Elbro & Stødkilde-Jørgensen, 1998).

Dual-task methodology is a procedure in experimental psychology that requires an individual to perform two tasks simultaneously, in order to compare performance with single-task conditions. This technique is used in several fields of science from therapy (Huang & Mercer, 2001) and cognitive psychology (Della Sala, Baddeley, Papagno & Spinnler, 1995; Olive, 2004) to linguistics (Brünken, Plass & Leutner, 2004; Emerson & Miyake, 2003).

Although potentially good and reliable measures of task complexity, the first two methods require very expensive machinery; therefore they could not be easily administered in typical classroom conditions. The dual-task methodology could be run with the use of computers, which already makes its implementation easier, however as a previous design and validation of the task itself are needed, it is not accessible for every researcher[14] either. Because of their practical administration during data collection in classrooms, time estimation and subjective ratings were selected for the present dissertation.

3.6 Summary of the chapter

Research into the construct of spatial reasoning demands and their impact on task complexity in the field of TBLT is not very common and few studies have specifically dealt with it as such. The present chapter has attempted to build up a general picture of this construct by answering some of the following questions: (1) how spatial reasoning demands are theoretically represented in psychology; (2) what some of the findings are regarding mental representations of space, mental transformations of space and selection of perspective; (3) and finally, how the issue of spatial relations and spatial reasoning is treated in applied linguistics and in TBLT in particular. The first section introduced the concept of space, where a distinction between framework and object was established. Furthermore, different frames of reference were presented, such as intrinsic, relative, and absolute frames of reference (Levinson, 2003). Regarding the psychological research into spatial reasoning five studies were selected to see what the evolution of the findings related to mental models of spatial identification was and also to analyze different ways of perspective and frame of reference selection. In the third section, the focus was on the studies in applied linguistics which had already dealt with spatial reasoning and spatial relations.

---

[14]In order to learn more about the independent measurement of task complexity, see Révész, Michel and Gilabert (2012, 2013), where they analysed three different ways of measuring task complexity, affective variable questionnaire, time estimation task and dual-task methodology. The interesting finding they obtained is that out of these three measures the most systematically reliable method is the affective variable questionnaire.

Orders of acquisition for static versus dynamic relations were presented and also a selection of studies in TBLT with the use of map tasks was described. Finally, different independent measures of task complexity were presented as being a necessary tool for independently justifying the operationalization of cognitive complexity of the tasks. Apart from a standard, widely used measure of subjective ratings (i.e. affective variable questionnaire), a brief list of newly proposed measurements was offered (i.e. time estimation, eye-tracking, dual-task methodology, brain-imaging techniques), from which time estimation was selected for the present study.

CHAPTER IV

TASK SEQUENCING

4.1 Introduction

In the previous chapters the focus was on tasks in general as tools for language production, assessment and learning, and their cognitive complexity which represents a good basis for distinguishing simple tasks from complex ones for research purposes as well as educational purposes. In recent years, several research and educational projects have come up with widespread use for tasks in order to encourage students to use and acquire the L2 (e.g. R. Ellis, 2003; García Mayo, 2007; Shehadeh & Coombe, 2012; Skehan, 1998; Van den Branden, 2006). However, despite a growing interest in syllabi based on pedagogic tasks (as opposed to linguistic units) in recent years, there is no consensus on which criteria to use to put tasks in a sequence so that the learning process is optimized. Thus, the need arose to find an optimal and logical task organization to foster learning. More than twenty years ago, Long and Crookes called for a need to create "valid, user-friendly sequencing criteria" (1992:46) which would be easily implemented in the classroom. While some scholars have suggested that difficulty of input or content should be the main criteria for sequencing, other scholars have already designed models of task sequencing that organize pedagogic tasks from cognitively simple to cognitively complex (Robinson, 2005, 2007b, 2010). Although cognitive complexity (see Chapter II for a detailed explanation of the construct) has been a well-established criterion for task sequencing in other fields of science, such as mathematics, mechanics, engineering, informatics, or physics among many others, in the field of applied linguistics, there is no general agreement to date as to what criteria should be used by designers and teachers to organize and sequence pedagogic tasks. The very limited number of studies on task sequencing in the field of applied linguistics, as well as the recent testable SSARC model of task sequencing by Robinson (2010) suitable for testing, became the main inspiration for the present research. As in previous chapters, the objective of the present chapter is to introduce the concept of task sequencing from different angles. For this reason, in what follows first an insight into the criteria of task sequencing in other fields of science (such as mathematics or computer-mediated education) is provided. Next, the SSARC model of task sequencing created by Robinson (2010) is described, which is to date the most complete proposal as to how tasks should be organized within a syllabus. Finally, a general overview of some of the scarce empirical studies on task sequencing in applied linguistics is provided.

4.2 Task sequencing outside applied linguistics

As noted before, the concept of task sequencing is not new in other fields of science (such as mathematics, physics, logistics and others) and it has been widely used in the literature to define task organization in many different senses. The present section does not aim to offer a detailed overview of the most important studies on task sequencing in all known science fields, rather the main objective is to show different facets of the construct of "task sequencing" in two different fields (mathematics and computer-based education) to facilitate the understanding of the concept as related to the field of applied linguistics.

In mathematics one basic principle was elaborated by Garey (1973). In an imaginary project a finite number of tasks was presented, where all the tasks had to be completed successfully, otherwise the project would fail. The organization of the tasks was based on the following principles – cost of task completion and probability of task completion. In that sense, a task with the least effort and the highest probability of completion would go first. This abstract representation of how task sequencing is operationalized could be illustrated by the following example: very frequently during the right candidate selection for a particular job there is a multi-level process of candidate testing. Each level consists of several tests, which are eliminatory. These tests are normally selected and organized according to their cognitive complexity. One of the main objectives of task sequencing mentioned by Garey (1973) is elaboration of task sequencing with minimal task cost and maximal task reward. Although Garey (1973) could only create a theoretical efficient optimization algorithm, the two main points he highlights are essential in the understanding of the core of task sequencing:

(1) task sequencing should be based on the cost and the probability of task completion;

(2) task sequencing should be done with minimal task cost and maximal task reward.

Much closer to syllabus design in our field is the research by Brusilovsky and Vassileva (2003) in the field of web-based education. In fact, it is directly related to intelligent tutoring systems in small-scale projects as well as in large-scale projects. They describe two models used in computer-based courses: the dynamic course generation system (DCG) and the concept-based course maintenance system (CoCoA). The two systems are especially designed to organize personalized educational computer-based courses and they do so by manipulating and sequencing tasks. DCG "looks like a traditional structured course. However, this course is generated individually for every student to achieve a certain learning goal (a concept or topic that has to be learned). The generation takes into account the already existing knowledge of the student and can accommodate differences in the individual's way and pace of acquiring the material" (Brussilovsky & Vassileva, 2003:78). CoCoA is a course maintenance system. "The core of the system is a course-sequencing engine that works in an "inverted way" to analyze the quality of sequencing in a static human-authored course […] the key to the sequencing power of CoCoA is a structured domain model and a refined approach to indexing the course material" (Brussilovsky & Vassileva, 2003:83). The main difference between the two systems is that DCG uses single-concept indexing, whereas CoCoA uses multi-concept indexing; therefore the first system is applied for simply organized courses and is not suitable for small groups whereas the second system is more powerful, but requires more sophisticated human resources. The systems were used in teaching such courses as training mechanical skills – typewriting, the theory of jazz or the functioning of a mechanical toaster. Interestingly, the criteria for task sequencing here is based on the following points:

(1) the previous knowledge of the students in relation to a given topic, on the basis of which the whole course is automatically organized;
(2) the progress of the students throughout the web-based course, where the sequences can eventually be restructured, if the results shown by a student do

not correspond to the knowledge required to go on with the next step and tasks.

In those two examples it can be seen that task sequencing was based on task complexity, hence starting with low complexity of tasks. Secondly, task complexity was based on the individual perception of learners and, where necessary and if possible, the sequencing of tasks should be adapted progressively to the rhythm and the capacity of the students. When switching to the area of language learning based on tasks, some similar proposals have come from several scholars (Long, 1985; Nunan, 2004; Prahbu, 1987). However, only one comprehensive model of task sequencing in the context of task-based language teaching has been advanced by Robinson (2010).

4.3 Task sequencing in TBLT

4.3.1 Early proposals on task sequencing

Before Robinson forwarded the first ever comprehensive model of task sequencing, which suggests that tasks should be graded according to their cognitive complexity, there had been previous proposals concerned with the issue of sequencing tasks in a syllabus. A brief overview of some of the crucial moments in the evolution of this issue is provided in this section.

The first early attempt at dealing with sequencing was presented and explicitly mentioned within the task-based language teaching project known as the Bangalore –Madras project (Prahbu, 1987). It was a large-scale project carried out in different schools of Bangalore, the main goal of which was to teach English as a foreign language through the use of communicative tasks. As this was a new approach, not systematically used, it was decided to start every new explanation with the use of a simple task, so as to make it easier for the students unfamiliar with the new approach to understand the functioning of task-based teaching. After the first experience with using tasks, the basic criteria for task sequencing were elaborated: "tasks within a given sequence (i.e. tasks of the same type forming the basis of several lessons) were ordered by a commonsense judgment of increasing complexity, the later tasks being either inclusive of the earlier ones or involving a larger amount of information, or an extension of the kind of reasoning done earlier" (Prahbu, 1987:40). While the principle of sequencing tasks from simple to complex had appeared, no clear scheme of how to actually identify a task as more or less complex was given, as teachers were guided by their intuition and they took the intuitive decision of how to sequence tasks rather than using objective sequencing criteria.

Some years later, another scholar, Nunan (2004), devoted a whole chapter to the grading and sequencing of tasks within syllabus design in his book on task-based learning. Apart from a general principle of task sequencing from simple to complex, he distinguished between a number of factors according to which tasks could be graded. The first factor was input, where a task was considered more complex when it was less explicit (e.g. a text with no visual support, such as photographs, highlighted text, tables and so on) or when it did not contain contextual support. Not having textual support made the task more complex to process and to carry out. The second factor was the learner who would perceive a task as more or less difficult depending on factors such as his or her background knowledge. Also following Brindley (1987), he suggested a list of affective and cognitive variables that could influence on the grading of task complexity (i.e. motivation, confidence, prior learning experience, learning pace, observed ability in language skills, cultural knowledge or awareness, and linguistic knowledge). The final set of factors had to do with task procedure, which was also likely to affect the cognitive complexity of the task. When it comes to this factor, task input should be stable and, by contrast, task procedure should be modified in order to gradually increase the cognitive complexity of the task.

Long (1985, 2005) and Long and Crookes (1992) raised the problem of a lack of empirical studies on task sequencing. They noticed that task sequencing processes mostly "appear to be arbitrary processes, left partly to real-time impressionistic judgments by the classroom teacher" (1992:37). Moreover, they pushed for defining operational terms for task sequencing which were not clearly identified. Importantly, before grading tasks Long and Crookes (1992) suggested that tests be selected according to needs analysis and to communicative criteria rather than linguistic functions.

Finally, Skehan (1996,1998) proposed a model of sequencing tasks in a logical way from simple to complex on the basis of the attentional resources that each task would require. Drawing on previous work by Candlin (1987) and Nunan (2004), he distinguished three sets of features to consider during task sequencing: code complexity, cognitive complexity, and communicative stress. Code complexity was considered the formal factor related to lexical and syntactic difficulty, whereas cognitive complexity and communicative stress were seen as individual characteristics affecting task performance. Cognitive complexity was related to content concerning two areas: processing and familiarity. Finally, communicative stress was referred to as related factors of task difficulty such as time pressure or task modality. Sequencing tasks with the proper selection of task features, according to Skehan, would lead to "an affective balance between fluency and accuracy" and "the opportunity for previous restructuring to be applied" (1996:53).

To sum up, previous proposals related to task sequencing have all agreed on the idea that task sequencing should be done by organizing tasks from simple to complex. Also, it has been agreed that tasks should be graded following some kind of notion of cognitive complexity and not only linguistic complexity as traditional syllabi have done. Up until the 21st century, however, no clear model of task grading according to cognitive complexity has been elaborated. In this context, Robinson (2010) made an attempt to fill this gap by creating the SSARC model, which represents the first model for task sequencing susceptible to testing.

4.3.2 Robinson's SSARC model of task sequencing

In order to understand the SSARC model, some assumptions which were thoroughly described in Chapter II are briefly mentioned at the beginning of this section. As seen previously, the Cognition Hypothesis, which constitutes the basis for the SSARC model of task sequencing, makes predictions regarding two aspects of second language acquisition: (1) it predicts the effects task complexity may have on L2 performance, which is typically measured in terms of fluency, linguistic complexity, and accuracy; and (2) it predicts how task complexity may impact L2 development through tasks graded according to their cognitive demands.

Robinson makes an important distinction between two dimensions of task complexity: resource-directing variables, which direct learners' attention to particular linguistic features needed to meet the demands of the task at hand, and resource-dispersing variables, which deplete learners' attention by dispersing resources over various aspects of task performance. An example of a resource-directing dimension is reasoning demands, where tasks with no reasoning demands but simple transmission of information may typically require simple linguistic devices (e.g. *and, but*), as opposed to tasks with some reasoning demands where at least subordination is needed to express cause-consequence relations (e.g. *because, therefore*). Similarly, in the case of map tasks, dealing with a well-organized map with all the landmarks in place is much easier and does not require as much of a mental effort as dealing with a confusing, fuzzy map with no clear landmarks and multiple perspectives and interpretations (Becker & Carroll, 1997). In contrast, a clear example of a resource-depleting variable is access to planning time before or during task performance. Giving no planning time before task performance simply disperses attentional resources during performance.

Associated with the Cognition Hypothesis there is the triadic componential framework (Robinson, 2001a, 2001b; Robinson & Gilabert, 2007), along with task conditions, and task difficulty. In the first version of the triadic componential framework (Robinson, 2001a), within the task complexity section Robinson distinguished between three resource-directing variables (i.e. +/- few elements; +/- here-and-now; +/- no reasoning demands) and three resource-depleting variables (i.e. +/- planning time; +/- single task; +/- prior knowledge). All these variables stood as a guide for decisions on task design and as recommendations for sequencing criteria.

Later, Robinson and Gilabert (2007) updated the existing version of the triadic componential framework by adding some new elements to it in order to create an operational taxonomy adequate for pedagogic design and use in classroom settings. Here again, Robinson reconfirmed his predictions regarding L2 production and L2 development as affected by task complexity. As claimed in previous descriptions of the model (Robinson, 2001a, 2001b, 2003, 2005) increased task complexity along resource-directing dimensions will promote L2 development of newly learned target items. In contrast, manipulating task complexity along resource-dispersing variables will contribute to further automatization of previously learned language features. In this way, both manipulating along resource-directing and resource-dispersing variables is crucial for L2 development.

In the SSARC model, Robinson (2010) suggests that the cognitive demands of tasks (i.e. task complexity features) should be the sole basis for sequencing decisions, and that tasks should first increase in complexity along resource-directing variables and only later along resource-dispersing variables. Robinson predicts that increasing cognitive complexity along resource-directing variables will cause more attention to and retention of input during task performance and therefore contribute to L2 development.

When grading tasks according to their cognitive demands, Robinson claims that it is important to pursue the natural order for sequencing, for example from tasks that require spatial reasoning first based on simple topological relations to tasks that include axis-based relations (Taylor & Tversky, 1996), which is the way children acquire these relationships (Becker & Carroll, 1997), or else from tasks with no casual relations to tasks which establish these relations (Bergman & Slobin, 1994). Robinson claims that L2 adult learners with their full cognitive apparatus in place have retained the notion of cognitive complexity and so organizing tasks from cognitively simple to cognitively complex will help adult L2 learners in at least three ways: 1) to more efficiently allocate their attentional and memory resources and; 2) to move from the use of simple linguistic devices to the use of complex ones and so help them through the process of development; 3) to play closer attention to input and achieve higher retention of that input.

Robinson proposed two instructional-design principles for task sequencing following the SSARC model, which are:

Principle 1. Only the cognitive demands of tasks contributing to their intrinsic conceptual and cognitive processing complexity are sequenced (Robinson, 2010:247). That means that tasks should be organized and sequenced not following the criteria of their linguistic complexity, but rather according to their cognitive complexity, whereby tasks should progress the way from cognitively simple to cognitively complex along such variables as +/- planning time, +/-single task, +/- number of elements, +/- reasoning demands.

Principle 2. Increase resource-dispersing dimensions of complexity first, and then increase resource-directing dimensions (Robinson, 2010:247).

The second principle explains the entire SSARC model. Sequencing should start with simple tasks along all dimensions, both resource-directing and resource-dispersing (e.g. +/- planning time, +/- few elements): simple and stable, "SS". Next, increase in complexity is along resource-depleting dimensions (e.g. - planning time; - single task), which promotes automatization of the already existing repertoire, "A". Afterwards, there is an increase in complexity both along resource-dispersing and resource-directing dimensions (e.g. – planning time; - causal reasoning), which on the one hand restructures an already existing repertoire, "R", and on the other hand introduces new forms and concepts of the target language, leading to overall maximum complexity, "C" (Figure 3 on p. 28 depicts schematically the SSARC model). These two principles along with a few steps (SS + A + RC) represent the heart of Robinson's SSARC model. As Robinson's model (2010) is a new proposal in

TBLT, the model has almost no empirical research behind it. In the next section, the first few studies testing the model are explained.

4.4 Empirical studies on task sequencing

To date, very few empirical studies have been carried out with regard to task sequencing. In the 4th Biennial Task-Based Language Teaching Conference in Auckland (New Zealand) a colloquium on task sequencing was organized where several scholars from different countries presented their findings in this new area. As a result of that meeting, a book on task sequencing is now in preparation (Baralt, Gilabert, & Robinson, in press). Some authors have kindly agreed to share the findings of their studies included in the book in this part of the dissertation.

All the selected studies deal with task sequencing in classroom contexts and most of them use intentional reasoning demands or spatial reasoning demands as a resource-directing variable along which task complexity is manipulated and then tasks are sequenced. All the studies follow Robinson's SSARC model of task sequencing (2010) in the design of their experiment.

Baralt (in press) conducted a study with the participation of 94 American university students in which the main goal was to analyze the impact task sequencing may produce on L2 development of the Spanish past subjunctive in two different learning contexts: traditional face-to-face classroom versus online classroom. As input students were provided with 12 small cards with L1 story blurbs (based on 6 comic strip pictures), which they had to explain in Spanish. Stories in the L1 used in the treatment were retold in L2. Two levels of cognitive complexity were established: simple (intentional reasoning was already provided) and complex (students were asked to explain the cause of some actions carried out by the stories' protagonists). Four different sequences were elaborated following Robinson's SSARC model (2010) and Robinson's proposition (2011) to compare complex sequences, simple sequences and U-shaped sequences: (1) simple – simple – complex (SSC); complex – complex – simple (CCS); complex – simple – complex (CSC); simple – complex – simple (SCS). As assessment, two tasks were created: oral and written story retell performed in pairs. Results showed that students who did sequences with more complex tasks than simple tasks, CCS and CSC, produced significantly more past subjective forms, than those with more simple tasks in the treatment sessions (SSC and SCS). This significant difference was only detected in the traditional, face-to-face group, as the results of the online classroom groups did not show any significant difference between any of the sequence treatment. The results showed that rather than task sequencing, the number of complex tasks during treatment was the decisive factor in the development of the use of Spanish past subjunctive by students.

Lambert and Robinson (in press) explored the effects task sequencing may produce on L2 development of learners' abilities to summarize short stories with the use of intentional reasoning. 120 Japanese university students organized into two groups were enrolled in the experiment, carried out during one academic semester. All the students were informed about the goal of the study – to improve their ability to retell short stories while justifying their opinion, explaining somebody's actions or intentions, and drawing general conclusions. In the design of tasks for sequencing three levels of complexity were established according to one resource-directing variable (intentional reasoning) and four resource-dispersing variables (planning time, prior knowledge, number of steps and multi-tasking). Treatment was organized into six sessions, where each level was performed in two sessions. One group followed the target treatment; the other was a control group with no specific treatment. Pretest and posttest retells were analyzed for syntactic complexity, intentional reasoning and grammatical accuracy. The results showed that in the posttest there were differences in short storytelling in favor of the target group for all measurements, however these differences were not statistically significant.

Finally, Levkina and Gilabert[15] (in press) have analyzed the effects of task sequencing on L2 development in the use of spatial expressions. 50 participants were involved in the study. They all were university students enrolled in the summer intensive course of English in the School of Modern Languages in Barcelona. They were randomly organized into three groups according to different types of sequencing: (1) from simple to complex, (2) from complex to simple and (3) randomized. Two assessment tests were employed: a two-part vocabulary test and a descriptive task. Before treatment students were given input materials to become familiar with target expressions of space. Treatment was organized in one session and contained three oral tasks. Task complexity was manipulated along spatial reasoning demands. The design of the experiment included a pretest, an immediate posttest, and a delayed posttest. The statistical analysis displayed significant results for the group with the complex to simple sequence, which in the immediate posttest outperformed the other two groups in both the two-part vocabulary test and the descriptive task. However, the results for the delayed posttest showed the significantly higher retention of the target items by the group in the sequence from simple to complex.

Several conclusions may be drawn from these first studies on task sequencing within the framework of Robinson's SSARC model of task sequencing

[15] This study corresponds to the second pilot described in Section 6.2 of the present dissertation.

(2010). Task sequencing from simple to complex has beneficial effects on L2 development in a short period of time (Lambert & Robinson, in press) and over a long period of time (Levkina & Gilabert, in press). On the other hand, as Robinson suggested in his article (2011), complex sequences and / or U-shaped sequences may make a difference in L2 development. Baralt's study (in press) and Levkina & Gilabert's study (in press) showed that these alternative sequences may also lead to significant improvements in the use of target structures immediately and over a longer period of time.

4.5 Summary of the chapter

In this chapter the key concept of the dissertation, task sequencing, was fully introduced. As in previous sections, task sequencing was first presented outside the boundaries of applied linguistics to draw a picture of how this concept is treated in other sciences such as mathematics or computer engineering. In that section it was pointed out that the main principle was sequencing tasks from simple to complex, where a simple task would contain less steps to accomplish the goal, while the complex task would require more steps; additionally no more complex tasks are offered if simpler tasks are not fully mastered. In applied linguistics task sequencing is still a relatively new concept that needs a considerable number of studies to be carried out in order to explore the optimal organization of tasks in the process of L2 learning. It was highlighted that over the last two decades several researchers stressed the need for a model or clear scheme of task sequencing. However their proposals were limited to offering a logical sequence from simple to complex without providing any objective, empirically verifiable model of task sequencing. In 2010 Robinson published an article where he offered a complete model of task sequencing – the SSARC model which is explained in Section 4.2.2. Finally, a few very recent studies based on this model concluded the chapter. According to recent findings, task sequencing from simple to complex is not the only optimal way of organizing tasks, though it yields a series of benefits in terms of immediate retention of information and for L2 development over time. Nevertheless, there are other ways of task sequencing (e.g. complex sequence, U-shaped sequence), which may also be potentially beneficial for L2 development of L2 specific target forms. As the other key component of the present study is working memory capacity and its influence on L2 development of spatial expressions by means of task sequencing, the next chapter will depict the construct of WMC both theoretically and empirically.

CHAPTER V

WORKING MEMORY CAPACITY

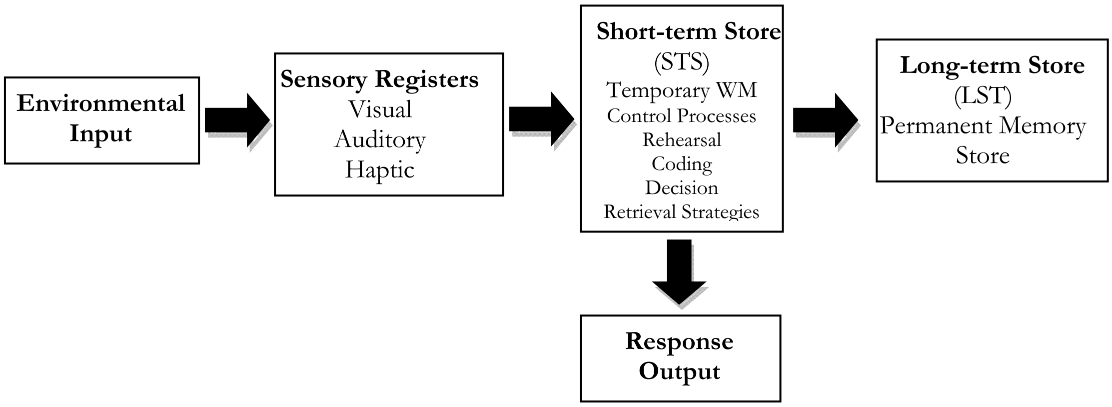
5.1 Introduction: from primary memory to working memory

So far we have seen the construct of task complexity, the construct of +/- reasoning demands and task sequencing in light of Robinson's perspective of the SSARC model. In this chapter, we turn to examining working memory capacity which is believed to mediate both L2 production and L2 development by means of the use of tasks. This is relevant because working memory is measured in this study and its role in conjunction with task sequencing is explored.

Although the term of "working memory" was first introduced into cognitive psychology in the 60s, the concept of a form of memory that serves to temporarily store information for ongoing mental processing is not new and has its own history which has evolved over at least the last hundred years. In 1890, James divided the notion of memory into two types: primary memory and secondary memory, where primary memory represented a reservoir of information from which it was always possible to retrieve immediately accessible information, whereas secondary memory, in order to obtain the required information, needed the activation of cognitive processes. After a fifty-year break, Miller (1956) returned to the issue of capacity of storage by analyzing some empirical studies and he came to the conclusion that short-term memory capacity was limited to "the magical number seven, plus or minus two" when measured by the span of absolute judgment or the span of immediate memory. Furthermore, Miller theorized on the item functions suggesting their flexible characteristics as they could be grouped and turned into "chunks", so that the initial number of seven selected items could be subsequently reduced by that technique. The idea of a "magical number" has been widely studied and discussed[16]. Later on two more characteristics of short-term memory were added: (1) its short duration without rehearsing, as tested by Brown-Peterson task (Brown, 1958; Peterson & Peterson, 1959) and (2) its immediate accessibility, as borne out in a set of studies by Sternberg (1966, 1969). All these assumptions have reaffirmed the initial distinction of two modes of storing information (short-term memory and long-term memory) which were next explored by Atkinson and Shiffrin (1968, 1971). They established a sequential way of processing and remembering information that, in their view, passed from short-term memory to be stored in long-term memory. Although currently the model is not as influential as it used to be, it is still very much cited in the cognitive literature. This decline in interest is basically due to the sequence principles of information processing, which was shown not to be linear, especially in studies with brain-damaged patients. Briefly, the main idea of that model consisted of the information flow that passed through visual and auditory registers and resulted in short-term storage where, after rehearsing, coding, and retrieving, that information was permanently stored in long-term memory (see Figure 7).

[16]In his recent work, Cowan (2010) claimed that when participants are prevented from using chunks, the maximum number of retained information bits in the short-term storage is reduced to 3 (+/- 2).

Figure 7. *The Atkinson-Shiffrin model of memory*



The shift in depicting short-term storage as a single system was also due to the experiment by Baddeley and Hitch (1974) where they administered a multi-task test on short-term storage to neurologically healthy people and saw that those people could handle it well. The test consisted of remembering a string of six to eight digits in each trial and retrieving them immediately after each true-false task. The results showed that people were able to both store information for a short period of time and process other bits of information. Balance between processing and storage initiated a new wave in "working memory" studies.

5.2 Models of working memory capacity

There are several models of working memory capacity which are theoretically discussed and empirically used and operationalized within cognitive psychology. Below three of these models are presented, beginning with Baddeley's Multi-Component Model, which has become one of the most widely used in our field.

5.2.1 Baddeley's model of WMC

WMC as a model was first forwarded by Baddeley and Hitch in 1974. It consisted of three main components: the central executive, as a control system, with other subordinate systems, the phonological loop (for processing auditory input) and the visuospatial sketchpad (for processing visual and spatial input). In 2000, Baddeley added a fourth component to the model – the third slave system, which is the episodic buffer. Below each of the components is presented.

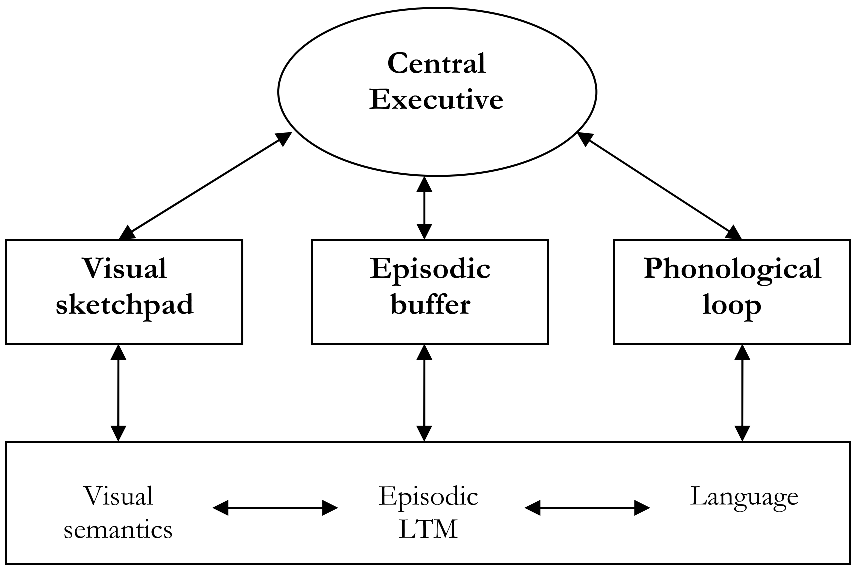
The central executive is responsible for the attentional control of working memory (Baddeley, 1986). The idea of control of the attentional system over other slave systems was developed from Norman and Shallice's model (1986), where one main component supervises the other components. The main functions of the central executive consist of processing input from visual and verbal sources, coordinating the three slave systems, and controlling shifts of attention between different kinds of information. It is important to note that the central executive is attentionally limited. In recent years, research conducted by Miyake, Friedman, Emerson, Witzki and Wager (2000) suggested that the central executive may have several subcomponents that execute different functions corresponding to each individual and which effectively can be partially or completely impaired by brain damage (Baddeley, Bressi, Della Sala, Logie & Spinnler, 1991).

The phonological loop also known as the "articulatory loop" is a subsystem of working memory which is responsible for the processing of verbal information. It consists of two parts: a short-term phonological store and an articulatory rehearsal component. Two types of information can enter in the phonological loop: auditory verbal information and also visually presented information which can be subsequently transformed into verbal information and be stored in a short-term phonological reservoir. Baddeley, Gathercole and Papagno (1998) showed that the phonological loop is crucial in the process of the acquisition of vocabulary by children, and it may also be important for learning lexis for L2 learners (Baddeley, 2003). Interestingly, it was found that people have more problems with remembering words that are phonologically similar, rather than with words which are semantically similar (Baddeley, 1966) speaking aloud impedes the articulatory rehearsal process and the correct functioning of the phonological loop (Baddeley, Thomson & Buchanan, 1975).

The visuospatial sketchpad is in charge of receiving visual information about shapes and colors from the environment in order to temporarily store it and manipulate it. The subsystem could also process complex information on movement into space through a complex map area or a multi-store building. As the visuospatial sketchpad is responsible for processing information on fixed characteristics and information on movements into space, Logie (1995) proposed an addition to the concept of the visuospatial sketchpad by dividing it into two subsystems: the visual cache, which would store the information about color and shape, and the inner scribe, responsible for movement processing. Several experimental studies supported the idea of dividing the visual and the spatial components of that slave system, because they are not always related and in some cases they are separable (Klauer & Zhao, 2004).

The final and most recent addition to the working memory model is the component referred to as the episodic buffer; it represents the third slave system of working memory and is controlled by the central executive. The main function of this component consists of helping interaction to take place between the phonological loop and the visuospatial sketchpad. It is also seen as "separate from long-term memory, but which forms an important stage in long-term episodic learning" (Baddeley, 2000, p. 421). In Figure 8 Baddeley's model of working memory capacity is presented in its entirety.

Figure 8. *The multi-component Working Memory revision (Baddeley, 2003:835)*



Baddeley's model provides an interpretation of mechanisms of short-term storage and processing. However, as in the case of any model, it has also received criticism, especially in the case of the phonological loop (Jones, Macken & Nicholls, 2004), where some later findings were not completely in line with the model initially made by Baddeley and Hitch (1974).

5.2.2 Alternative models of WMC

Although Baddeley's multi-component model of working memory capacity seems to encompass all the mechanisms that can be encountered during information processing and storage, several alternative ways of representing working memory exist. This is mainly due to the fact that there are some questions regarding working memory, where the answers do not coincide with those given by Baddeley. Some of these questions are: (1) is working memory a unitary system or a multi-component system; (2) assuming that working memory is a multi-component system, how should it be divided?; (3) is working memory completely separate from long-term memory?

Although, according to Baddeley's model, working memory capacity consists of multiple components which interact amongst each other, there are other researchers who claim that working memory is a domain-free, unitary system which is closely related to general intelligence (Conway, Cowan, Bunting, Therriault & Minkoff, 2002; Conway, Kane & Engle, 2003; Engle, Kane & Tuholski, 1999). Based on a series of studies, they came to the conclusion that cognitive processes that contribute to WMC "critically tap an executive attention-control process" (Conway et al., 2003:550).

Finally, Eriksson and Kintsch (1995) argued that due to the extended use of working memory in everyday tasks such as large text comprehension or expert performance, it is not sufficient to use short-term storage alone, but the long-term storage system should also be involved. Consequently, they introduced a new term for the combination of long-term storage with processing, namely the model of "long-term working memory". In this way, the results on text processing, understanding and performance would not only depend on the immediate capacity of retaining one piece of information while processing another one, but also on differences in prior knowledge reflected by long-term storage.

5.2.3 Summary of the models of WMC

Throughout the description of different models of working memory capacity, it has been shown that researchers are not unanimous in their interpretations of the elements of working memory. Though many aspects of working memory are still unresolved, Baddeley's multi-component model is well established among researchers both in the fields of cognitive psychology and SLA, and many experimental studies are based on it. The necessity of a well-designed test to measure WMC is an undisputable necessity, so in what follows a description of different spans to measure working memory capacity is given.

5.3 Measures of WMC

Although working memory capacity is a theoretically well-defined construct, there is no absolute consensus on which test should be chosen to measure it. The main goal of any of the spans elaborated to measure working memory capacity was to combine two main processes involved in working memory: (1) processing of information and (2) retention of information. With these two conditions in mind, several spans have been created which are presented below.

5.3.1 WM span tests

While creating the span tests several issues related to the kind of information to be processed and the kind of information to be retained had to be tackled. Some spans included arithmetical operations or counting to be processed, i.e. Operation Span (OSPAN) or Counting Span (CSPAN).

OSPAN (Engle, Cantor & Carullo, 1992; Turner & Engle, 1989) is a computer-driven test organized into sets. In each set participants have to solve an arithmetic problem aloud and qualify the answer given to them as “true” or “false” (a processing component); after that they have to remember the letter that was shown on the screen and say it aloud. Immediately after that, they are offered the next arithmetic problem so as to prevent any kind of rehearsal. At the end of each set they are asked to write down the letters they had just seen in the order they appeared on the screen. An automated, computerized, mouse-driven version of OSPAN was created by Unsworth, Heitz, Schrock and Engle (2005).

The CSPAN (Case, Kurland & Goldberg (1982) - version for school-aged children; Engle, Tuholsky, Laughlin & Conway (1999) – version for adults) consists of counting different shapes of different colors in a display with multiple distractors including the same shape or the same color, so that the task requires processing. Participants are asked to count aloud and after they finish counting they pass immediately to the next item. At the end of the set, they are asked to recall the numbers of the shapes in the order they were presented.

Finally, a slightly different test that instead of math processing deals with spatial processing is the spatial span task - SPAN (Shah & Myake, 1996), which includes processing and retention of spatial information, where participants are asked to remember the orientation of a letter given to them (angles of 0°, 45°, 90°, etc.) and at the same time to decide whether the letter is normal or mirror-imaged. At the end of each set the participants have to remember the orientation of the letter in a serial order.

For the present study, the reading span test was selected to measure the working memory capacity of the participants as it is considered to be a reliable measurement as well as a widely used tool in cognitive psychology and in applied linguistics. In what follows a more through description of this measure will be given.

5.3.2 Reading span

The reading span test represents the standard test for measuring general working memory capacity. It was also one of the first tests created for this purpose. This test has also been selected and included in the present research. For all these reasons, a separate section is dedicated to the evolution and functioning of this test.

5.3.2.1 Daneman and Carpenter's reading span test

In the original version of the reading span test created by Danenam and Carpenter (1980), participants were given a series of sentences, which they had to read aloud and remember the last word of each sentence. Participants were also asked to decide whether the sentence was plausible or not, which would involve sentence processing and would impede rehearsal of words to remember. After each set, they were asked to recall all the words in the correct order. The whole test included 60 sentences with a length of between 13 and 16 words. The sentences were typed on cards and each set had from two to six sentences. The sets of two, three, four, five and six sentences were organized by increasing the number of sentences and they appeared consecutively. Participants were also warned that the number of sentences in each set would gradually increase. If a participant failed to remember the words of all the sets at a particular level, the test was stopped at that point and that level determined the WMC of that particular student. Since then, more versions of the reading span test were presented (Engle et al., 1999; Turner & Engle, 1989). Recently, an automated version of reading span test (Unsworth et al., 2005; Unsworth, Redick, Heitz, Broadway & Engle, 2009) was created, which is described below.

5.3.2.2 Automated version of reading span

The automated version of the reading span test is experimenter-independent and fully mouse-driven. It has three blocks of practicing before the main test. In the first block participants are asked to remember a series of letters, which appear on the screen in the order of their appearance. In the second part of the practice phase, participants are trained to identify sentences as plausible or implausible. At that stage the program also computes the mean time of reading each sentence, which is subsequently applied to the main experiment. In the third practice participants are trained to read sentences, answer the question after it and remember the letter that appears after the sentence identification. In the main experiment participants are shown their ratio of correct responses for sentence identification, which cannot be below 85 %. This condition is necessary to prevent people from rehearsing. The test itself consists of a series of unrelated sentences which participants are asked to read and then decide whether each sentence makes sense or not by clicking on the corresponding button which appears on the screen. At the end participants have to remember unrelated letters after each sentence (out of the following: B, F, H, J, L, M, Q, R, and X) and recall them all in the correct order in the grid offered after each set. Half of the sentences are plausible and the other half are implausible. Each sentence is 10 to 15 words in length. Sets and sentences within each set are randomized, so each participant completes the whole test. As this version of the reading span test is much more complex than the original, a different kind of grading is needed. Several scoring methods are explained in the next section.

5.3.3 WM scoring methods

Traditionally, participants were assigned a quasi-absolute score, as in the case of Daneman and Carpenter (1980), where the threshold was established at the level of participant's ability to correctly remember the recalled items. The main problem detected by Conway, Kane, Bunting, Hambrick, Wilhelm and Engle (2005) consisted of the fact that "the difficulty of a span item may vary on many dimensions, thus threatening span reliability across different tasks" (Conway et al., 2005:774). Therefore, Conway et al. (2005) distinguished four scoring methods to be taken into consideration:

(1) partial-credit unit scoring (PCU) – grading is at the level of a set, so if a whole set is recalled correctly, one point is assigned (1), if a set is, for example, half-recalled, half-point is assigned (0,5).

(2) all-or-nothing unit scoring (ANU) – if the whole set is recalled correctly, one point is assigned (1), if the set is partially recalled a zero is assigned (0).

(3) partial-credit load scoring (PCL) –grading is at the level of an item, so if, for example, a set of 3 is totally recalled, three points are assigned (3), if a set of 4 is half-recalled, two points are assigned (2).

(4) all-or-nothing load scoring (ANL) – if a set of 3 is totally recalled, three points are assigned (3), if a set of 3 is partially recalled, a zero is assigned (0).

Due to the loss of information collected with the test while using all-or-nothing scoring, Conway et al. (2005) recommended using partial-credit scoring. As for the use of unit versus load scoring, results did not demonstrate a preferred use of one scoring method over the other, so the decision should be made by the researcher.

5.4 Empirical studies on WMC and L2 development

Working memory capacity has been recently incorporated in the studies within the field of second language acquisition. Several studies have been conducted in relation to WMC and L2 production (e.g. Fortkamp, 1999, 2003; Gilabert & Muñoz, 2010; Mizera, 2006; Mota, 2003; Trebits & Kormos, 2008). However, fewer studies have been done regarding the effects of WMC on L2 development. A selection of five studies which are relevant to the goals of the present research is reviewed below.

Mackey, Philp, Egi, Fujii and Tatsumi (2002) conducted a study aimed at analyzing the effects that WMC may cause on the amount and the quality of retention of corrective feedback (recasts in interaction) during L2 interaction. 30 Japanese students took part in the three-session experiment. They were asked to complete different tasks (from picture drawing to storytelling) and they were given a series of span tasks to measure their WMC. Those tasks included three tests: a non-word repetition test to measure phonological short-term memory, a listening span test in English and a listening span test in Japanese. In the non-word repetition test participants had to repeat 16 pairs of non-words, whereas in the listening span tests participants had to listen to a sentence and identify it as plausible or implausible while at the same time remembering the initial word (in the case of Japanese) or the final word (in the case of English) in the sentence. At the end of each set they were asked to recall all the words in a serial order. For WMC a composite z-score was created. The results showed that students with higher working memory capacity were better at noticing and processing corrective feedback, although the effect was not significant.

Robinson (2005) analyzed the role of WMC in implicit and incidental learning. 54 Japanese learners of English as an L2 were involved in the study. They were given 27 vocabulary items in Samoan to memorize. Afterwards in the training session they were shown 45 Samoan sentences and in the main experiment they were asked to identify the meaning of the sentences with "yes" or "no" questions. WMC was measured by means of a reading span task where students had to read sentences in Japanese and remember the underlined words. Robinson did not find any significant effect of WMC on implicit or incidental learning of artificial grammar.

Weissheimer and Mota (2009) conducted a study to see the role of WMC on L2 development of speech production. Thirty-two undergraduate students of English took part in the experiment. To measure WMC authors used the speaking span test following Daneman (1991) and Daneman and Green (1989). Each set consisted of a number of words which appeared one by one on the screen. The participants had to remember first the words and then at the end of each set recall them in the correct order and generate some sentences (the exact number was given on the screen) with the use of the words seen in the set. Two memory scores were applied: strict, where the sentences should contain the word in the correct form and position, and lenient, where the sentence could contain a word in a different form from the one seen in the set. To measure L2 speech production a speech generation task was used. Two data collections with a two-month interval were carried out. Weissheimer and Mota (2009) discovered that working memory capacity predicted gains in fluency and complexity, however, contrary to what had been expected, no significant effect of WMC on accuracy was detected.

Goo (2012) looked at the role of corrective feedback and WMC on L2 learning. 54 learners of EFL were involved in the study. They were divided into three groups according to the type of learning: the recast group, the metalinguistic feedback group and the control group. The study used a pretest-posttest design. To measure WMC two spans were employed: a reading span task and an operation span task. The experiment consisted of two treatment sessions where the gains in the use of *that*-trace filter were controlled by two tests: a written production test and a grammaticality judgment test. The results showed that WMC is a predictor of the effects mediated by recasts but not by metalinguistic feedback which, according to the author, was explained by the fact that attention control is responsible for the noticing of recasts rather than in the noticing of metalinguistic feedback.

Révész (2012) carried out a study to establish the relationship between WMC and recasts on different L2 outcome measures. 90 learners of English took part in the experiment. They were divided into three groups (recast group, non-recast group and control group). The design of the experiment included pretest, posttest and delayed posttest. The control tasks consisted of (1) a grammaticality judgment test, (2) a written description, and (3) an oral description. During the treatment session with the recast group while performing the descriptive tasks a recast technique was applied. Three tests of working memory capacity were administered to the participants: two tests of phonological short-term memory (digit span and non-word span) and a reading span. The results showed that participants from the recast group who scored higher on the reading span task were better at completing the written tests, whereas those with higher digit and non-word spans outperformed their counterparts on the oral task. No relation between WMC and L2 development was found for the non-recast group.

In Table 6, a summary of the five studies on WMC and L2 development is given. Overall, most studies found a positive relationship between higher working memory on L2 development through recasts (Goo, 2012; Mackey et al., 2002; Révész, 2012). Higher working memory capacity seems also to positively affect fluency and complexity of L2 production (Weissheimer & Mota, 2009). However, evidence showed a marginal role of WMC in L2 development when analytical abilities are required (learning of a new language), such as analysis of grammatical rules or accuracy (Robinson, 2005; Weissheimer & Mota, 2009).

and L2 development in SLA

udy	Measures	Results
on f uring L2	WM: non-word repetition test, listening span in English and listening span in Japanese L2: different tasks from picture drawing to story telling	Students with higher WM were better at noticing and processing corrective feedback, although the differences were not statistically significant.
implicit g	WM: reading span test L2: sentence identification task (based on an unknown language (Samoan)	No significant effect of WMC on implicit or incidental learning of artificial grammar was detected.
on L2	WM: speaking span test L2: speech generation task	WMC predicted gains in fluency and complexity, however, contrary to what had been expected, no significant effect of WMC on accuracy was detected.
e n L2	WM: reading span task and operation span task L2: written production test and grammaticality judgment test	The results showed that WMC is a predictor of the effects mediated by recasts but not by metalinguistic feedback.
veen e	WM: digit span, non-word span and reading span L2: a grammatically judgment test, written description and oral description	Participants from the recast group who scored higher on the reading span task were better at completing the written tests, whereas those with higher digit and non-word spans outperformed their counterparts on the oral task.

5.5 Summary of the chapter

Although working memory has been recently introduced into the field of SLA as a factor of cognitive ability, it has already proved to be a potentially important component of L2 production and development. As seen, several theories and models of working memory capacity exist in psychology. The most commonly used theory in our field is Baddeley's multi-component model of WMC, according to which WMC contains four components (central executive – the main control system, and three slave components - phonological loop, audiovisual sketchpad and episodic buffer). However, there are some other theories depicting the construct of WMC (e.g. Conway et al. 2003; Eriksson and Kintsch, 1995; Just and Carpenter, 1992; Myiake & Shah, 1999). Whatever model of working memory capacity is adopted, two components of WMC are always present: retention of information and processing of information. In order to empirically measure working memory capacity a series of spans have been developed. They include the reading span (Daneman & Carpenter, 1980), the operation span (Engle et al., 1992; Turner & Engle, 1989), and the spatial span (Shah & Myiake, 1996) among others. As WMC is considered "the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities" (Baddeley, 2003:189), it has been lately seen as a crucial component in L2 production and development. Several studies have already been conducted that have shown the role that WMC might play in noticing and processing of corrective feedback (Mackey et al. 2002), in the development of L2 speech production (Weissheimer & Mota, 2009), in L2 incidental and implicating learning (Robinson, 2005) and in the reception of information from recasts (Goo, 2012; Révész, 2012). Considering the potential role WMC plays in L2 development, it was decided to include this construct the present study. Moreover, to date, no research has been carried out looking at the relationship of WMC and task sequencing. In that sense, this study fills a gap in the literature. In the following chapter the methodology of the present study is described, including the description of the most relevant pilots by means of which the final version of the experiment was constructed.

5.6 Purpose of the study

After having revised the literature on task complexity, task sequencing, spatial reasoning demands and WMC, it has been shown that, firstly, there is very little research on task sequencing (for an exception see Baralt, Gilabert & Robinson (Eds.), in press). Secondly, when dealing with task complexity, more specifically with the +/- spatial reasoning variables in the context of L2 research, most studies used map tasks in their experiments, whereas many studies in cognitive psychology used tasks involving the perspective-taking component, such as the organization of objects in a closed space. Finally, in the last decade WMC has been incorporated into SLA studies as playing a significant role in L2 production and L2 development; however, as task sequencing is heavily under-researched, no studies on WMC as a mediating factor in L2 development through task sequencing have been carried out. The purpose of the present research is to explore the effects of task sequencing on L2 development of spatial expressions. Motivated by the lack of previous studies and the call for the justification of the independent measurement of task complexity before undergoing a study (see Section 6.2). Several pilots have been conducted to clearly establish the operationalization of task complexity in the study. Additionally, the mediating role of WMC in L2 development of spatial expressions by means of task sequencing in laboratory contexts is also examined, as the ultimate goal of the study is to address the crucial question of whether sequencing as mediated by WMC plays any role in the development of the L2.

5.7 Research questions

Based on the theoretical and empirical studies described above, the present study aims to explore whether sequencing tasks from simple to complex on the basis of their cognitive complexity is more beneficial for L2 learning and development than alternative way of sequencing. The study therefore puts the SSARC model by Robinson (2005, 2007b, 2010) to test. Additionally, the study examines whether individual differences in working memory capacity are involved in and mediate L2 learning processes. According to the Cognition Hypothesis, sequencing tasks from simple to complex will promote learners' L2 language development in terms of its linguistic complexity and accuracy and subsequently will lead to learners' acquisition of targeted forms; L2 learning will depend on and be affected by individual differences (e.g. attention, working memory) and affective factors (e.g. motivation, interest).

On the basis of the literature reviewed in the preceding chapters, the following research questions are proposed:

1) *Does manipulating task complexity from simple to complex affect learners' reported perception of cognitive complexity?*

This question addresses the issue of whether predicted task complexity will actually be matched by learners' perception. As seen in Section 3.5, there has been a number of calls in the literature (Norris & Ortega, 2003) for the independent measurement of task complexity. Two techniques (subjective ratings and time estimation) will be used to independently assess task complexity (see Section 6.3).

Hypothesis for the independent measurements of cognitive complexity: Based on the prediction of the Cognition Hypothesis greater task complexity along +/- spatial reasoning will make learners perceive tasks as more difficult.

2) *Is there any effect of task sequencing from simple to complex on L2 development? If so, what is the effect of task sequencing on L2 development of spatial expressions?*

The second question tackles the issue of which task sequence may be more beneficial for L2 development. According to Robinson, sequencing tasks from cognitively simple to cognitively complex will result in a more efficient allocation of attentional resources and therefore in a potentially higher degree of retention of new target information (As will be seen in Section 6.2, the target information in this dissertation consists of spatial expressions). This question will be answered in both a laboratory and a classroom context. The decision to use both settings is motivated by the fact that in TBLT research, studies are often criticized for being removed from the classroom context. In this way, by using both settings, it will be possible to see whether the classroom results are consistent with the results obtained by a more carefully controlled laboratory setting.

Hypothesis for L2 development: Based on the predictions of the Cognition Hypothesis, sequencing tasks from simple to complex will be beneficial for L2.

3) *Does Working Memory Capacity moderate L2 development as affected by task sequencing on the basis of task complexity?*

This question explores the moderating role of WMC in the development of the L2 as a consequence of task sequencing. As learners are asked to perform tasks orally and with different degrees of cognitive task complexity, tasks with a higher cognitive load will require more attentional and memory resources, therefore, working memory capacity may serve as a mediator for a more efficient performance of cognitively more complex tasks and subsequent greater retention of target information.

Hypothesis for WMC: Based on the Cognition Hypothesis, learners with higher WMC will learn more spatial expressions when tasks are organized from simple to complex.

CHAPTER VI

STUDY DESIGN AND METHODOLOGY

6.1 Introduction

This chapter presents the methodology used in the current study. First, the operationalisation of task complexity and sequencing is discussed. Specifically, an explanation is provided of the selection of lexical target items used to measure L2 development, and the technical review of the evolution of decisions made on task complexity through three pilots which preceded the main study is offered. Following is the description of the participants who signed up for the main study and the instruments employed to measure L2 development and working memory capacity. Finally, a description of procedures, coding, and scoring methods is given, which ends with the summary of the statistical analyses selected for the present study.

6.1.1 Operationalisation of task complexity

As we saw before, it is of utmost importance to guarantee that during task design we indeed obtain different levels of cognitive complexity of tasks. This need springs from the fact that authentic target tasks (e.g. a job interview, furnishing a new apartment) are highly complex so as a syllabus should present pedagogic tasks in order of increasing cognitive complexity to prepare learners for their performance in real life conditions (Long, 2005; Robinson, 2001b). In the current study, a resource-directing variable (see Chapter II, Section 2.4), +/- spatial reasoning, is operationalized from simple to complex in task design based on the guidelines provided by Robinson (2001b, 2003, 2005, 2007a, 2007b, 2010) and the recommendations by Norris and Ortega (2003, 2009). According to Robinson (2003:46), there are dimensions "which can be manipulated to push learners to go beyond this [their repertoire – M.L.] to meet the demands of a task by extending an existing L2 repertoire (such as making increasing demands on the conceptual/linguistic distinctions needed to resort to spatial location, temporality, or causality)". In other words, when manipulating task complexity along resource-directing variables, learners' attention is potentially mobilized and geared towards stretching beyond the learners' current repertoire as a consequence of the design of the task. For instance, tasks that use more causal reasoning will push for the use of more causal subordination such as "as", "because", "so that" etc. to successfully complete the task.

According to Robinson (2001b, 2003) manipulating task complexity along a resource-directing variable such as spatial reasoning demands promotes learners' processing and retention of new information in the input. By engaging learners in complex spatial reasoning, they may be pushed to focus their recourses on their spatial reasoning knowledge to be able to solve a particular problem. As said in Section 3.4, some of the examples of tasks with a spatial reasoning component are the so-called map tasks, widely used in TBLT studies (Gilabert 2005, 2007a; Gilabert et al.; 2009 Robinson 2001a).

In the present study, a descriptive, decision-making task was used, which consisted of furnishing different spaces in a new flat. This kind of task is common in everyday life, which enhances its ecological validity. In the performance of such a task, when one needs help or advice on what to buy and where to put it, one has to imagine the space with its dimensions, to take into account the size of the pieces of furniture to put in it and also to think carefully of the ideal places for each piece of furniture in a new flat. Depending on different factors such as the size of the space, the number of elements involved in decision-making, the unanimity in the decision on object placement and the selection of the objects themselves, the task may become more or less complex (see Chapter III on spatial reasoning and Section 6.4 on the decisions made on task complexity manipulation throughout the three pilots).

In order to operationalize task complexity, two theoretical and empirical sources were used: (1) the cognitive literature and (2) a series of pilots previous to the main experiment[17]. After having analyzed several studies in the cognitive literature (see Chapter III) the following list of dimensions to consider when manipulating task complexity were produced: topological (simple) versus axis-based (complex) description (Becker & Carroll, 1997), egocentric (simple) versus allocentric (complex) spatial frame of reference (Nori et al., 2004), common knowledge of object properties (Becker & Carroll, 1997), relative frames versus intrinsic frames and perspective-taking frame (Levinson, 1996; Rapp & Taylor, 2004; Zacks, Mires, Tversky & Hazeltine, 2000). The empirical source used to define task complexity stemmed from the three pilot studies carried out prior to the main experiment. The instruments used in the analyses of the results obtained from the pilots were a retrospective feedback questionnaire, an affective variable questionnaire, and a time estimation task. The results, the analyses, and the decisions taken on the basis of these three instruments are reported in Section 6.4.

[17] In Section 6.2 a justification of target items based on two pilots is given and in Section 6.4 the elaboration of three tasks for treatment manipulated along +/- spatial reasoning is explained. If the reader wishes, he could pass directly to Section 6.5, which is devoted to the description of the main study.

6.1.2 Operationalisation of sequencing

Scholars in applied linguistics investigating task sequencing have agreed on the basic premise that tasks should be organized from simple to complex (Long 1985; Nunan, 2004; Robinson 2005, 2007b, 2010; Skehan 1996, 1998). However, the next step to take, which determines the scale of cognitive task complexity, has traditionally been a difficult issue to deal with (see Section 4.3.1). In the current study the decision was made to consider and test the SAARC model proposed by Robinson (2005, 2007b, 2010) according to which a sequence of tasks should be organized solely by means of cognitive complexity, which must be gradually manipulated first, through resource-dispersing variables in order to automatize an already existing repertoire and, second, through resource-directing variables in order to promote the use of new target features of language. Robinson (2005:6) suggests that "[…] sequencing cognitive demands from simple to complex along the resource-directing dimensions of L2 tasks […] would be complementary to adult learners' own initial dispositions, and also helpful in prompting them to move beyond them". Resource-directing variables include +/- here and now, +/- few elements, +/- spatial reasoning, +/- causal reasoning, +/- intentional reasoning and +/- perspective taking (Robinson & Gilabert, 2007). As suggested in Section 2.4.1.2, all of these dimensions have been extensively tested in terms of L2 production (Gilabert, 2005, 2007a, 2007b; Gilabert, Barón & Levkina, 2011; Ishikawa, 2007; Kuiken, Mos & Vedder, 2005; Kuiken & Vedder, 2007a, 2007b, 2007c, 2008; Michel, Kuiken & Vedder, 2007; Révész, 2008), whereas there is a lack of studies that have used what it is known about task complexity and its effects to create sequences of tasks.

6.2. Justification for target items

As discussed above, the cognitive dimension selected for task sequencing operationalisation is +/- spatial reasoning demands. At the initial stage of experimental design, it was predicted that the nature of the task (their final version will be explained in detailed in Section 6.5.3), putting some pieces of furniture in an unfurnished flat, would lead to an extensive use of spatial expressions. In order to check whether the expectations were fulfilled by performance, the first piloting study was carried out.

6.2.1. First pilot: eliciting lexical target items

The goal of the pilot was two-fold: (1) to elicit and to select the target lexical items used in measuring L2 development and (2) to test the operationalisation of task complexity within a task sequence. 13 participants took part in the pilot: 2 Catalan native speakers, 2 Spanish native speakers, 4 English native speakers and 5 intermediate non-native speakers of English. Such a varied pool of participants is justified by the following reasons:

- in order to test the cognitive complexity of tasks, participants had to be able to judge on the cognitive complexity of tasks independently from their level of language proficiency;
- in order to adequately select target lexical items, it was necessary to have a list of lexical expressions of space used in the given context by Spanish, Catalan, and English native speakers and Spanish and Catalan learners of English for their further comparative analysis.

Table 7. *Lexical items of space detected in the first pilot*

ITEMS	Native Speakers	Non-native speakers		
		Total	Incorrect use	Correct Use
On top of	5	1	0	1
On the wall	5	5	3	2
Above something	3	2	0	2
In the corner	5	3	2	1
In the centre	**2**	**0**	**0**	**0**
On the right (left) – hand side	5	3	2	1
On the left (right)*	5	4	0	4
Opposite (something)	5	1	0	1
In front of *	5	5	0	5
Behind (something)	**1**	**0**	**0**	**0**
Under something	**5**	**5**	**0**	**5**
Underneath	**0**	**1**	**0**	**1**
Against the wall	3	0	0	0
Along the wall	4	0	0	0
At the back	4	0	0	0
At the forehand	**2**	**0**	**0**	**0**
Next to*	5	4	0	4
Near something	4	2	1	1
Close to something	**2**	**0**	**0**	**0**
Around*	4	4	0	4
Over the side of something	**1**	**0**	**0**	**0**

In Table 7 is a list of 21 lexical items which were used by English native speakers and non-native speakers. In the table, three columns for the non-native speakers correspond to (1) the number of participants who used a particular item, (2) the number of students who used the item incorrectly; (3) the number of students who used the item correctly. The selection of the items for the target items' test was based on the following conditions:

- whether the native speakers employed a particular item in their explanations (the frequency of the use – at least 3 out of 5 native speakers used a particular expression);
- if so, whether non-native speakers had problems using a given item or did not use it at all [18] . The items which were taken out are highlighted.

Apart from the target items, four distractor items were also included in the test (marked with an asterisk) to avoid students' frustration with having to deal with many unknown words and expressions. The "known" words and expressions were those that were widely and correctly used by the non-native speakers in the pilot defined by their use by at least 4 out of 5 native speakers and correct use for 4 out of 5 non-native speakers.

An additional analysis of the target items was undertaken in terms of its similarity to or difference from Spanish / Catalan equivalents. For this purpose, two Catalan and two Spanish native speakers were asked to perform the same tasks in their L1. Table 8 contains the results of the comparative analysis for English versus Spanish / Catalan target items.

Below the justification for the selection of each item is provided.

On top of: Although this expression has a very similar structure in both Catalan and Spanish languages (i.e. a sobre de / encima de), only one occurrence of its use was detected among the non-native speakers of English, whereas five out of five native speakers used it.

[18]If the non-native speakers of English used the selected items without hesitation, these items were also excluded.

Table 8. *Selected items with their Catalan and Spanish equivalents*

Selected Items	Catalan Equivalents	Spanish Equivalents
On top of	a sobre de	encima de / sobre
On the wall	a la paret	encima de la pared / en la pared
Above something	*a sobre de*[19]	*encima de*
In the corner	en el racó / en la cantonada	en la esquina
On the right (left) – hand side	a la banda dreta de	en el lado derecho
On the left (right)*	a la dreta de	a la izquierda
Opposite (something)	oposada a	enfrente de
In front of *	davant de	delante de
Against	*contra la paret*	contra la pared
Along the wall	a la paret	en la pared
At the back	*darrera*	*atrás*
Near something	a prop de	*cerca de*
Next to*	al costat de	al lado de
Around*	al voltant de	alrededor de

On the wall / along the wall: While in English the separation of two spatial concepts "on something" meaning "hanging ON something" and "along something" meaning "placed ALONG something" are fixed by two commonly used prepositions – "on" versus "along", the pilot Spanish / Catalan speakers used a shared preposition "a" (in Catalan) and "en" (in Spanish) extensively for both situations instead of another preposition – equivalent to "along", which is "al llarg de " (in Catalan) or "a lo largo

[19] The words in italics are those not originally used by the native speakers of Spanish or Catalan in their narratives; however, they were included in the table to give an equivalent in Spanish / Catalan for some of the selected items.

de" (in Spanish). This is likely the reason the non-native English speakers did not show the use of "along" in their productions.

On top of / above: in Spanish / Catalan at the lexical level there is no distinction between the concepts of "on top of" and "above", but rather a single prepositional expression "a sobre de" / "encima de" is used instead. For this reason, and also due to an infrequent use of both prepositions by Spanish/Catalan students of English as an L2, it was decided to include these two prepositions in the list of the target items.

In the corner: almost all non-native speakers of English made a mistake by using a different preposition instead of "in", such as "at" or "on".

On the right (left) – hand side: this spatial expression stands as another element that is widely used by native speakers, but it was observed to have a limited and incomplete use among non-native speakers (incorrect use of the preposition or omission of one or several elements of this expression).

Opposite something / in front of: it was detected that Spanish/ Catalan non-native speakers of English tended to employ "in front of" much more (which corresponds to the Spanish / Catalan preposition "delante de" / "davant de") than "opposite", whereas in the samples taken from English native speakers this preposition (i.e. "opposite") is commonly used.

Against: although in Spanish / Catalan there is an exact equivalent of this preposition "contra", no instances of it were detected in non-native language speech, in contrast to native speech where three out of five participants used it in their explanation. Consequently, the decision was made to include it.

At the back: this is another spatial expression that was not detected in the non-native English speech, and neither was its equivalent detected in Spanish / Catalan speech. This may be explained by the fact that it involves taking perspective, something that speakers usually simplify or avoid altogether.

Near something: due to the syntactic differences between "near" and "cerca de / a prop de" (Spanish / Catalan), non-native speakers of English displayed problems with its correct use by adding "to" after "near" as a clear case of transfer from their mother tongue.

On the right (left), in front of, next to, around: the spatial expressions were seen as already acquired by non-native learners of English in terms of their use and functions. In this sense, they were included in the text as the additional items to avoid participants' frustration due to being presented with too many unknown target items.

After taking the described above steps (i.e. selecting target items according to their use by native speakers and adjusting the list of target items based on both their use by the non-native users of English as well as their similarity and difference from their equivalents in Spanish / Catalan), 14 items (10 targets and 4 distractors) were selected as listed in Table 9:

Table 9. *Selected target items*

Targets	Distractors
On top of	On the right (left)*
On the right (left) - hand side	Next to*
Near	In front of*
Opposite	Around*
In the corner	
On the wall	
Against	
Above	
At the back	
Along the wall	

6.2.2. Second pilot: adjusting lexical target items

The main objective of the second pilot was to see whether the selection of target items was appropriate in terms of their recognition and use. Additionally, after the pilot some design features of the task were re-analysed and modified according to the misunderstandings and the failures yielded throughout the pilot and during the analysis of the collected material.

6.2.2.1 Participants of the second pilot

48 adult students enrolled in the summer course of English as a foreign language organized by the School of Modern Languages (University of Barcelona) took part in the second pilot. The intensive one-month course of English consisted of 25 hours of instruction per week. The level of the participants was tested by means of the school's own placement test before the beginning of the course and they were all assigned to the fourth level, which corresponds to an upper-intermediate level (B2). The current pilot was integrated into the course, and so all the parts of the experiment (pretest - posttest - delayed posttest) were completed within three classroom sessions. From the original pool of the participants (48), nine of them were eliminated, as they missed one of the parts of the pilot experiment.

6.2.2.2 Procedure of the second pilot

The pilot was distributed in three days: day 1 – pretest (20-30 minutes); day 2 - treatment and immediate posttest (1 hour – 1 hour 15 minutes); day 3 – delayed posttest (30 minutes), two weeks after treatment and immediate posttest.

The whole group was split into three subgroups according to three sequencing conditions (simple – complex, complex – simple, randomized). Participants were randomly assigned to one of the three subgroups. All of them received the same input before treatment which was comprised of three tasks which included all target items selected in the first pilot. Each of the target items appeared three times throughout the input. In the input tasks participants worked in pairs and were asked to follow the instructions of the teacher first and then of their partners in each dyad and to draw the furniture items onto the floor plan of a flat. As for treatment, three sequences[20] were assigned to the students, but this time they had to work individually.

Sequence 1: three tasks organised from simple to complex;

Sequence 2: three tasks organized from complex to simple;

Sequence 3: three tasks organized randomly (more complex – simple – most complex).

Therefore, the independent variable was task sequencing. The within-group measures included pretest - posttest - delayed posttest tasks, affective variable questionnaire, and time estimation task. In what follows, the analysis of pretest – posttest - delayed posttest, which determined the adjustment of the final list of the target items, is presented.

[20] A detailed explanation of the nature and content of the task is given in Section 6.5.

6.2.2.3 Two-part vocabulary test

This test consisted of a productive part (i.e. translation) and a receptive part (i.e. multiple-choice test). The two-part vocabulary test in the pilot contained 14 items which included 10 target items and 4 distractors. The results obtained from the students were analysed according to the following criteria: the number of people who knew the target items before treatment should be no more than the half of the whole pool (19,5); the gains should be more than 15 % (at least 6 participants); the tasks should not contain any trace of ambiguity.

Table 10. *Item analysis of the vocabulary test*[21]

Items	Pretest	Posttest	Delayed Posttest
on top of	10	21	12
on the right-hand side	17	31	30
near	15	26	25
opposite	9	30	21
in the corner	16	35	29
at the back	16	35	29
on the right*	23	37	39
along the wall	17	35	37
next to*	27	30	32
on the wall	14	29	27
in front of*	36	37	37
against the wall	26	30	29
around*	34	37	39
above	7	30	29

[21]It reflects the number of items used by the participants.

As observed from Table 10, “against” met neither the first nor the second criterion as more than 50 % of participants knew this item and, moreover, less than six people showed gains after the treatment. This item was consequently removed from the final version. Regarding the four distractors, two of them (“on the right” and “next to”) were not familiar enough to most participants to be considered a good distractor. The decision was made to replace these distractors with the alternative “in the middle of” which, as will be seen later, was widely used in the descriptive task. In the end, the 12 items listed in Table 11 below were preserved for the main experiment.

Table 11. *Re-defined target items*

Targets	Distractors
On top of	Around*
On the right (left) - hand side	In front of*
At the back	In the middle*
Opposite	
In the corner	
Above	
Near	
Along the wall	
On the wall	

Regarding the ambiguous items of the test, especially in the “translation” part, all the words that caused confusion (such as “un cuadro encima de la cama” which could be translated in two different ways “a picture on top of the bed” or “a picture above the bed”) were specified or replaced.

6.2.2.4. Descriptive task

In the descriptive task the participants were asked to explain to a friend of theirs how was a living room (with an indication of the exact placement of the objects) they had just seen in London that they really liked. They were instructed to indicate the exact position of the objects by using a list included in the materials.

The main objective of the descriptive task was to make the participants use as many target items as possible. Each element to be described was designed to elicit the use of one target item. Through the second pilot, based on the renovated list of target items with three distractors, an adjustment in the pieces of furniture to mention was also needed. In this way, some of the items were removed (such as "a lamp" AT THE BACK which was almost never mentioned in the description and, if it was, it was done so by other means and not through the use of any target items) and some others were added (such as a laptop which was "ON TOP OF" the table or a coffee table which was "IN THE MIDDLE OF" the room).

6.3. Independent measures of cognitive task complexity

When dealing with the cognitive load of tasks it is crucial to have a series of instruments that measure the cognitive complexity of the tasks used in the experiment in an independent manner. This has been a claim in the field in the last few years in order to avoid circularity[22]. Up until now, researchers who opted to measure the complexity of tasks independently have normally used subjective ratings by means of an

[22] If it is predicted that a more complex task will cause disfluency and indeed learners turn out to be disfluent at doing a more complex task the confirmation is made that the task was more complex. However, in this case, task complexity would not have been tested independently.

affective variable questionnaire, which has been considered a good indirect indicator of the level of task complexity. Recently, however, some scholars (Norris & Ortega, 2003) have suggested that this measure is not sufficient to determine the degree of complexity of the task. In this context, several alternative instruments have been made available to researchers (see Section 3.5). In the present study, apart from the self-reported data questionnaire, it was decided to use subjective time estimation after completing each task. Subjective time estimation is considered to be the most practical and useful measure for delivery in a classroom context as it does not require special expensive equipment as in the case of brain-imaging techniques or eye-tracking. In what follows a detailed description of the two instruments used to independently measure cognitive task complexity is given.

6.3.1 Affective variable questionnaire

There are different versions of the self-reported questionnaire already used in TBLT literature (Gilabert et al. 2009, Robinson 2001a, 2007a). All the questionnaires employed in these studies are based on a Likert scale (6-point, 9-point, 10-point or 15-point). Normally, the statements which appear in the questionnaires concern the degree of difficulty of task completion, the level of stress and anxiety caused by the task and the interest and motivation that the task generated.

Before making a list of necessary statements to determine learners' anxiety and their perceived difficulty during task performance a factor analysis with the data collected from several earlier research projects (Gilabert 2005; Gilabert et al. 2009; Gilabert et al. 2011; Levkina & Gilabert, 2012) was run.

The 5 items of the affective variable questionnaire were subjected to factor analysis using SPSS Version 19. Prior to performing statistics the suitability of data for factor analysis was assessed. Results of the correlation matrix revealed the presence of several coefficients of .3 and above. The Kaiser-Meyer-Oklin value was .63 and Barlett's Test of Sphericity reached statistical significance ($p=0,001$), supporting the factorability of the correlation matrix. Principal components analysis revealed the presence of two components with eigenvalues exceeding 1, explaining 42,34 % and 30.46 % of the variance respectively. Therefore, it was decided to retain two components for further investigation.

Table 12. *Pattern/structure for coefficients (rotated component matrix)*

Item	Component 1	Component 2
Difficulty	-.686	
Confidence	.857	
Stress	.808	
Interest		.931
Motivation		.923

To aid in the interpretation of these two components, Varimax rotation was performed. The two-component solution explained a total of 73.00 % of the variance, with Component 1 contributing 42.34 % and Component 2 contributing 30.46 %. The results of the factor analysis showed that stress and confidence on the one hand, and interest and motivation, on the other hand, can be grouped as they go in the same direction (see Table 12). Therefore, in the present study, the affective variable questionnaire (see Appendix M) included three statements:

- on **task difficulty** perceived by the participants;
- on **stress / anxiety** the task might have caused;
- on **interest / motivation** the task generated.

6.3.2 Time estimation task

This measure has been extensively used in cognitive psychology (Block, Hancock & Zakay, 2010); however, it is relatively new as far as the field of applied linguistics is concerned (Baralt, 2010; Recio, 2011; Révész, Michel & Gilabert, 2012). The main goal of the instrument is to capture the difference between the perceived and real time of task completion, since it has been shown that people can easily keep track of time when the task they perform is simple; however, when things get more complex, people are more likely to lose track of time. The indicator of task complexity is the difference between perceived time and real time of task completion. In the present study, after each task (three tasks in total) the students were asked how much time they believed the task took them to complete (Appendix M). Further, three different analyses were carried out: (1) comparison of estimated times of tasks; (2) comparison of real times of tasks; (3) comparison of differences between perceived time and real time.

Additionally, a retrospective feedback questionnaire was also employed in the first pilot to obtain an explicit explanation from the participants regarding their assessment of task difficulty. Next, a brief history of the development of task complexity on the basis of the three pilots is presented.

6.4 Operationalizing +/- spatial reasoning: task elaboration

6.4.1 Input

In order to build up materials for the input section of the treatment, the transcriptions from the pilot were selected, reviewed and adapted to the overall design. Input contained three texts – instructions to furnish three different areas of a flat (living room, bedroom, and kitchen). The students were asked to follow the instructions and place all the furniture items on the layout given to them (see Appendix G and Appendix H). The following rules were established:

- 12 items from pretest (9 target items and 3 additional items) were distributed throughout the three texts;
- each item appeared three times in the texts;
- target items were underlined and presented in boldface (input enhancement

technique[23]).

6.4.2 Treatment tasks

Treatment tasks included three tasks which students were asked to complete individually. The difference among tasks was based on their complexity, which was manipulated according to the suggestions of the Cognition Hypothesis (Robinson, 2001b, 2003, 2005). The dimension selected for the tasks' manipulation is the amount of spatial reasoning demands. To make a decision on which criteria to choose in order to gradually increase task complexity along the three tasks, two sources were used:

[23] Input enhancement has shown its potential to draw learners' attention to highlighted items in a text.

- theoretical, many previous studies on psychology (see Chapter III) carried out research on the effects that varying cognitive demands may have on spatial reasoning. Table 13 with theoretical criteria is given below.

- empirical, three pilot studies preceded the main experiment in order to determine and to justify task operationalization as well as to determine the significant difference in cognitive task complexity among the three tasks (see Section 6.4.2.1, Section 6.4.2.2 and Section 6.4.2.3) .

es	Description	SIMPLE TASK	+ COMPLEX TASK	++ COMPLEX TASK
arroll	The fundamental principle for spatial acquisition for static relations is **"topological before axis-based"** order. Shape, size and distance are irrelevant in **topology**. The object of study is rather the "connectivity" of a figure. In **the axis-based description,** direction plays a constitutive role. In the conceptualisation of space, three directions have a privileged status: the up-down direction (vertical), the front-back-direction (sagittal) and the right-left-direction (lateral). In this case shape and size should be also considered. *Therefore, it is assumed that an axis-based conceptualization is associated with more complex reasoning than topological one.*	topological w/o size w/o shape	axis-based w/ size w/shape	axis-based w/ size w/shape
003 004), , (2004)	The choice of the system to use in the description / localization is orientation-dependent. People have been shown to be faster and more accurate when they deal with the same imagined perspective as the study's or when their observer's view coincides with the room-size layout position. *So, it is assumed that when the room-size layout position does not coincide with the participants' point of reference, the task is more complex.*	same position	same position	different position

rroll	The task can be easily completed if the participants know enough about the properties of the objects to be located. *So, it is assumed that if the participants are asked to place some unusual objects whithin a well-known place (such as a living room), it will make their task more complex.*	common objects	common objects	common & uncommon objects
003) (2004) ersky	A "relative frame" incorporates the speaker's coordinate system, involves a tertiary spatial relationship, and uses projective terms (left, right, front, back, above, below). An "intrinsic frame" uses an object's or another person's coordinate system, describes a binary relationship, and uses the same projective terms. If the participant naturally tends to use object-to-object perspective based on the room-size layout given to him, being forced to use an instrinsic frame will complexify the task.	no intrinsic frame	no intrinsic frame	with intrinsic frame

6.4.2.1. First pilot

The objective of the first pilot was to establish the relative cognitive complexity of the tasks manipulated along spatial reasoning demands, a resource-directing variable.

6.4.2.1.1 Participants

The participants of the study, a total of 13 people, were mostly University undergraduate and graduate students with a mean age of 30.15 (SD = 6.93). Eight of them were asked to perform the tasks in their L1 (2 in Catalan, 4 in English, 2 in Spanish) and the other five (non-native speakers of English) were asked to carry out the tasks in English as their L2 (for a more detailed description of the participants see Section 6.2.1).

6.4.2.1.2 Methodology

The experiment was carried out on an individual basis in a silent room. It took approximately 40 minutes to complete. All tasks were audio-recorded. Two series of the tasks were administered to the students and each series was based on different conditions created by manipulating cognitive task complexity along spatial reasoning demands.

The first series consisted of four tasks that were manipulated on the basis of spatial reasoning demands and also the number of elements involved in the task. In this way, the easiest task only dealt with basic spatial reasoning and the hardest task included higher spatial reasoning demands and also contained many elements (see Table 14).

Table 14. *Task complexity distribution through the four tasks*

++ Simple	+ Simple	+ Complex	++ Complex
Low Spatial Reasoning Few Elements	Higher Spatial reasoning Few elements	Low Spatial Reasoning Many elements	Higher Spatial Reasoning Many elements

In this way, lower spatial reasoning included (1) furnishing an empty space with a list of furniture items (always 7 items / categories), (2) considering room size and furniture size when placing items. In the higher spatial reasoning conditions, a perspective-taking factor was added (a delivery man was already in the room, so the student had to take into account the position of the delivery man while giving him instructions). Regarding +/- few elements, in the simple version of the tasks the students were already provided with the final list of the furniture items they had to put in the flat, however in the case of many elements, for each category (7 in total) they had to choose a piece of furniture among three options (see Appendix C for the simple task and Appendix D for the complex task).

The second series of tasks included three tasks, manipulated only in terms of spatial reasoning demands with the key factor being the points of reference. In this case, the simple task had two points of reference in the room and the most complex task did not have any point of reference[24]. The number of items was always the same (6 items per task). The students also had to take into account the size of the room and

[24] When some points of reference are already available, students can easily use them in order to exactly place new pieces of furniture. On the contrary, when no points of reference are available students have to find some other complex ways of explaining where to place new furniture items in order to be fully understood by the delivery man.

the size of the furniture items. However this time neither a perspective-taking factor nor a higher number of elements was included (see Appendix E for the simple task and Appendix F for the complex task). After each series of tasks the participants were asked to estimate the amount of time it took them to complete each task in the series (see Appendix A) and they were also given an affective variable questionnaire containing 10 statements to report their perception of task difficulty, self-confidence, motivation, time pressure etc. (see Appendix A) Finally, a retrospective questionnaire was also employed in order to retrieve information on how the participants found the tasks, what exactly seemed more complex or less complex to them, as well as some weak points of the tasks, and their general impression of the tasks (see Appendix B).

6.4.2.1.3 Results

Non-parametric statistical analysis[25] (Friedman test) of the affective variable questionnaire and time estimation showed an overall significant difference for stress and anxiety statements ($p = .022$ and $p = .026$ respectively) as well as for time estimation ($p=.001$) in the first set of tasks. No significant difference was detected among the task of the second series (see Table 15).

By means of a Wilcoxon pairwise test a significantly different pair of tasks were detected. For stress and for anxiety a significant difference was displayed between Task 1 – Task 3 ($p = .009$; $p = .028$), Task 2 – Task 3 ($p = .048$; $p = .023$). The descriptive statistics are shown in Figure 9 below where the lower numbers mean more “disagreement” and higher numbers “agreement” with the statement (for stress “This task was stressful for me” and for anxiety “The task made feel me anxious”).

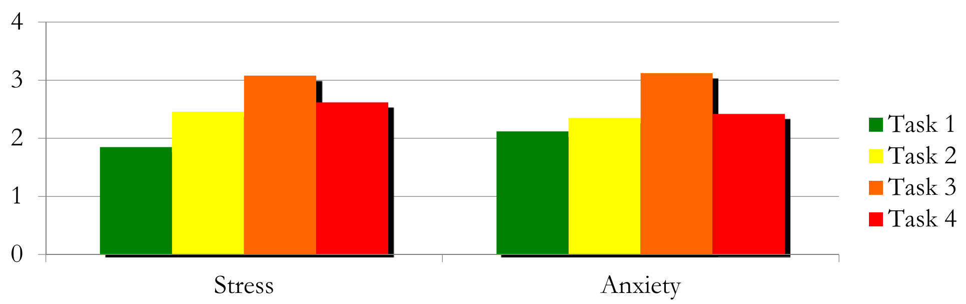
[25] Non-parametric statistics was chosen due to the low number of participants.

Table 15. *Differences at perception of the tasks with manipulated cognitive task complexity*

Statement	First series (4 tasks)	Second series (3 tasks)
This task was difficult.	.100	.812
I felt like I didn't have enough time before I responded.	.750	.135
I felt rushed during the task.	.595	.913
I was relaxed and comfortable completing the task.	.231	.116
This task was stressful for me.	**.022***	.143
I enjoyed doing the task.	.757	.097
The instructions of the task were clear.	.200	.097
The task made me feel anxious.	**.026***	.144
I could complete the task easily.	.612	.368
I had no problem in completing the task.	.052	.174

*p<0,05

Figure 9. *Descriptive statistics: stress and anxiety (pilot 1)*



As for time estimation, Wilcoxon pairwise statistical analysis showed a significant difference between Task 1 – Task 3 (p = .001), Task 1 – Task 4 (p = .001), Task 2 – Task 3 (p = .001), Task 2 – Task 4 (p = .001) in the first series of tasks. Here again, a significant difference is detected between the tasks manipulated along the number of elements.

Overall, no significant difference was reported in the affective variable questionnaire as far as the difficulty of the tasks is concerned. The participants reported being more anxious and stressed during the performance of the third task in comparison to the first two tasks which contained many elements to deal with. The same result was displayed in terms of time estimation. The significant difference was uniquely yielded in the first series of tasks and between Task 1, Task 2 and Task 3. This suggests that task complexity was determined by the number of elements and not by spatial reasoning required to complete the tasks. To discover some possible reasons for these results, the answers of the retrospective questionnaire were analysed.

All the answers were organized according to the category of task complexity manipulation along spatial reasoning.

Perspective-taking axe

In general, the participants admitted to finding the task where they had to take the perspective of a delivery man into account more difficult than the others.

"En general, les tasques del primer set on hi ha el repartidor i la tasca de l'habitació del segon set. Amb el repartidor costa més la referència i la perspectiva".[26]

"La perspectiva - arriba un moment que se t'oblida que has d'anar amb la perspectiva i directament fas el mateix que havies fet a l'altra tasca"[27].

"Em costa una mica el tema de l'esquerra i la dreta i a més si ho he de fer des de la perspectiva d'un altre"[28].

"It was a little messy when you had a guy and you had to indicate where to put the objects".

---

[26]In general, the tasks of the first set where a delivery man appears and the task of the double room of the second set seemed to me more difficult. With the delivery man it becomes more difficult to deal with the reference and the perspective (translation from Catalan – Mayya Levkina (M.L.).

[27] The perspective – sometimes it happens that you just forget you must include the perspective and you do exactly the same as in the previous tasks » (translation from Catalan – M.L.).

28 It's hard for me to distinguish the left from the right and even more if I have to do it from the other's perspective (translation from Catalan – M.L.).

More versus fewer references

It was observed that the points of reference on the map were beneficial, as some students reported them to be helpful and others said when they did not have explicit points of reference they had to find alternatives.

"La de la cuina com que hi havia dues referències doncs m'ha costat menys que no pas lo de l'habitació que n'hi havia una"[29].

"It was smaller and you didn't have any reference point, only the perspective of the door".

Unusual objects to put inside

An additional factor of perception of task difficulty was that of the unusual items to be placed in the flat. As their place was not immediately obvious, but required a bit of imagination, some people proved to have problems with those kinds of objects.

"Especialment la cuina, perquè si et demanen que col·loquis fora una sèrie d'objectes que no compraries mai si tens una cuina tant petita."[30].

"Más difícil en la cocina es intentar pensar en objetos espacialmente faltándome objetos de estos objetos[31]*"*.

Small space, many objects

Looking at the task with many elements almost all participants reported feeling stressed by such a large amount of information to process in 2 minutes. They

[29] The kitchen task was easier for me as it had two points of reference, compared to the double room which only had one" (translation from Catalan – M.L.).

[30]Especially the kitchen, because if you are required to put a number of items that you'd never buy if you had such a small kitchen (translation from Catalan – M.L.).

[31]The most difficult thing is thinking of the objects in the kitchen when you lack some other objects for the kitchen (translation from Spanish – M.L.).

considered the third task to be the most complex (according to the affective variable questionnaire and time estimation task) due to the number of elements. Subsequently, task 4 was perceived as having the same difficulty as task 3, since it was not new anymore. The other problem that participants reported sometimes was not having enough space in some of the rooms to place all the objects from the list, which was another point of task difficulty.

"La más pesada me pareció la 3 la que tenía menos espacio y mas objetos"[32].

"I've got the problems with bedroom 1, because it was smaller".

6.4.2.1.4 Conclusions for further task design

Based on the first pilot, a series of decisions were made regarding task design and task complexity distribution:

1. To manipulate only one variable over tasks (spatial reasoning demands) and keep the number of elements as a stable variable (unchangeable over tasks).
2. To exclude the size of the room and the size of the objects due to being confusing.
3. To use only a two-dimensional image of the room for the sake of convenience and ease.
4. To exclude the perspective – taking component as being too difficult to deal with.
5. To reduce treatment sessions from four to three in order to ensure a clear task difference in terms of cognitive complexity.

[32]The hardest one seemed to be task 3, the one which had less space and more objects to put inside" (translation from Spanish – M.L.).

6. To reduce time for preparation to one minute as it was considered sufficient for rehearsal yet and not long enough to process spatial expressions beforehand.

After having concluded the analysis, a new version of the treatment tasks was elaborated. Additionally, based on the oral production of English-native speakers of the tasks from the first pilot, materials for the input portion were designed together with the two-part vocabulary test and the descriptive task. Therefore, the second pilot served as a pilot of the whole experiment with a pretest, an immediate posttest and a delayed posttest.

6.4.2.2 Second pilot

The objective of the second pilot was to test the entire experiment (pretest, treatment, posttest, delayed posttest) and to see whether the obtained results showed any gains of the target items at the immediate posttest and delayed posttest. Moreover, it was crucial to determine whether the modified series of tasks (3 tasks) was correctly operationalized in terms of sequencing. The detailed description of the participants and the procedure is given in Sections 6.2.2.1 and 6.2.2.2.

6.4.2.2.1 Task design

Following the guidelines of the first pilot, in the second version of tasks only one resource-directing variable was retained (spatial reasoning demands). The number of elements (furniture items) was always kept consistent (8 items per task). In addition, several cognitive factors were chosen to operationalize task complexity based on the previous pilot (more versus fewer references; small space with many objects to place in etc.). The new distribution of task complexity is shown in Table 16.

Table 16. *Distribution of task complexity (+/- reasoning demands) in pilot 2*

+ Simple	+ Complex	++ Complex
No reasoning demands	Semi-furnished flat	Unfurnished Flat
Already Furnished Flat	Some points of reference	No points of reference
Big space	Big space	Small space

In the simple version of the task the participants were asked to simply give the instructions based on the layout with the objects already placed in it. In the more complex task they had four furniture items already in the room, but they still had to place the rest of the objects and explain the position of all eight objects to the delivery man. And finally in the last, most complex task, they had to completely furnish a given space. No points of reference were provided and the space itself was small so it had an additional difficulty of fitting everything inside. As for the sequences, the tasks were organized in three different ways: (1) from simple to complex; (2) from complex to simple; (3) randomized.

6.4.2.2.2 Results

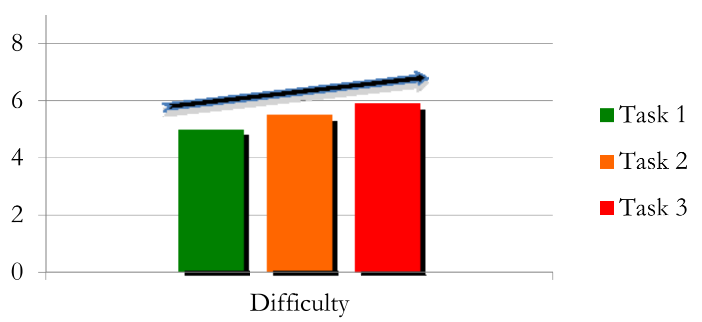
As seen in Table 17, repeated measures ANOVAs for the three affective variables displayed a significant effect for the perception of confidence and motivation; however, no significant effect was found for difficulty.

Table 17. *Main effects for affective variable questionnaire (pilot 2)*

Dependant Variable	Df	Sum of squares	F-value	p-value	η^2
Difficulty	2	1.826	.905	.408	.020
Confidence	2	14.812	6.720	.002*	.130
Motivation	2	18.290	6.281	.003*	.122

Pairwise comparisons for confidence and motivation showed that the students felt more confident and more motivated when performing the most complex task, and also more confident doing the third task in comparison with the second task and more motivated doing the second task than the first task, while in terms of difficulty they did not show any statistically different perception of task difficulty among the three conditions. Nevertheless, when looking at the descriptive statistics, perception of task difficulty followed the expected pattern from simple to complex (see Figure 10).

Figure 10. *Descriptive statistics: difficulty (pilot 2)*



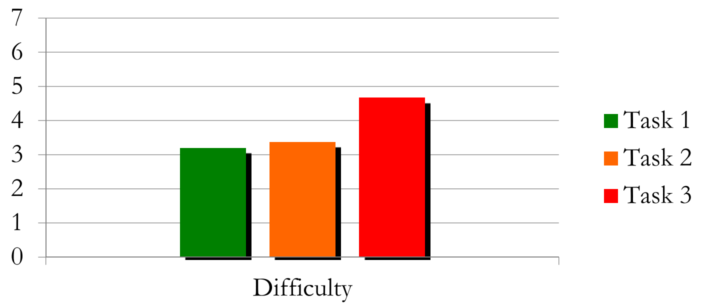
Repeated-measure ANOVA for time estimation task indicated the significant overall effect among the three tasks (see Table 18).

Table 18. *Main effect for time estimation task (pilot 2)*

Dependent Variable	Df	Sum of squares	F-value	p-value	η^2
Time Estimation	2	30.029	3.913	.023*	.080

Pairwise comparisons showed a significant difference between the simple task and the most complex task (Task 1 - Task 3). No other differences were detected. As before, when analysing descriptive statistics, a clear tendency from simple to complex was observed (see Figure 11).

Figure 11. *Descriptive statistics: time estimation task (pilot 2)*



6.4.2.2.3 Decisions and conclusions for further modifications of tasks

In the second pilot the difference among the tasks was more clearly established, however the difference was not significant enough to state that the three tasks differed in cognitive load. Some modifications needed to be introduced in the design to make the difference among them absolutely salient.

1. The distribution started with no spatial reasoning demands and no intermediate steps in elaborating tasks were included, e.g. Task 1 (fully furnished) had no reasoning demands, whereas Task 2 (completely empty space to furnish) already had a high cognitive load.

2. More conditions of spatial reasoning were introduced (i.e object size, room size, perspective) in order to clearly distinguish Task 2 from Task 3 (in terms of cognitive complexity). However, as opposed to the first pilot, the sizes for room and for furniture items were given less precisely so as to facilitate easy measurement calculations (such as 2m. or 2.5, and not 2.35m or 1.65; furthermore the perspective element appeared only in the third task, with the aim of making it much more complex for all participants compared with the second task.

3. Participants were not given time to plan their subsequent explanations, which allowed to the controlling for any rehearsal effects.

6.4.2.3 Third Pilot

The goal of the third pilot study was to test the distribution of task complexity along the three tasks. For this reason, 10 participants (graduate university students) were asked to do three modified treatment tasks in their native languages and to complete the affective variable questionnaire and time estimation task after each treatment task.

6.4.2.3.1 Results

This time, repeated-measure ANOVAs did display a significant difference for difficulty, but not for confidence or motivation (see Table 19). Pairwise comparisons for task difficulty showed a significant difference for all pairs (task1 – task2, p = .001; task1 – task3, p = .001; task2 – task3, p = .022).

Table 19. *Main effects for affective variable questionnaire (pilot 3)*

Dependant Variable	Df	Sum of squares	F-value	p-value	η^2
Difficulty	2	46.815	60.551	.001	.883
Confidence	2	8.259	2,018	.165	.201
Motivation	2	.111	.229	.798	.028

Repeated-measure ANOVAs for the time estimation task indicated significant differences for time estimation and real time variables, while no significant difference was found for time difference (see Table 20). Pairwise comparisons showed a significant difference among all tasks for both time estimation (task1 – task2, p = .007; task1 – task3, p = .012; task2 – task3, p = .037) and real time (task1 – task2, p = .011; task1 – task3, p = .011; task2 – task3, p = .033).

Table 20. *Main effects for time estimation task (pilot 3)*

Dependant Variable	Df	Sum of squares	F-value	p-value	η^2
Time Estimation	2	19969.593	17.731	.001	.689
Real Time	2	4808.926	11.244	.001	.584
Time Difference	2	1339.693	2.018	.165	.201

Based on the results obtained from the third pilot, the tasks were considered to be valid in terms of task complexity distribution and ready to be included in the main study.

6.4.2.4 Summary of the decisions made on the basis of the three pilots

Across the three pilots a series of decisions on the following criteria had to be made: (1) number of dimensions to include; (2) number of conditions to introduce; (3) room type to offer; (4) allotted time for preparation. In Table 21 an overview of task characteristics and their modifications are presented. The final version of the tasks included (1) only one variable along which the tasks were operationalized (spatial reasoning demands) to ensure the effect of this variable on L2 development (in the case L2 development was to take place); (2) conditions related to spatial reasoning were gradually introduced in the tasks, starting with the most simple task with no conditions; (3) room type was modified from furnished to unfurnished; (4) no time for preparation was given to the students to avoid task rehearsal before starting performance.

Table 21. *Treatment tasks across three pilots*

Category	Pilot 1 (series of 4 tasks)	Pilot 1 (series of 3 tasks)	Pilot 2	Pilot 3
Number of variables	Spatial reasoning (all tasks) Nº of elements (tasks 3,4)	Spatial reasoning (all tasks)	Spatial reasoning (all tasks)	Spatial reasoning (all tasks)
Number of conditions	Size (all tasks) Perspective (all tasks)	Points of reference (all tasks)	Two points of ref. (task 1) One point of ref. (task 2) No points of ref. (task 3)	Size (task 3) Perspective (task 3) Unusual objects (task 3)
Room type	Unfurnished (all tasks)	Partially furnished (task 1) Sparsely unfurnished (task 2) Unfurnished (task 3)	Furnished (task1) Semi-furnished (task 2) Unfurnished (task 3)	Furnished (task 1) Unfurnished (task 2) Unfurnished (task 3)
Time for preparation	2 minutes	2 minutes	1 minute	No time

6.5 The current study: design and methodology

6.5.1 Participants

Originally, 89 participants split into four groups took part in the current study with two groups in a laboratory setting and two groups in a classroom setting. The decision to use two contexts in the analysis of the effects of spatial reasoning demands on L2 development was primarily motivated by the fact that laboratory studies are usually criticized for the lack of ecological validity under real classroom conditions. Participants were adult students of English as a foreign language at the University of Barcelona. The first laboratory group was enrolled in the Applied Linguistics course as part of their four-year English studies bachelor's degree. The second laboratory group was enrolled in the English teaching course as part of their four-year English studies degree. The third and the fourth classroom groups were students of English in the School of Modern Languages (EIM, UB). The students in the laboratory context were given a proficiency test in English (X-Lex and Y-Lex[33]) to determine their level, which placed them at an intermediate level. The students in the classroom context had an intermediate level of English, which would correspond to B2 (Common European Framework). They were assigned to the fourth level (maximum level being 6) according to their results obtained either (1) after having done a level test or (2) after having successfully completed the previous stage (level 3).

From the original pool, a number of participants were eliminated from the study for the following reasons: (1) missing sessions; (2) demonstrating extensive prior knowledge of the target spatial expressions; (3) not meeting 85 % accuracy on the processing component in the working memory test. In the end, a total of 28 participants were eliminated for the above-mentioned reasons. The total number of participants in the study was 61.

The age of the four groups is distributed as shown in Table 22. 6 participants from Group 2 were Erasmus students: 1 Belgian, 2 Germans, 1 Estonian, 1 Dutch, and 1 Chinese. Several participants of both laboratory groups had previously travelled to an English-speaking country for study purposes (13 people in Group 1 and 7 people in Group 2). Regarding the classroom groups in Group 4 only two participants reported that they had been in a study abroad program in an English-speaking country. In the laboratory groups most students started learning English in primary school with the mean starting age being 6.50 (SD=2.09) for Group 1 and 8.42 (SD=2.81) for Group 2. In the classroom groups the variety of participants as far as their starting age is concerned was higher, so they ranged from 3 to 42 years. When

[33] X-Lex (Meara & Milton, 2005 ; Y-Lex (Meara & Miralpeix, 2006)

looking the starting ages, Group 3 had a mean score of 12.50 (SD=9.47) whereas Group 4's mean was 12.22 (SD=11.58). In terms of communicative use of English, laboratory groups reported higher levels of engagement in oral and written communication in English, than the two classroom groups (see Table 22).

Table 22. *Background information per group*

Group	N	Age M(SD)	Gender	Age of Onset	Use of English (written)		Use of English (Oral)	
					N	Frequency M (SD)	N	Frequency M(SD)
Lab 1 (Simple – Complex)	20	19.65 (1.18)	16 female 4 male	6.50 (2.90)	16	3.13 (.885)	11	3.45 (.820)
Lab 2 (Randomized)	19	23.92 (3.30)	14 female 5 male	8.42 (2.81)	6	3.83 (.753)	9	3.22 (.833)
Class 1 (Simple – Complex)	11	27.80 (6.10)	10 female 1 male	12.50 (9.47)	1	1	4	2.25 (1.50)
Class 2 (Randomized)	11	27.67 (9.77)	8 female 3 male	12.22 (11.58)	4	2.75 (1.71)	3	2.67 (.577)

6.5.1.1 Group comparability

It was necessary to ensure group comparability:

1) within each context (laboratory and classroom) so as to see whether the students from every setting were comparable at the level of their bio-data in order to proceed to further analyses of sequence effects on L2 development;

2) between the two contexts (laboratory and classroom) in terms of the bio-data reported by the participants. One-way ANOVA on the following variables were run: age, gender, age of onset, use of English (written), and use of English (oral). The results are shown in Table 23 and in Table 24.

Table 23. *Differences between the two laboratory groups*

Variable	SS	df	MS	F	Sig.
Age	106.408	1	106.408	27.176	.001*
Gender	27.552	1	27.552	4.865	.421
Age of onset	2.189	1	2.189	3.003	.035*
Use of written English	.267	1	.267	.392	.099
Use of oral English	.131	1	.131	.662	.539

There was a significant difference in the age and in the age of onset between the two groups. Looking back at the descriptive statistics, the raw difference of means for age (19.65 versus 23.92) and age of onset (6.50 versus 8.42) was not thought to be a drawback for the objectives of the present study. So, the decision was made to continue with the quantitative analysis of the two laboratory groups since there was no significant difference detected at the level of the use of written or oral English.

Table 24. *Differences between the two classroom groups*

Variable	SS	df	MS	F	Sig.
Age	.057	1	.057	.001	.979
Gender	.182	1	.182	1.176	.291
Age of onset	.214	1	.214	.002	.967
Use of written English	2.450	1	2.450	.840	.427
Use of oral English	.298	1	.298	.201	.673

Regarding the two classroom groups, no significant differences were shown in terms of any of the analysed variables. Looking at the overall comparability of the two contexts (see Table 25), there was a significant difference for age (p=.001), age of onset (p=.013) and use of oral English (p=.028).

Table 25. *Differences between the two contexts*

Variable	SS	df	MS	F	Sig.
Age	430.920	1	430.920	16.680	.001*
Gender	.078	1	.078	.431	.514
Age of onset	239.406	1	239.406	6.684	.013*
Use of written English	3.435	1	3.435	3.070	.092
Use of oral English	4.402	1	4.402	5.431	.028*

Taking into account the importance of the use of English outside class, the significant difference for the use of English forced us to make the preliminary decision, (subject to further revision on the basis of the pretests) not to engage in a comparison of the two different contexts. Also, the final number of participants in the classroom groups is much lower than in the laboratory groups (11 versus 20), which is another impediment for comparison. As a consequence, our analysis will focus on the laboratory groups and the results from the classroom groups will only be used to confirm or reject the ecological validity of the tasks employed in the main experiment.

6.5.2 Design

The study was based on a pretest – immediate posttest - delayed posttest design. Each experiment group corresponded to its real educational group.

1. **Lab 1**: from simple to complex
2. **Lab 2**: randomized
3. **Class 3:** from simple to complex
4. **Class 4:** randomized

The group classification was based on the following criteria: (1) sequence, whether the treatment tasks were organized following the simple to complex model or they were randomized; and (2) context, whether the treatment sequence was done in classroom (collectively) or in the laboratory (in small controlled groups). In that case, the between-subject factor was task sequence. The within-subject factors included the results and the gains obtained at each point in time (pretest, immediate posttest and delayed posttest), and learners' affective variable questionnaires, as well as learners' time estimations of the duration of task completion. Finally, working memory capacity was a covariate, considered a mediating variable.

6.5.3 Treatment session

The treatment session consisted of two parts. The first was input and the second included three tasks organized according to their cognitive task complexity.

6.6.3.1 Input

During the input phase the students were presented with three tasks. The objective was always the same and was explained by asking them to imagine that they were delivery men and they had just come to their client's home and now by following the instructions of the client, they had to furnish the required space. The students worked in pairs. Each of them was given a photocopy of the input tasks with three descriptions and three plans (empty and then furnished). On an empty layout of a flat they had to draw the furniture items (simply a box with the label, in order to save time and effort). Before task execution, the teacher was asked to give example instructions for the first task to ensure that the students understood what was required. The subsequent instructions were given by the students (Task 2 – Student A and Task 3 – Student B) in a way that each student gave instructions for placement of furniture once and then also followed the instructions of his / her peer (see Appendix G and Appendix H for the full version of Input materials). Importantly, the three texts of the input had an equal number of items (9 target items + 3 distractors) which appeared three times in the different texts. This was done to try to balance and control the number of exposures to input for each target item the participants would be exposed to.

6.5.3.2 Treatment tasks

The oral treatment consisted of three tasks created on the basis of the spatial reasoning demands they imposed on the participants' processing (a resource-directing variable of cognitive task complexity). The tables were organized from simple to complex in the case of Lab 1 and Class 1 or randomly in the case of Lab 2 and Class 2 (see Table 26).

Table 26. *Two task sequences*

	Task 1	**Task 2**	**Task 3**
Sequence 1 (from simple to complex)	simple	+ complex	++complex
Sequence 2 (randomized)	+ complex	simple	+ + complex

These tasks were to be done individually. The students were instructed to imagine they had just moved to London and they needed to furnish their new flat. The whole space had a double room (simple task), a kitchen (+ complex task), a living room (++ complex task), and a single room. One by one, except for the single room, they had to furnish the whole space. To do so, in each case they were provided with a list of items to place inside (the number of items was always the same and equalled twelve).

In **Task 1** (simple task) the students were already given the placement of the objects drawn on the layout they had. So, the only thing they had to do was to explain (to verbalize) the furniture distribution in the double room by giving instructions to their peer delivery man. Here, no spatial reasoning demands were involved in the task.

In **Task 2** the students were asked to furnish the kitchen. This time they did not have a finished layout with the whole distribution already included, but rather they had to decide whether to put each of the items in the kitchen. So, some spatial reasoning demands were required in the task.

In **Task 3** the students again were asked to furnish a new space (a living room) but here once again they did not have a pre-established layout to be described; apart from that, while putting the furniture items into the living room they had to take into account the size of the room given in meters, the size of the objects and also the perspective of a delivery man who was also in the layout. In terms of spatial reasoning, this task was considered to be the most complex (see Appendix I and Appendix J for the full version of the tasks) (see Table 27).

Table 27. *Distribution of task complexity through the three tasks*

	Task 1	Task 2	Task 3
Criteria of Task Complexity	no reasoning demands	empty space	empty space size perspective

As seen in Section 6.4, to ensure the significant cognitive complexity of the tasks the three tasks were previously carefully piloted and subsequently modified in order to obtain a significant difference of their difficulty perceived by the participants of the pilots. The description and the results of these pilots are demonstrated in Section 6.4 of this chapter.

6.5.4 Control tests

Two different tasks were used to assess participants' L2 development in the study: a descriptive task (productive) and a receptive two-part vocabulary test (part I - translation of target expressions from Spanish into English; part II - multiple-choice test).

6.5.4.1 Descriptive task

A descriptive task (see Appendix K) was employed to assess participants' previous knowledge of the target items, as well as the gains immediately after treatment and two weeks after treatment. It was a simple picture descriptive task. The story to relate was the same as in the treatment session: they had to imagine they were moving to London. They had just come from a friend's place, which seemed fabulous to them. So, once at home they decided to write an e-mail to their boyfriend/girlfriend who was about to come join them describing exactly what they saw in the other flat. It was a written task, with no reasoning demands (as the students only needed to describe the picture). The decision not to include reasoning demands in the control test was taken to avoid any possible effect of the preliminary task on the main sequence (by potentially moving the treatment series to the pretest set). As in the treatment, the task had a list of 12 items to be mentioned in the description, designed to elicit the use of each of the target items. This was also piloted beforehand with 4 native speakers of English (as described in Section 6.2).

6.5.4.2 Two-part vocabulary test

This test (see Appendix L) was provided after the productive descriptive task in order to prevent the participants from using some target items in the descriptive task. The first part of the test consisted of a simple translation from Spanish into English of the target items together with the three distractors (a total of 12 expressions to translate). The distractors were included in the text in order to avoid discouraging participants with a high number of unknown items to be translated.

The second part of the test was a multiple-choice task, which also included the same number of items to deal with (12 items, 9 target items and 3 distractors). For each gap, the students had four options where only one was correct. The test was previously revised by native speakers and piloted to ensure its reliability. As an example, a test item is given below:

The stools will go just ____________ stove.

a) opposite to the   b) opposite   **c) opposite the**   d) opposite of the

6.5.4.3 Background questionnaire

Before starting the experiment, all participants were asked to complete a background questionnaire (see Appendix N). This tool was important to retrieve basic information about age, gender, age of onset, years of learning English, stays abroad in English-speaking countries, knowledge of additional foreign languages apart from English, different L1s, reasons for learning English, turning points in the learning of the language, and frequency of reading, writing, listening and speaking in English. Since the four groups did not have the same linguistic trajectory or educational background (laboratory groups came from English university degree courses and classroom groups came from a whole variety of professions, such as medicine and physics), the questionnaire was essential to be able to establish the level of comparability in the present study.

6.5.4.4 Measurements of task difficulty

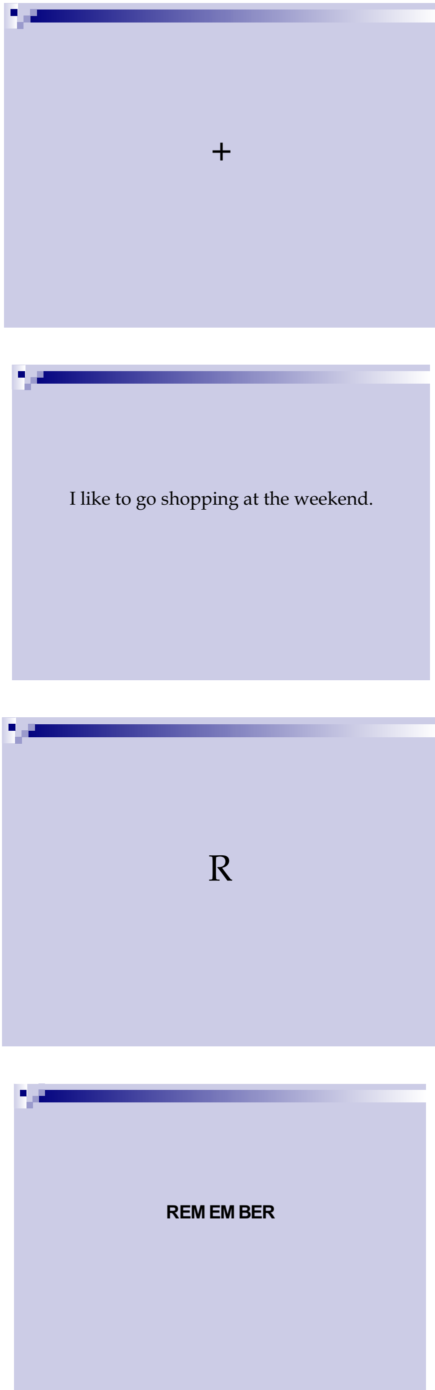
As seen in Section 3.5, the most frequently used instruments to reach this goal are subjective ratings, which allow for the obtaining of participants' perception of task difficulty, their level of anxiety, confidence, interest, and motivation. The statements within the questionnaire could vary depending on the purposes of the study. In the present study the decision was made to use only three statements related to task difficulty, self-confidence and motivation. Therefore, it slightly differs from earlier versions of the same test where additional questions on interest and stress were normally included. However, after having run a factor analysis with the use of multiple previous data, it became possible to combine self-confidence with stress on the one hand and interest with motivation on the other hand (see Section 6.3.1 for a detailed statistical report on this). In the questionnaire, learners were asked to circle a number on a 10-point Likert scale to indicate to what degree they agreed or disagreed with each of the three statements.

6.5.4.5 WM Span – automated reading span

In the present study the decision was made to use an advanced version of the reading span for measuring working memory capacity of the participants. The automated version of the reading span was created at the research laboratory at Georgia Technical University (Unsworth et al., 2005)[34]. This version of the span was designed to be used with software called E-Prime. The task was entirely mouse-driven with all the explanations given on screen. The researcher's assistance was not necessary. The test consisted of three blocks of practice and an experimental block. In this test, the participants were first shown a sentence and had to judge whether it was plausible or not and then after each sentence a letter appeared on the screen. At the end of the set, they were required to recall all the letters in the order they appeared. The sets ranged from 3 to 7 sentences, with three trials for each size and a total of 75 sentences and letter. An example of an ARSPAN Stimuli is presented in Figure 12 below:

[34] The test was adapted for native speakers of Spanish and Catalan, and subsequently piloted with results showing high internal consistency (α =.876).

Figure 12: *Example of ARSPAN stimuli*



6.6 Procedure

This experiment was carried out within the fall semester (laboratory groups) and the spring semester (classroom groups). All 61 participants attended three to four sessions each (the number of sessions depended on whether they did the ARSPAN on the same day as when they completed the treatment or not). All participants took part in the experiment on a volunteer basis and they were given credit for their participation. The participants were recruited via their teachers who explained the objective and the organization of the study. The students of the laboratory groups did the pretest and the delayed posttest in class, whereas the treatment session was carried out in small groups (up to 4 people). The classroom students did the entire experiment during their class hours (see Figure 13).

Figure 13. *Experimental procedure*

<table>
<tr><th></th><th>LAB 1, LAB 2</th><th>CLASS 1, CLASS 2</th></tr>
<tr><td>Day 1</td><td colspan="2">Background questionnaire
Pretests: a descriptive task, a two-part vocabulary test</td></tr>
<tr><td colspan="3">one week</td></tr>
<tr><td rowspan="2">Day 2</td><td colspan="2">Input + three treatment tasks
Immediate posttests: a descriptive task, a two-part vocabulary test</td></tr>
<tr><td>Affective variable questionnaire
Time estimation of the three tasks
Working memory span (ARSPAN) – LAB 1</td><td></td></tr>
<tr><td>Day 3</td><td>Working memory span (ARSPAN) – LAB 2</td><td></td></tr>
<tr><td colspan="3">10-14 days</td></tr>
<tr><td>Day 4</td><td colspan="2">Delayed posttests: a descriptive task, a two-part vocabulary test</td></tr>
</table>

On the **first day** participants were given the background questionnaire (with a signed consent form to take part in the experiment) and the two pretests (a descriptive task, followed by the two-part vocabulary test). The average time of the session was approximately half an hour.

The **second day** (a week after the pretest) consisted of a treatment session, immediate posttest and, in the case of Group 1, the day concluded with the WM test, ARSPAN. The duration of the session ranged from one hour and a half to two hours. In all cases participants were first provided with the instructions and explanations of what they were going to do. The laboratory task performances were audio-recorded. Each task was preceded by a researcher / teacher[35] giving detailed instructions on what they were expected to do. In the case of Group 1, ARSPAN was collected individually with the use of DELL laptops which had E-Prime installed. It normally took half an hour to complete the test and a separate room was booked to ensure a quiet place for better concentration. In the case of the laboratory groups, after completing each of the three tasks the participants were asked to fill in an affective variable questionnaire to indicate their perception of the task and also to provide an estimated time of task completion (see Appendix M). Group 2 had an additional session (2b) to complete the ARSPAN. This task was carried out in a computer room with the programme installed on each computer. Due to the number of participants (22 people) and to make sure the instructions were clear, the researcher came through every screen with instructions to give an additional explanation of what the students were expected to do. This session lasted 35-40 minutes. The last session (two weeks after treatment) was administered on the **third day** within class hours and lasted half

[35]Researcher – for the laboratory groups, Teacher for the classroom groups

an hour. Once again, the students were given the descriptive task and the two-part vocabulary test.

Regarding the classroom groups, the experiment was also split into three days. Participants completed all parts of the experiment within the classroom sessions. The working memory test, the affective variable questionnaire and the time estimation task were not included in the sessions because of lack of time (for the affective variable questionnaire) and the impossibility of recording the participants (time estimation) or giving them an ARSPAN (lack of necessary number of laptops).

6.7 Coding and scoring procedures

6.7.1 Background questionnaire

The background questionnaires (see Appendix N) were offered in Spanish and in Catalan in accordance with the dominant L1 of the participants. The answers from the background questionnaire were qualitatively and quantitatively coded in order to be run for group description and group comparability. The categories and the correspondent codings are shown in Table 28.

Additionally, they were asked to mark the factors that influenced their learning of English among the following: age of onset, family, friends, effort, aptitude, importance of results, motivation, teachers, extra-curricular classes, and stay abroad; to describe a turning point in their learning of English in the cases where they had one and finally to rate their knowledge of all the languages including their native(s) one(s) in terms of reading, writing, listening and speaking on the scale of ten (where ten was the highest score). This information was not included in the last version of the thesis analysis for the sake of space.

Table 28. *Codification of the background questionnaire*

Categories	Coding
Gender	1 – female 2 - male
Extra-school classes	a) number of years b) frequency (times a week)
Stay abroad	a) country (1 – UK, 2 – USA, 3 – Ireland, 4 – other English-speaking countries) b) duration (in months)
Other foreign languages	a) language (1 – French, 2 – German, 3 – Italian, 4 – Japanese, 5 – other foreign languages) b) duration (in years)
TV programmes in English	1 – once a year 2 – once a semester 3 – once a month 4 – once a week 5 – every day
Reading of long texts in English	1 – once a year 2 – once a semester 3 – once a month 4 – once a week 5 – every day
E-mail writing	1 – once a year 2 – once a semester 3 – once a month 4 – once a week 5 – every day
Use of English outside the University	1 – once a year 2 – once a semester 3 – once a month 4 – once a week 5 – every day

6.7.2 Descriptive task

The descriptive task was scored in four ways. Absolute scoring, percentage absolute scoring, frequency scoring and frequency percentage scoring were applied for this task assessment.

In the absolute score, 1 point was assigned to three categories: (1) a correctly used expressions of space, (2) each target expression of space; (3) each target expression of space used correctly; and 0 was given to (1) an incorrectly used expression of space, (2) a non-target expression of space, (3) a target expression of space used incorrectly.

Figure 14. *Example of scoring in the descriptive task*

Dear Sonia,

Yesterday I put all my furniture on my wonderful living room! In this e-mail I sent you a photo. ***At the back of (Target, T)*** *the room there is a big sofa very comfortable.* ***Above (T)*** *the sofa there are 2 pictures.* ***In the middle of (Correct, C)*** *the room there is a rug,* ***on the top (Target Incorrect, TI)*** *a coffee table and a laptop and* ***above (T)*** *the table there is a red lamp.* ***On the right-hand side (T) in front of (C)*** *the coffee table there is an armchair, and another armchair* ***on the left side (T)****. Also* ***near (T)*** *the sofa* ***on the left side (T)*** *there is a plant.* ***On the right-hand side (T)*** *there is a table with white chairs. Finally there is a wooden chair* ***behind(C)*** *the sofa. What do you think about my living room?*

	All Correct expressions	**All target expressions**	**Correct target expressions**
Absolute Scoring	12	9	8
Percentage Scoring	91.67 %	75%	66.67%

	All target items	**Correct target items**
Frequency Scoring	5	4
Frequency Scoring in %	55.56%	44.44%

In the percentage scoring the decision was made to use a fixed maximum number of prepositions equal to the number of objects to be mentioned in the task, 12. The percentages were counted for (1) correctly used expressions of space, (2) each target expression of space; (3) each target expression of space used correctly.

Finally, frequency scoring worked as follows: the first score corresponded to (1) all target items used in the task, and (2) the target items used correctly. The maximum score was equal to 9. The second score corresponded to the percentage of the first two values, where the maximum (100%) was 9. Above an example of all scoring methods (see Figure 14) is given.

6.7.2.1 Interrater agreement

A second researcher coded 10% of the data. To test the reliability of the scoring, the two codings were subsequently compared and the results showed a high reliability for all of the assessments: absolute scoring (all correct expressions 98 %, all target expressions 99 %, correct target expressions 100%); percentage scoring (all correct expressions 98 %, all target expressions 99 %, correct target expressions 100%); frequency scoring (all target items 96 %, correct target items 97 %; Frequency Scoring in %, all target items 96 %, correct target items 97 %).

6.7.3 Two-part vocabulary test

For both parts of the vocabulary test, 1 point was given to a correct answer, whereas 0 points were assigned to incorrectly translated, untranslated or unanswered items in the first part and to incorrectly answered or unanswered items in the second part. The total score obtained for each of the parts was 12. Additionally, it was considered important to look at the gains in the immediate posttest and delayed posttest. For this reason, an additional scoring was added, i.e. gains at immediate posttest (posttest - pretest) and gains at delayed posttest (delayed posttest - pretest).

6.7.3.1 Interrater agreement

As in the case of the descriptive task, a second researcher coded 10% of the two-part vocabulary test. Interrater reliability was 99%.

6.7.4 Automated reading span

Firstly, the researcher had to ensure that the processing component of the working memory span was fulfilled, which was essential for the test. Following the guidelines of Conway at al. (2005), all the participants were checked for at least 85 % of the correct identification of sentence plausibility. In the present study, all the subjects scored more than 85%, so their results were considered as valid for further analyses and therefore no one was excluded.

The automated version of the test used in the present study (Unsworth et al., 2005) computed the score for each subject at the end of the test, which was also recorded in the .txt result document generated by the E-Prime programme. As opposed to the early versions of the test, where absolute scoring was used, the present scoring took into account all the information obtained from the participants (not only from the sets recalled correctly). Up until now, four scoring methods have been used:

1) **Partial-credit unit scoring**: the mean proportion of elements within an item that were recalled correctly.
2) **Partial-credit load scoring**: the sum of correctly recalled elements from all items, regardless of whether the items are perfectly recalled or not.
3) **All-or-nothing unit scoring**: the proportion of items for which all the elements were recalled correctly.

4) **All-or-nothing load scoring**: the sum of the correctly recalled elements from only the items in which all the elements are recalled in a correct serial order.

In the present test, partial-credit unit scores were computed as they were considered the most convenient and advantageous for the measurement of WMC.

6.8 Statistical analyses

To address the first research question on the independent measures of cognitive task complexity, various statistical tests were run. For the affective variable questionnaire, repeated-measures ANOVA was first performed to find overall differences in the affective perception of the tasks (difficulty, confidence and motivation) along the three levels of complexity. Next, pair-wise comparisons were applied to investigate any significant difference among the levels. For time estimation, first repeated-measure ANOVA was used to detect any overall significant effects between the three levels of cognitive task complexity as measured by estimated time, real time and time difference. Afterwards, pairwise comparisons were carried out to establish differences among the tasks.

To address the research question on L2 development a one-way ANOVA on all the pretest measures for the descriptive task and the two-part vocabulary test was performed to determine the comparability of the two contexts (laboratory and classroom). As the results displayed significant differences between the two contexts, the decision was then made to present the results on task sequencing and L2 development separately for the two contexts. First, the results are shown for the laboratory setting, followed by the results for the classroom setting Regarding the measures, the results for the descriptive task are presented before the results of the two-part vocabulary test. Two series of tests were used: in order to determine L2 development over time and to find out whether task sequencing from simple to complex is more beneficial over other kinds of task sequencing. For the descriptive task, results were given for the general score, the frequency score and the gains over time. In order to address the first part of the research question, inferential statistics included a repeated-measure ANOVA to determine an overall significant difference over time, after which pairwise comparisons were run to detect any significant difference among the tasks. To address the second part of the research question, one-way ANOVA was used to detect any significant differences between the two sequence groups in the immediate posttest, delayed posttest and in the gains (pretest - immediate posttest; pretest – delayed posttest). The same tests and procedures were performed in the case of the two-part vocabulary test, where the results were offered for part I, part II and the total.

Finally, to address the research question on the mediating role of WMC in L2 development of spatial expressions by means of task sequencing, Pearson correlations were performed between the two measures of WMC (WM1 and WM2) for the students of the laboratory group altogether, the five measures of the descriptive task (all correct expressions, all target, correct targets, frequency all targets, frequency correct targets), and finally, for the three measures of the two-part vocabulary test (part I, part II, total). Next, the students were split into two groups according to the level of their WMC (high versus low) and the same correlations were repeated. To perform statistical analyses, the statistical package for the social sciences (SPSS) 19 was used. An alpha level of $p < .05$ was set for all tests.

CHAPTER 7

RESULTS

In this chapter all the statistical analyses related to the three research questions are presented. First, the results of independent measures of task complexity (i.e. affective variable questionnaire and time estimation task) for the laboratory groups are presented. Next, the results of L2 development are given (separately for laboratory and classroom contexts). For a better visualization of results on L2 development, they are split into two parts (see Section 7.2.1 and Section 7.2.2): the first part provides the quantitative results of the laboratory groups, while the second provides the quantitative results of the classroom groups. The chapter finishes with the description of the results on the relationship of WMC as a mediating variable of L2 development in the context of task sequencing.

7.1 Research questions on independent measures of cognitive complexity

1) Does manipulating task complexity from simple to complex affect learners' reported perception of cognitive complexity?

As stated previously, in order to carry out a deeper and more precise analysis of task complexity two measures were selected: (1) an affective variable questionnaire and (2) a time estimation task. In the following sections the results of these variables are presented.

7.1.1 Affective variable questionnaire

The affective variable questionnaire was comprised of three statements related to (1) difficulty, (2) self-confidence, and (3) motivation. These three variables were defined through factor analysis, which is thoroughly explained in the methodology section (see Section 6.5.4.4). In Table 29 we report the descriptive statistics of the three statements (difficulty, self-confidence and motivation) for the three tasks.

Table 29. *Descriptive statistics: affective variable questionnaire*

	Task 1 Simple task	Task 2 More complex task	Task 3 Most complex task
Difficulty	4.85 (1.70)*	4.67 (2.30)	6.94 (1.45)
Confidence	4.51 (2.13)	4.85 (2.33)	4.03 (2.61)
Motivation	6.51 (1.96)	6.59 (1.80)	6.05 (1.84)

**Mean (SD)*

First, repeated-measure ANOVA was performed to explore overall differences in the affective perception of the tasks through three levels of cognitive complexity. A significant main effect was found for difficulty ($F=30.52$, $p = .001$). No other significantly different effect for confidence or motivation was found (see Table 30).

Table 30. *Affective variable questionnaire: within-subject effects*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Difficulty	101.543	2	50.771	30.519	.001	.601
Confidence	6.533	2	3.267	2.093	.131	.010
Motivation	3.790	2	1.895	2.437	.095	.056

Next, pairwise comparisons were performed to analyze any significant difference among the three tasks for the "difficulty" statement. The results showed a significant difference between Task 1 and Task 3 ($p = .001$) and Task 2 and Task 3 ($p = .001$). In addition, a significant difference was found between Task 2 and Task 3 for confidence ($p = .010$) and motivation ($p = .033$) – see Table 31. This confirms that predicted task complexity was matched by the perception of difficulty, with the more complex task being perceived as more difficult.

Table 31. *Affective variable questionnaire: pairwise comparisons*

Measure	Task1 – Task2	Task1 – Task3	Task2 – Task3
Difficulty	.560	.001	.001
Confidence	.261	.291	.010
Motivation	.520	.087	.033

7.1.2. Time estimation task

As described in the methodology section, participants were asked to estimate how much time they thought they had spent to complete each of the three tasks. Their time estimation together with the real time of task performance was recorded. Subsequently, time difference was calculated by subtracting the real estimated time. In Table 32 below, the descriptive statistics (mean and SD) for estimated time, real time and time difference (Task1 – Task 2 – Task 3) are provided.

Table 32. *Descriptive statistics: time estimation task*

	Task 1	Task 2	Task 3
Estimated Time	128.72 (80.92)*	126.84 (74.77)	135.00 (57.07)
Real Time	98.41 (27.23)	106.36 (34.13)	139.26 (54.81)
Time Difference	29.97 (83.72)	20.71 (69.25)	-6.38 (94.00)

**Mean (SD)*

Repeated-measure ANOVA reported a main significant effect only in the case of "real time" values (p = .001), whereas no effect was found for "estimated time" or "time difference" (see Table 33). Pairwise comparisons revealed a significant difference between Time 2 and Time 3 (p = .001) and Time 1 and Time 3 (p = .001) for "real time", whereas no main effect was found for "estimated time" or "time difference" (see Table 34).

Table 33. *Time estimation task: within-subject effects*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Estimated Time	2230.808	2	1115.404	.394	.676	.012
Real Time	22770.081	2	11385.040	12.667	.001	.284
Time Difference	12097.051	2	6048.525	1.500	.231	.045

Table 34. *Time estimation task: pairwise comparisons*

Measure	Task1 – Task2	Task1 – Task3	Task2 – Task3
Estimated Time	.774	.558	.368
Real Time	.124	.001*	.001*
Time Difference	.752	.345	.234

7.2 Research question on L2 development

The research question related to L2 learning during treatment through task sequencing was as follows: *Is there any effect of task sequencing from simple to complex on L2 development of spatial expressions? If so, what is the effect of task sequencing on L2 development of spatial expressions?*

The results on the effects of task sequencing on L2 development of spatial expressions are presented separately for laboratory (see Section 7.2.1) and classroom settings (see Section 7.2.2), where the results from the latter one are used in attempt to confirm the results of a much more controlled laboratory group. To address the question of which sequence is more beneficial for L2 development, a one-way ANOVA was performed for the results of immediate posttest and delayed posttest.

7.2.1 Laboratory context

The results for the laboratory context are first presented for the descriptive task (see Section 7.2.1.1) and then for the two-part vocabulary test (see Section 7.2.1.2). In each section descriptive statistics are followed by inferential statistics.

7.2.1.1 Descriptive task

7.2.1.1.1 L2 development over time – overall results

7.2.1.1.1.1 Descriptive statistics

The mean scores and the standard deviations for the descriptive task (three general scores and two frequency scores) are given in Table 35 and they are visually displayed in Figure 15 and Figure 16.

Table 35. *Descriptive statistics: descriptive task - laboratory groups*

	Pretest	Immediate Posttest	Delayed Posttest
All Correct Expressions - General Score	7.05* (2.92)	9.67 (2.66)	8.95 (1.78)
All Targets - General Score	2.46 (1.54)	5.62 (1.93)	4.78 (1.77)
Correct Targets - General Score	1.62 (.99)	3.87 (1.47)	3.57 (.261)
All Targets - Frequency	2.03 (1.11)	3.87 (1.49)	3.62 (1.01)
Correct Targets - Frequency	1.49 (.85)	2.92 (1.27)	2.76 (.96)

**Mean (SD)*

Figure 15. *Mean scores for laboratory groups (descriptive task – general score)*

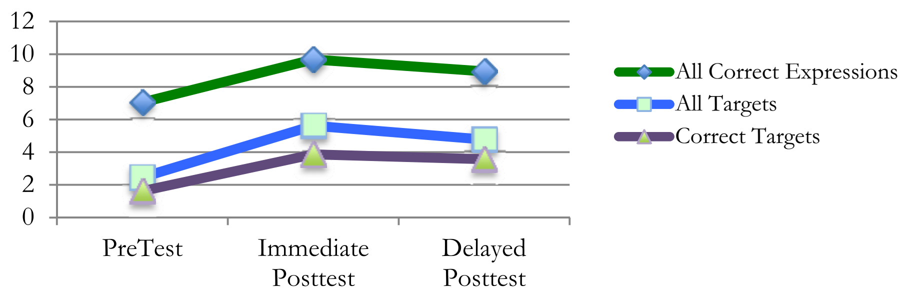
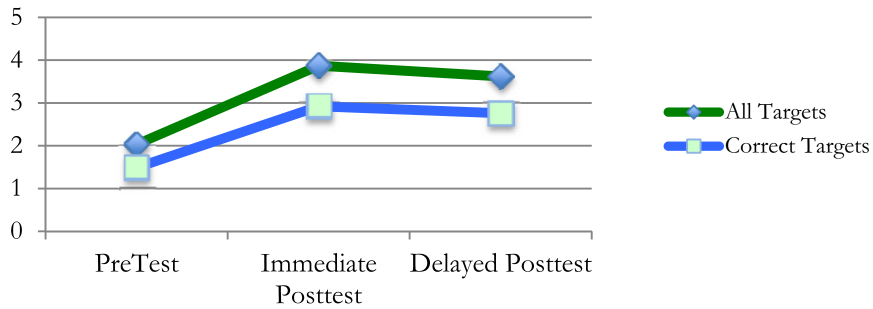


Figure 16. *Mean scores for laboratory groups (descriptive task – frequency score)*



As shown in Figure 15 and Figure 16 above, laboratory groups had a significant improvement as measured by all correct expressions, all targets and correct targets at posttest, which slightly decreased over time; however students still scored higher on the delayed posttest, than on the pretest. Overall, laboratory groups followed the predictable trajectory for both measures (all targets and correct targets). The participants scored higher on the immediate posttest and the delayed posttest than on the pretest, where the delayed posttest results were lower compared to the immediate posttest results.

7.2.1.1.1.2 Inferential statistics

A repeated-measure ANOVA was computed and displayed a significant main effect for "all correct expressions" ($F = 17.92$, $p = .001$, partial $\eta^2 = .33$), "all targets" ($F = 36.61$, $p = .001$, partial η^2 .50), "correct targets" ($F = 34.76$, $p = .001$, partial $\eta^2 = .49$), "frequency – all targets" ($F = 70.92$, $p = .001$, partial $\eta^2 = .66$) and "frequency – correct targets" ($F = 8.30$, $p = .001$, partial $\eta^2 = .19$). The ANOVAs of the descriptive task for laboratory groups are reported in Table 36 below:

Table 36. *Within-subject effects: descriptive task – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	152.000	2	76.000	17.921	.001*	.332
All Targets	200.342	2	100.171	36.612	.001*	.504
Correct Targets	116.919	2	58.459	34.763	.001*	.491
Frequency – All targets	120.721	2	60.360	70.920	.001*	.663
Frequency – Correct targets	14.991	2	7.495	8.302	.001*	.187

Given the significant differences between different times for all the measures of the descriptive task pairwise comparisons were calculated. The results showed a significant difference for all measures between pretest (Time 1) – posttest (Time 2) and pretest (Time 1) – delayed posttest (Time 3), the statistical difference being p = .001. However, only two measures ("all correct expressions" and "all targets") reached a significant difference in the comparison at Time 2 – Time 3 (p = .031, p = .030, respectively) – see Table 37.

Therefore, it can be stated that in the descriptive task carried out by the laboratory groups, learners used significantly more "correct expressions", as well as more "targets" and "correct targets", and they used them more frequently. This was true for the comparison between the pretest, the posttest and the delayed posttest. Interestingly, the results from the posttest and the delayed posttest were virtually indistinguishable in terms of the use of correct targets and their frequency.

Table 37. *Pairwise comparisons: descriptive task - laboratory groups*

Measure	Time1 – Time2	Time1 – Time3	Time2 – Time3
All Correct Expressions	.001*	.001*	.031*
All Targets	.001*	.001*	.030*
Correct Targets	.001*	.001*	.242
Frequency – All targets	.001*	.001*	.303
Frequency – Correct targets	.001*	.001*	.431

The results showed that participants significantly improved at posttest along all the measures of the descriptive task and that they maintained their acquired knowledge with no significant difference in the case of correct targets and frequency – all targets.

7.2.1.1.2 L2 development over time – two sequences

First, a one–way ANOVA analysis was used to check group comparability on all the measures for the descriptive task. The results did not reveal any significant difference between the two groups: "all correct expressions" ($F(3,35) = 2.69$; $p = .110$), "all targets" ($F(3,35) = 2.66$; $p = .112$), "correct targets" ($F(3,35) = 1.75$; $p = .194$), "frequency – all targets" ($F(3,35) = 3.65$; $p = .064$) and "frequency – correct targets" ($F(3,35) = 2.54$; $p = .120$), which means that any significant differences in the results of the immediate posttest and the delayed posttest compared to the pretest between the two sequence groups after treatment could be attributed to the treatment effects.

7.2.1.1.2.1 Descriptive statistics

The following are descriptive statistics for all the measures of the descriptive task (general score) in Table 38, descriptive task (frequencies) in Table 39, descriptive tasks (gains) in Table 40 and five Figures (one per measure for general scores and for frequencies) graphically displaying the results.

Table 38. *Descriptive statistics: descriptive task (general score) - two sequences, laboratory context*

	Pretest			Immediate Posttest			Delayed Posttest		
	All Correct Expressions	All Targets	Correct Targets	All Correct Exs.	All Targets	Correct Targets	All Correct Exs.	All Targets	Correct Targets
Group 1	6.25*	1.95	1.35	9.20	5.55	3.60	9.05	5.20	3.85
(L_SC**)	(2.88)	(1.36)	(.99)	(2.91)	(1.85)	(1.47)	(1.70)	(1.70)	(1.53)
Group 2	7.82	2.65	1.76	10.06	5.29	4.18	8.80	4.20	3.33
(L_R)	(2.94)	(1.22)	(.90)	(2.44)	(1.65)	(1.51)	(2.04)	(1.86)	(1.72)

**Mean (SD)*
*** L_SC: laboratory group carrying tasks in the simple-to-complex sequence*
L_R: laboratory groups carrying tasks in the randomized sequence

Figure 17. *Mean scores for two sequences – laboratory context (descriptive task – all correct expressions)*

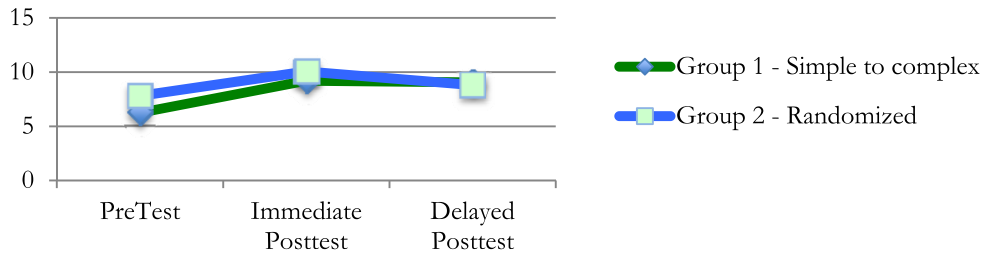


Figure 18. *Mean scores for two sequences – laboratory context (descriptive task – all targets)*

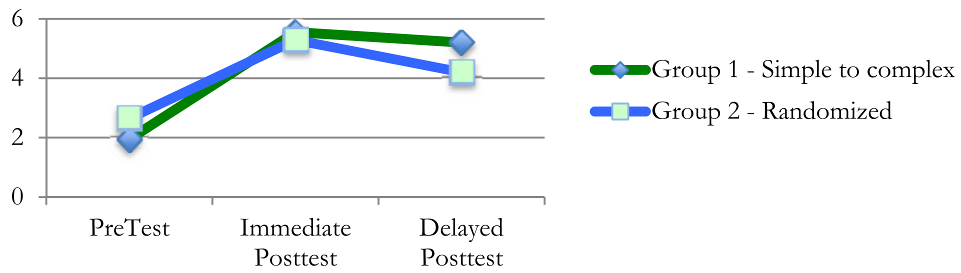
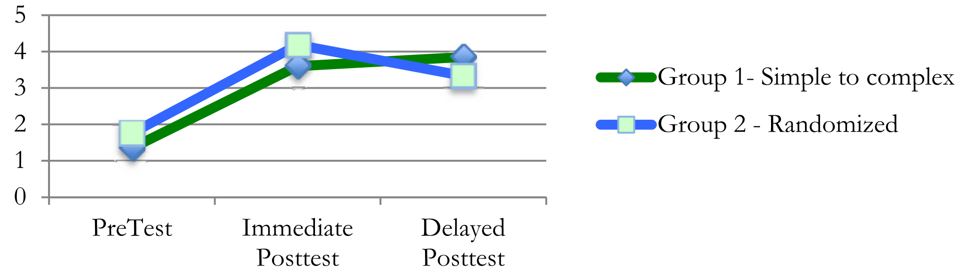


Figure 19. *Mean scores for two sequences – laboratory context (descriptive task – all correct targets)*



As observed in Figures 17, 18, and 19 above, Group 1 (simple to complex), while displaying scores equally higher on the immediate posttest as Group 2 (randomized), compared to the pretest, seems to retain the newly learned target items on the delayed posttest with more success than Group 2. The difference in decrease is especially notable in the case of the following measures: "all targets" and "correct targets" measures. The same behaviour is observed in the case of the frequency results (see Figures 20, 21). Group 1 scores higher on the delayed posttest than Group 2, particularly for "correct targets".

Table 39. *Descriptive statistics: descriptive task (frequency score) - two sequences, laboratory context*

	Pretest		**Immediate Posttest**		**Delayed Posttest**	
	All Targets	Correct Targets	All Targets	Correct Targets	All Targets	Correct Targets
Group 1 (L_SC**)	1.60* (.995)	1.10 (.718)	3.60 (1.39)	2.60 (1.14)	3.70 (.979)	2.95 (.945)
Group 2 (L_R)	2.18 (.809)	1.65 (.61)	3.88 (1.41)	3.18 (1.33)	3.60 (1.06)	2.60 (.910)

* *Mean (SD)*
*** L_SC: laboratory group performing tasks in the simple-to-complex sequence*
L_R: laboratory groups perfomirng tasks in the randomized sequence

ve task (gains) - two sequences, laboratory context

Pretest - Immediate Posttest					Pretest - Delayed Posttest			
Targets	Correct Targets	Frequency All	Frequency Correct	All Correct Expressions	All Targets	Correct Targets	Frequency All	Frequency Correct
0 (2.46)	2.25 (1.77)	2.00 (1.75)	1.50 (1.10)	2.80 (2.78)	3.25 (2.51)	2.50 (2.01)	2.10 (1.45)	1.85 (1.23)
5 (1.41)	2.41 (1.28)	1.71 (1.31)	1.53 (1.46)	.87 (2.90)	1.47 (2.13)	1.53 (1.36)	1.40 (1.18)	.93 (.88)

imple-to-complex sequence
ndomized sequence

Figure 20. *Mean scores for laboratory groups (descriptive task – frequency score – all targets)*

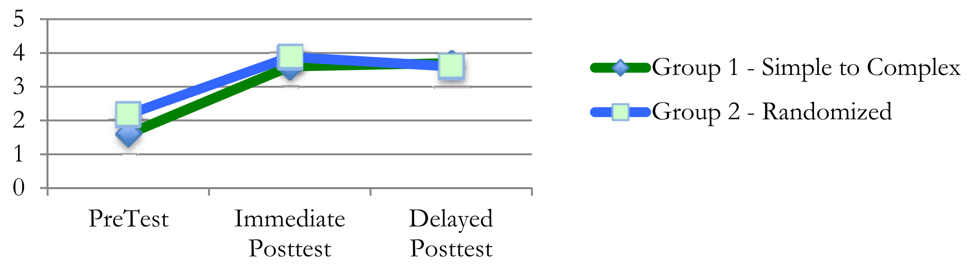
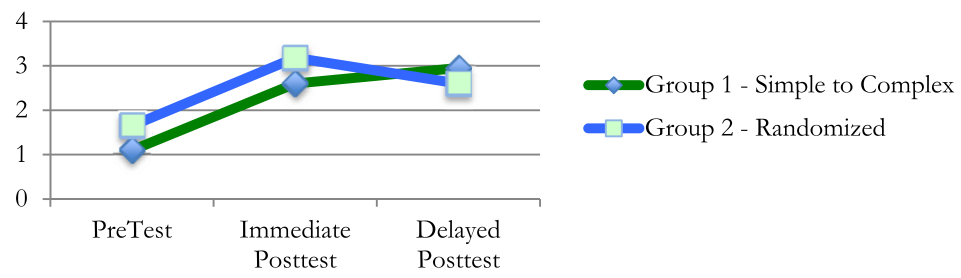


Figure 21. *Mean scores for laboratory groups (descriptive task – frequency score – correct targets)*



Results in Table 40 display gains immediately after treatment for all the measures of the descriptive task and retention of correct forms and target items two weeks after treatment.

7.2.1.1.2.2 Inferential statistics

Next, one-way ANOVAs were calculated on the mean scores of the immediate posttest as well as the delayed posttest with all the measures of the descriptive task and sequence as the between-subject factor. The results are reported in Table 41 and Table 42 for general scores and frequencies and in Table 43 and Table 44 for gains.

As displayed in Tables 41-44, there was a significant difference between the two groups only for "all targets" (general score) and "correct targets" (frequency) when looking at the gains between pretest and delayed posttest, which means that Group1 (simple to complex) maintained the acquired knowledge of the target items significantly better than Group 2 (randomized) in the case of both the "all targets" measure (p = .001) and the "frequency – correct targets" measure (p = .001).

Table 41. *One-way ANOVA: descriptive task (immediate posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.[36]	η^2
All Correct Expressions	6.778	1	6.778	.926	.342	.03
All Targets	.602	1	.602	.194	.662	.01
Correct Targets	3.054	1	3.054	1.383	.247	.04
Frequency – All targets	.733	1	.733	.374	.545	.04
Frequency – Correct targets	3.054	1	3.054	2.006	.165	.07

Table 42. *One-way ANOVA: descriptive task (delayed posttest) – laboratory* groups

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	.536	1	.536	.156	.695	.01
All Targets	8.571	1	8.571	2.730	.108	.06
Correct Targets	2.288	1	2.288	.879	.355	.04
Frequency – All targets	.086	1	.086	.084	.774	.01
Frequency – Correct targets	1.050	1	1.050	1.214	.279	.05

[36] Due to the multi-measuring of the descriptive task Bonferroni correction to the p-value was applied.

Table 43. *One-way ANOVA: descriptive task (gains: pretest - immediate posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	4.694	1	4.694	.399	.532	.01
All Targets	8.345	1	8.345	1.991	.167	.05
Correct Targets	.240	1	.240	.098	.756	.01
Frequency – All targets	.795	1	.795	.325	.572	.01
Frequency – Correct targets	.008	1	.008	.005	.945	.01

Table 44. *One-way ANOVA: descriptive task (gains: pretest - delayed posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	32.038	1	32.038	3.991	.054	.11
All Targets	27.260	1	27.260	4.903	.001*	.13
Correct Targets	8.010	1	8.010	2.573	.118	.07
Frequency – All targets	4.200	1	4.200	2.333	.136	.06
Frequency – Correct targets	7.202	1	7.202	6.020	.001*	.15

7.2.1.2 Two-part vocabulary test

The two-part vocabulary test consisted of a translation of 12 items from Spanish into English (9 target items and 3 distractors) and of a multiple-choice test also in relation to 12 items (9 target items and 3 distractors). Below, the results for both laboratory and classroom settings are presented.

7.2.1.2.1 L2 development over time – overall results

7.2.1.2.1.1 Descriptive statistics

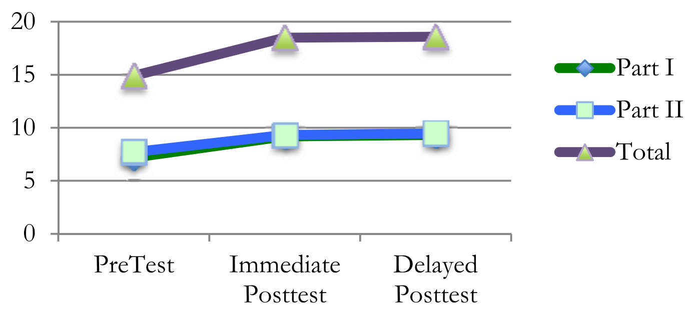
The mean scores and the standard deviations for the two-part vocabulary test are given in Table 45 (general score and frequency score) and are visually displayed in Figure 22.

Table 45. *Descriptive statistics: two-part vocabulary test - laboratory context*

	Pretest	Immediate Posttest	Delayed Posttest
Part I	8.66 (1.21)*	9.67 (1.01)	10.00 (1.05)
Part II	9.00 (1.48)	9.61 (1.48)	9.73 (1.02)
Total	17.66 (2.16)	19.28 (2.09)	19.73 (1.53)

**Mean (SD)*

Figure 22. *Mean scores for laboratory groups (two-part vocabulary test)*



As seen from Figure 22, the results obtained in the immediate posttest are higher than in the pretest and, moreover, they seem to be stable, as they maintain almost at the same level in the delayed posttest.

7.2.1.2.1.2 Inferential statistics

A repeated-measure ANOVA was computed and it displayed a significant main effect for part I ($F = 12.18$, $p = .001$, partial $\eta^2 = .30$), part II ($F = 7.76$, $p = .001$, partial η^2 .22), total ($F = 16.55$, $p = .001$, partial $\eta^2 = .37$). The ANOVAs of the two-part vocabulary test for laboratory groups are reported in Table 46 below.

Table 46. *Within-subject effects: two-part vocabulary test - laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	27.885	2	13.943	12.178	.001*	.303
Part II	13.885	2	6.943	7.758	.001*	.217
Total	80.989	2	40.494	16.551	.001*	.372

After having obtained the significant differences over time for all the parts of the two-part vocabulary test, pairwise comparisons were computed. The results showed a significant difference for all the parts between Time 1 (pretest) – Time 2 (posttest) and Time 1 (pretest) – Time 3 (delayed posttest), the statistical difference being $p = .001$. However, none displayed a significant difference in the case of the comparison for Time 2 – Time 3 (see Table 47). That means that the students retained new target information immediately after the treatment and they did not forget a significant amount of it over time.

Table 47. *Pairwise comparisons: two-part vocabulary test - laboratory groups*

Measure	Time1 – Time2	Time1 – Time3	Time2 – Time3
Part I	.001*	.001*	.097
Part II	.025*	.001*	.371
Total	.001*	.001*	.119

7.2.1.2.2 L2 development over time – two sequences

One–way ANOVA was calculated to ensure group comparability for all the parts of the two-part vocabulary test. The results did not reveal any significant difference between the two groups: part I ($F(1,33) = .002$; $p = .969$), part II ($F(1,33) = 1.35$; $p = .253$), total ($F(1,33) = .658$; $p = .423$), which meant that any differences in the results of the immediate posttest and the delayed posttest compared to the pretest between the two sequence groups after treatment could be attributed to the treatment effects.

7.2.1.2.2.1 Descriptive statistics

What follows are descriptive statistics for all the measures of the two-part vocabulary test in Table 48 (general score) and in Table 49 (gains) and three Figures (one per part) graphically displaying the results (see Figures 23, 24, 25). Interestingly, in part II, Group 1 (from simple to complex) performed even better in the delayed posttest than Group 2 (randomized). Nevertheless, the overall results show that both sequences seem to be beneficial for long-term improvement.

Figure 23. *Mean scores for two sequences (two-part vocabulary test – part I)*

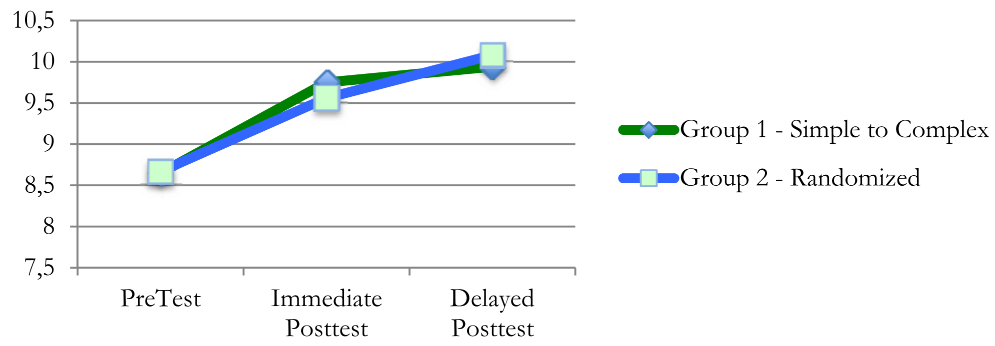


Figure 24. *Mean scores for two sequences (two-part vocabulary test – part II)*

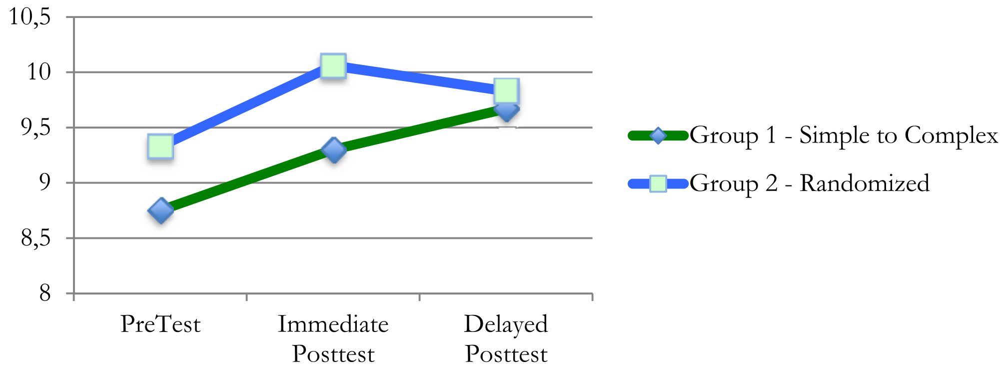
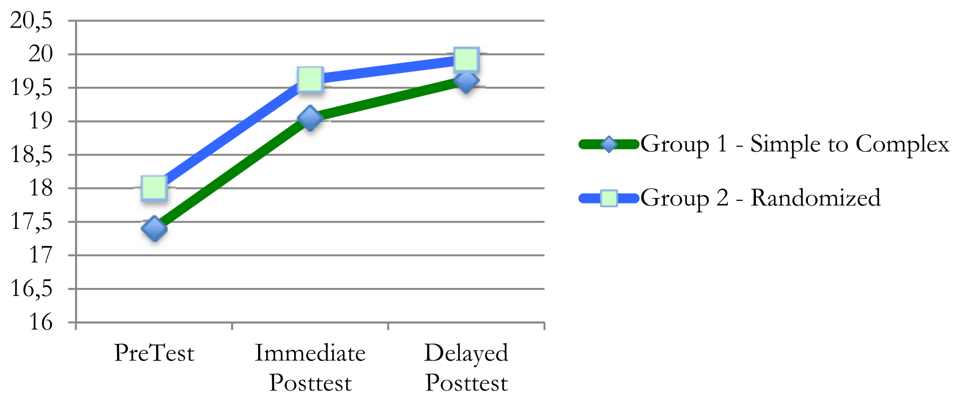


Figure 25. *Mean scores for two sequences (two-part vocabulary test – total)*



t vocabulary test - two sequences, laboratory context

etest		Immediate posttest			Delayed posttest		
rt II	Total	Part I	Part II	Total	Part I	Part II	Total
(1.52)	17.40 (2.35)	9.75 (1.02)	9.30 (19.05)	19.05 (1.11)	9.94 (1.11)	9.67 (1.03)	19.61 (1.58)
(1.40)	18.00 (1.89)	9.56 (1.03)	10.06 (1.44)	19.62 (2.13)	10.08 (.996)	9.83 (1.03)	19.92 (1.50)

nple-to-complex sequence
domized sequence

t vocabulary test (gains) - two sequences, laboratory context

	Pretest - Immediate Posttest			Pretest - Delayed Posttest		
	Part I	Part II	Total	Part I	Part II	Total
**)	1.10 (1.55)*	.55 (1.10)	1.65 (1.93)	1.22 (1.77)	1.06 (.94)	2.28 (2.35)
	.93 (1.22)	.67 (1.76)	1.60 (1.96)	1.55 (1.92)	.82 (1.54)	2.36 (2.84)

ry group performing tasks in the simple-to-complex sequence
groups performing tasks in the randomized sequence

7.2.1.2.2.2 Inferential statistics

Next, one-way ANOVAs were calculated on the mean scores of the immediate posttest as well as the delayed posttest with sequence as the between-subject factor. The results are reported in Table 50 (immediate posttest) and Table 51 (delayed posttest) - general score, and in Table 52 (pretest - immediate posttest) and Table 53 (pretest - delayed posttest) – gains, below.

Table 50. *One-way ANOVA: two-part vocabulary test (immediate posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	.313	1	.313	.298	.589	.009
Part II	5.168	1	5.168	2.698	.110	.07
Total	2.939	1	2.939	.691	.412	.02

Table 51. *One-way ANOVA: two-part vocabulary test (delayed posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	.139	1	.139	.122	.729	.004
Part II	.200	1	.200	.189	.667	.006
Total	.672	1	.672	.280	.601	.01

No significant differences were found between the two groups in the posttest and the delayed posttest, which suggests that long-term retention was equal.

Table 52. *One-way ANOVA: two-part vocabulary test (gains: pretest - immediate posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	.238	1	.238	.118	.734	.004
Part II	.117	1	.117	.058	.811	.002
Total	.021	1	.021	.006	.940	.001

Table 53. *One-way ANOVA: two-part vocabulary test (gains: pretest - delayed posttest) – laboratory groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	.713	1	.713	.214	.647	.008
Part II	.385	1	.385	.269	.608	.01
Total	.050	1	.050	.008	.930	.001

Again, no significant differences were found between the two groups in the gains observed in the posttest or the delayed posttest in comparison with the pretest. We now turn to the results of the classroom groups in order to confirm whether the same trends in the results hold outside the laboratory.

7.2.2 Classroom group

As seen earlier, the results for the classroom context are displayed for the descriptive task first (see Section 7.2.2.1) and then for the two-part vocabulary test (see Section 7.2.2.2). In each section descriptive statistics precede inferential statistics.

7.2.2.1 Descriptive task

7.2.2.1.1 L2 development over time – overall results

7.2.2.1.1.1 Descriptive statistics

The mean scores and the standard deviations for the descriptive task (three general scores and two frequency scores) are given in Table 54 and visually displayed in Figure 26 and Figure 27.

Table 54. *Descriptive statistics: descriptive task - classroom groups*

	Pretest	Immediate Posttest	Delayed Posttest
All Correct Expressions - General Score	4.32 (2.46)	9.09 (3.90)	6.86 (4.57)
All Targets - General Score	1.41 (1.37)	7.77 (2.91)	5.18 (3.53)
Correct Targets - General Score	.68 (.78)	4.77 (2.63)	3.09 (2.67)
All Targets - Frequency	1.18 (1.26)	5.09 (1.63)	4.72 (1.45)
Correct Targets - Frequency	.50 (.74)	2.55 (1.10)	3.00 (1.53)

**Mean (SD)*

Figure 26. *Mean scores for classroom groups (descriptive task – general score)*

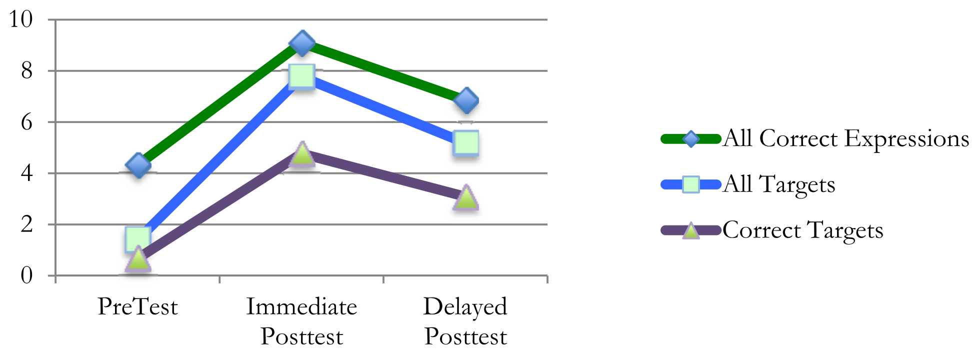
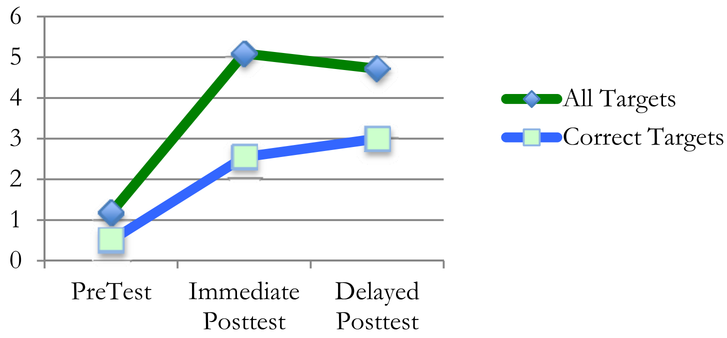


Figure 27. *Mean scores for classroom groups (descriptive task – frequency score)*



Overall results of the two classroom groups show that, as expected, the students obtained higher results on the immediate posttest for all the measures of the descriptive task. On the other hand, their scores decreased over time for all the measures, except for "correct targets – frequency".

7.2.2.1.1.2 Inferential statistics

A repeated-measure ANOVA was applied and it displayed a significant main effect for "all correct expressions" ($F = 21.94$, $p = .001$, partial $\eta^2 = .56$), "all targets" ($F = 75.35$, $p = .001$, partial η^2 .82), "correct targets" ($F = 25.64$, $p = .001$, partial $\eta^2 = .60$), "frequency – all targets" ($F = 52.28$, $p = .001$, partial $\eta^2 = .76$) and "frequency – correct targets" ($F = 26.42$, $p = .001$, partial $\eta^2 = .61$). The ANOVAs of the descriptive task for classroom groups are reported in Table 55 below:

Table 55. *Within-subject effects: descriptive task – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	267.815	2	133.907	21.939	.001*	.563
All Targets	427.000	2	213.500	75.353	.001*	.816
Correct Targets	174.778	2	87.389	25.639	.001*	.601
Frequency – All targets	169.037	2	84.519	52.283	.001*	.755
Frequency – Correct targets	60.037	2	30.019	26.421	.001*	.608

Given the significant differences over time for all the measures of the descriptive task, pairwise comparisons were computed. The results showed a significant difference for all the measures between pretest (Time 1) – posttest (Time 2) and Time 1 – Time 3 (delayed posttest), the statistical difference being p= .001. However, only three measures ("all correct expressions", "all targets", and "frequency – all targets") reached a significant difference for the comparison Time 2 – Time 3 (p = .009, p = .002, p = .001 respectfully) – see Table 56.

Table 56. *Pairwise comparisons: descriptive task - classroom groups*

Measure	Time1 – Time2	Time1 – Time3	Time2 – Time3
All Correct Expressions	.001*	.001*	.009*
All Targets	.001*	.001*	.002*
Correct Targets	.001*	.001*	.061
Frequency – All targets	.001*	.001*	.001*
Frequency – Correct targets	.001*	.001*	.177

The results showed that participants significantly improved at the posttest along all measures of the descriptive task and that they maintained their acquired knowledge in the case of “correct targets” and “frequency – correct targets”.

7.2.2.1.2 L2 development over time – two sequences

Again, one–way ANOVA was used to check group comparability on all the measures for the descriptive task. The results did not reveal any significant difference between the two groups: “all correct expressions” ($F(1,20) = .598$; $p = .448$), “all targets” ($F(1,20) = .210$; $p = .651$), “correct targets” ($F(1,20) = .662$; $p = .426$), “frequency – all targets” ($F(1,20) = .110$; $p = .744$) and “frequency – correct targets” ($F(1,20) = .738$; $p = .401$), which meant that any differences in the results of the immediate posttest and the delayed posttest compared to the pretest between the two sequence groups after treatment could be attributed to the treatment effects, and not to any initial differences.

7.2.2.1.2.1 Descriptive statistics

The following are descriptive statistics for all the measures of the descriptive task (general score - in Table 57, frequencies – in Table 58, and gains in Table 59) and five figures (one per measure) displaying the results graphically (see Figures 28 – 32).

Figure 28. *Mean scores for two sequences (descriptive task – all correct expressions)*

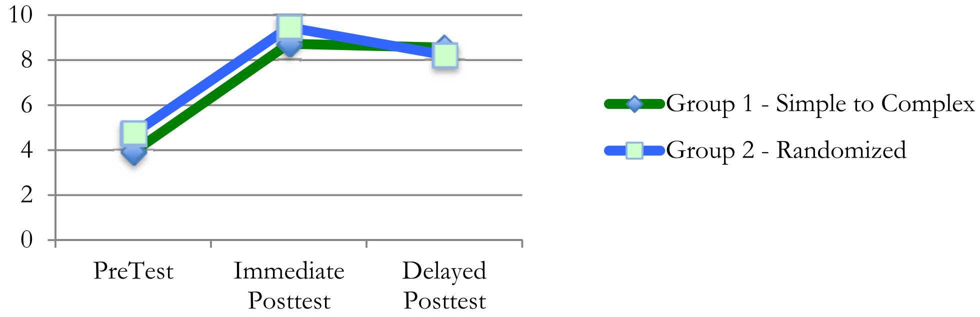


Figure 29. *Mean scores for two sequences (descriptive task – all targets)*

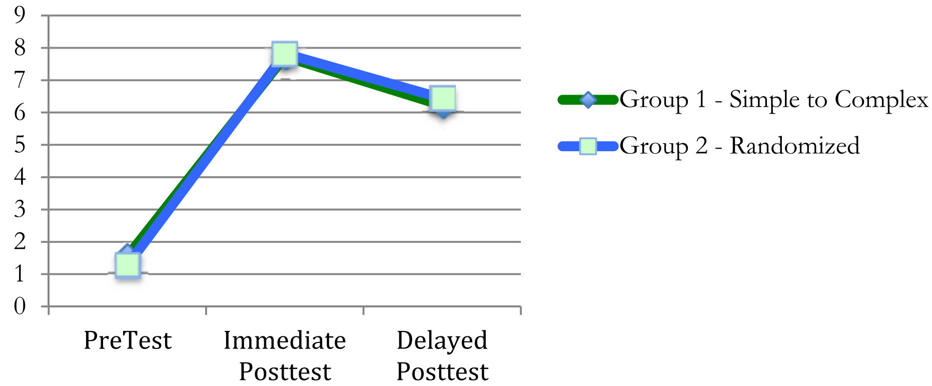
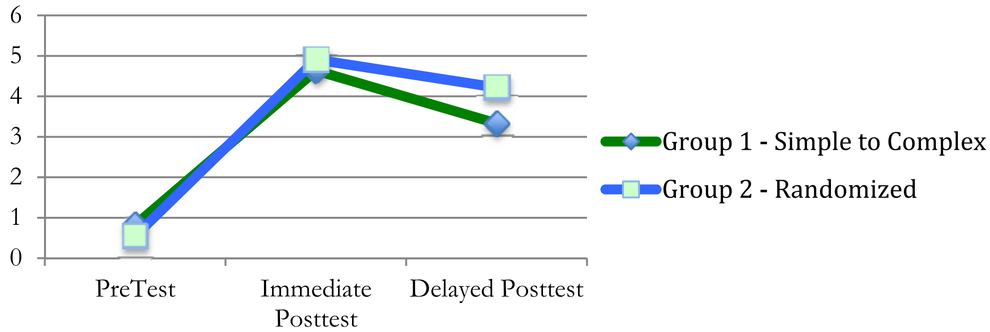


Figure 30. *Mean scores for two sequences (descriptive task – all correct targets)*



ve task (general score) - two sequences, classroom context

Pretest		**Immediate posttest**			**Delayed posttest**		
All Targets	Correct Targets	All Correct Expressions	All Targets	Correct Targets	All Correct Expressions	All Targets	Correct Targets
1.55 (1.29)	.82 (.75)	8.73 (3.32)	7.73 (2.61)	4.64 (2.45)	8.56 (3.28)	6.22 (2.44)	3.33 (1.94)
1.27 (1.49)	.55 (.820)	9.45 (4.55)	7.82 (3.31)	4.91 (3.08)	8.22 (3.90)	6.44 (3.21)	4.22 (2.50)

e simple-to-complex sequence
randomized sequence

The descriptive statistics showed that in the case of the classroom groups, both groups had an increase in the knowledge of target items at the posttest, whereas a slightly decrease is observed at the delayed posttest. Interestingly, Group 1 (simple to complex) scored higher in the delayed posttest for "all correct expressions", while Group 2 (randomized) appears to have retained more "correct target items" than Group 1.

Table 58. *Descriptive statistics: descriptive task (frequency score) - two sequences, classroom context*

	Pretest		**Immediate Posttest**		**Delayed Posttest**	
	All Targets	Correct Targets	All Targets	Correct Targets	All Targets	Correct Targets
Group 1 (L_SC**)	1.27* (1.49)	.64 (.809)	4.91 (1.14)	2.64 (1.21)	5.22 (1.20)	3.22 (1.56)
Group 2 (L_R)	1.09 (1.04)	.36 (.67)	5.27 (2.05)	2.45 (1.04)	4.22 (1.56)	2.78 (1.56)

** Mean (SD)*
*** L_SC: laboratory group performing tasks in the simple-to-complex sequence*
L_R: laboratory groups performing tasks in the randomized sequence

Figure 31. *Mean scores for classroom groups (descriptive task – frequency score – all targets)*

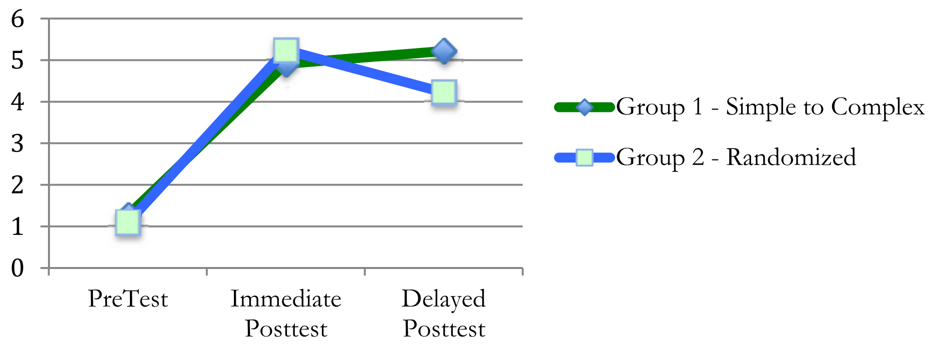
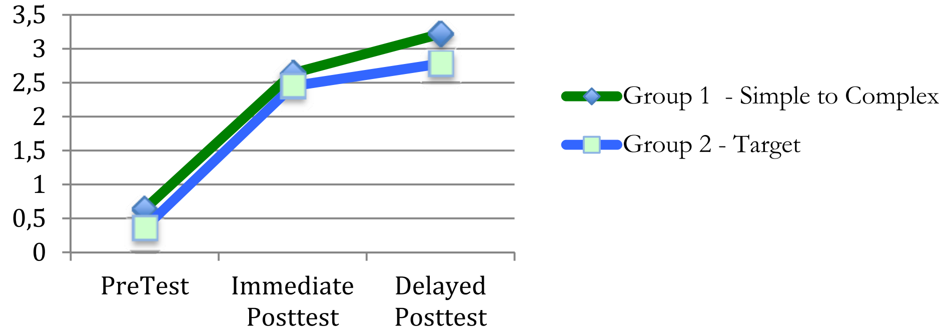


Figure 32. *Mean scores for classroom groups (descriptive task – frequency score – correct targets)*



The descriptive statistics for two frequency measures showed that for the "all targets" measure Group 1 (from simple to complex) scored better on the delayed posttest than on the immediate posttest, whereas Group 2 (randomized) scores actually decreased in the delayed posttest. At the same time both groups seemed to improve their knowledge of "correct target items" between the two times (immediate posttest and delayed posttest) as evidenced by the improved results of both groups in the delayed posttest in comparison to their results in the immediate posttest.

ve task (gains) - two sequences, classroom context

retest - Immediate Posttest					**Pretest - Delayed Posttest**			
Targets	Correct Targets	Frequency All	Frequency Correct	All Correct Expressions	All Targets	Correct Targets	Frequency All	Frequency Correct
(1.94)	3.82 (2.14)	3.64 (1.75)	2.00 (1.27)	4.56 (3.61)	4.56 (2.46)	2.44 (1.67)	3.89 (1.97)	2.56 (1.33)
(3.64)	4.36 (3.64)	4.18 (2.14)	2.09 (1.45)	3.11 (4.34)	5.11 (3.26)	3.67 (3.16)	3.00 (2.18)	2.33 (2.00)

ple-to-complex sequence
domized sequence

7.2.2.1.2.2 Inferential statistics

Next, one-way ANOVAs were run on the immediate posttest, as well as the delayed posttest with sequence as the between-subject factor. The results are reported in Table 60 and Table 61 (general score and frequencies) and in Table 62 and Table 63 (gains) below.

Table 60. *One-way ANOVA: descriptive task (immediate posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	2.909	1	2.909	.184	.673	.01
All Targets	.045	1	.045	.005	.944	.002
Correct Targets	.409	1	.409	.056	.815	.003
Frequency – All targets	.727	1	.727	.264	.613	.01
Frequency – Correct targets	.182	1	.182	.144	.708	.007

Table 61. *One-way ANOVA: descriptive task (delayed posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
All Correct Expressions	.500	1	.500	.039	.847	.002
All Targets	.222	1	.222	.027	.871	.002
Correct Targets	3.556	1	3.556	.571	.461	.03
Frequency – All targets	4.500	1	4.500	2.314	.148	.12
Frequency – Correct targets	.889	1	.889	.364	.555	.02

No significant difference was found for any of the five measures at the posttest as or for the delayed posttest. That means that the sequence from simple to complex does not explain longer-term retention.

Table 62. *One-way ANOVA: descriptive task (gains: pretest - immediate posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η²
All Correct Expressions	.045	1	.045	.003	.954	.001
All Targets	.727	1	.727	.085	.773	.004
Correct Targets	1.636	1	1.636	.215	.648	.01
Frequency – All targets	1.636	1	1.636	.430	.520	.02
Frequency – Correct targets	.045	1	.045	.025	.877	.001

Table 63. *One-way ANOVA: descriptive task (gains: pretest - delayed posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η²
All Correct Expressions	9.389	1	9.389	.589	.454	.03
All Targets	1.389	1	1.389	.167	.688	.01
Correct Targets	6.722	1	6.722	1.052	.320	.06
Frequency – All targets	3.556	1	3.556	.826	.377	.05
Frequency – Correct targets	.222	1	.222	.077	.785	.004

Again, no significant difference was found for any of the five measures as far as the overall gains in the posttest or delayed posttest are concerned.

7.2.2.2 Two-part vocabulary test

The results for the two-part vocabulary test are first presented for the two classroom groups together (see Section 7.2.2.2.1) and then split according to two sequences (see Section 7.2.2.2.2).

7.2.2.2.1 L2 development over time

7.2.2.2.1.1 Descriptive statistics

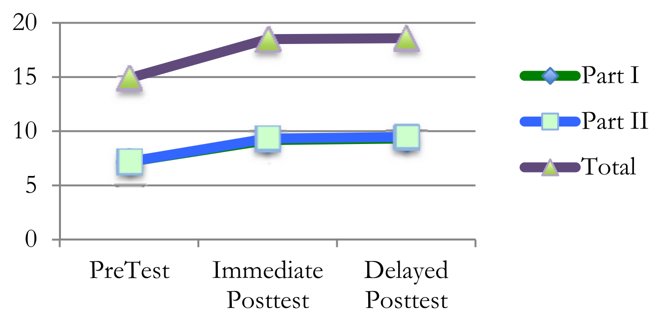
The mean scores and the standard deviations for the two-part vocabulary test are presented in Table 64 and are visually displayed in Figure 33.

Table 64. *Descriptive statistics: two-part vocabulary test - classroom context*

	Pretest	Immediate Posttest	Delayed Posttest
Part I	7.19 (1.81)	9.18 (1.71)	9.32 (.75)
Part II	7.19 (1.81)	9.32 (1.89)	9.47 (1.02)
Total	14.90 (2.97)	18.50 (3.17)	18.58 (1.54)

**Mean (SD)*

Figure 33. *Mean scores for classroom groups (two-part vocabulary test)*



As seen from Figure 33, the overall results showed that students achieved higher results on the two-part vocabulary test immediately after treatment and their results did not decrease over time.

7.2.2.2.1.2 Inferential statistics

A repeated-measure ANOVA was computed and displayed a significant main effect for part I ($F = 19.69$, $p = .001$, partial $\eta^2 = .43$), part II ($F = 16.74$, $p = .001$, partial η^2 .50), total ($F = 24.65$, $p = .001$, partial $\eta^2 = .59$). The ANOVAs of the two-part vocabulary test for classroom groups are reported in Table 65 below.

Table 65. *Within-subject effects: two-part vocabulary test - classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	39.370	2	19.685	12.880	.001*	.431
Part II	37.370	2	18.685	16.735	.001*	.496
Total	152.704	2	76.352	24.654	.001*	.592

Given the significant differences over time for all the parts of the two-part vocabulary test, pairwise comparisons were run. The results showed a significant difference for all the parts between pretest (Time 1) – posttest (Time 2) and pretest (Time 1) – delayed posttest (Time 3), the statistical difference being $p = .001$. However, none displayed a significant difference between Time 2 – Time 3 (see Table 66). This means that the students retained newly learned information over time.

Table 66. *Pairwise comparisons: two-part vocabulary test - classroom groups*

Measure	Time1 – Time2	Time1 – Time3	Time2 – Time3
Part I	.001*	.001*	.570
Part II	.002*	.001*	.076
Total	.001*	.001*	.139

7.2.2.2.2 L2 development over time - two sequences

One–way ANOVA was calculated to check group comparability on all the parts of the two-part receptive vocabulary test. The results did not reveal any significant difference between the two groups: part I ($F(1,20) = .598$; $p = .448$), part II ($F(1,20) = .210$; $p = .651$), total ($F(1,20) = .662$; $p = .426$), which meant that any differences in the results of the immediate posttest and the delayed posttest compared to the pretest between the two sequence groups after treatment could be attributed to treatment effects, and not to initial differences between the two groups.

7.2.2.2.2.1 Descriptive statistics

The following are descriptive statistics for all measures of the two-part vocabulary test in Table 67, in Table 68 (gains) and three Figures (one per part) graphically displaying the results (see Figures 34 – 36).

Figure 34. *Mean scores for two sequences (two-part vocabulary test – part I)*

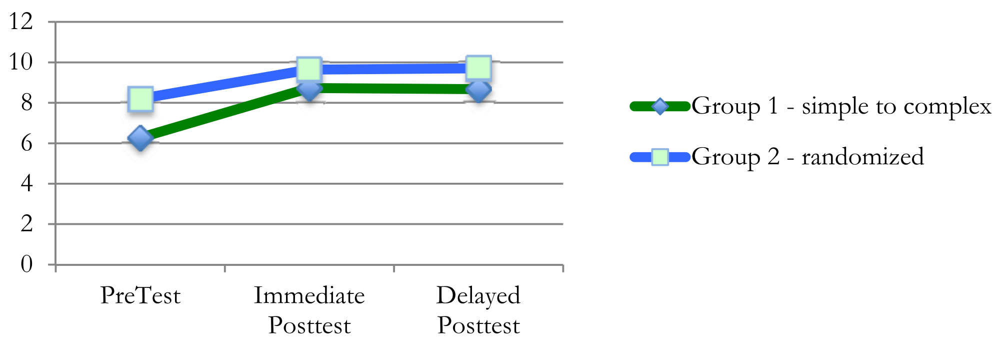


Figure 35. *Mean scores for two sequences (two-part vocabulary test – part II)*

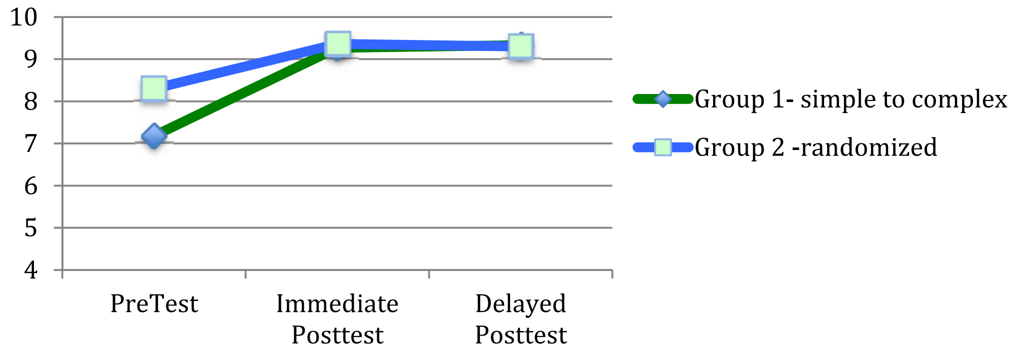
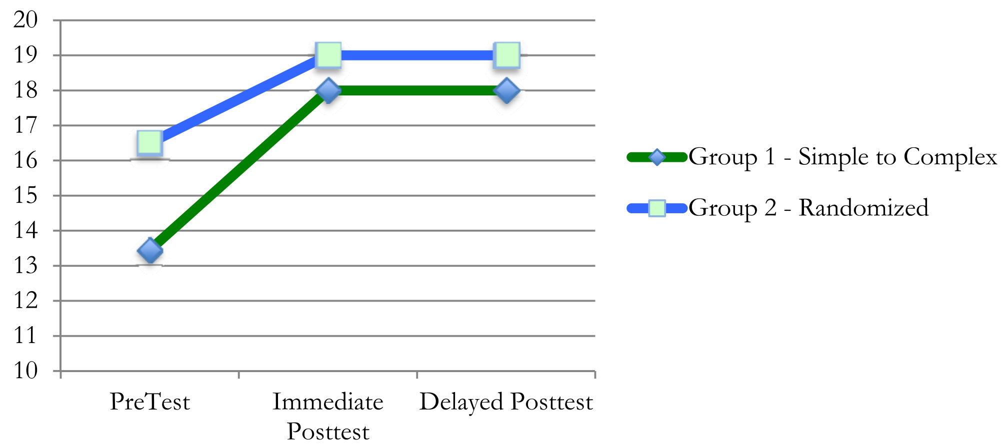


Figure 36. *Mean scores for two sequences (two-part vocabulary test – total)*



For both parts (part I and part II) and the total result, both groups behaved in a very similar way – they gained higher scores at the immediate posttest compared to the pretest and maintained their results at the delayed posttest.

-part vocabulary test - two sequences, classroom context

Pretest		Immediate Posttest			Delayed Posttest		
Part II	Total	Part I	Part II	Total	Part I	Part II	Total
7.18 (1.66)	13.45 (2.30)	8.73 (1.70)	9.27 (2.28)	18.00 (3.44)	8.67 (.50)	9.33 (1.00)	18.00 (.707)
8.30 (1.89)	16.50 (2.88)	9.64 (1.69)	9.36 (1.50)	19.00 (2.97)	9.70 (.675)	9.30 (1.16)	19.00 (1.49)

s in the simple-to-complex sequence
in the randomized sequence

Descriptive statistics: two-part vocabulary test (gains) - two sequences, classroom context

	Pretest - Immediate Posttest			Pretest - Delayed Posttest		
	Part I	Part II	Total	Part I	Part II	Total
_SC**)	2.45* (2.07)	2.09 (2.47)	4.55 (3.53)	2.11 (1.83)	2.22 (1.56)	4.33 (2.24)
_R)	1.3 (1.83)	.90 (.99)	2.20 (2.20)	1.33 (1.12)	.78 (1.20)	2.11 (1.54)

)
boratory group performing tasks in the simple-to-complex sequence
oratory groups performing tasks in the randomized sequence

7.2.2.2.2.2 Inferential statistics

Next, one-way ANOVAs were calculated on the immediate posttest, as well as the delayed posttest with sequence as the between-subjects factor. The results are reported in Table 69 and Table 70 (general score) and in Table 71 and Table 72 (gains) below.

Table 69. *One-way ANOVA: two-part vocabulary test (immediate posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	4.545	1	4.545	1.603	.220	.07
Part II	.045	1	.045	.012	.913	.001
Total	5.500	1	5.500	.534	.473	.03

Table 70. *One-way ANOVA: two-part vocabulary test (delayed posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	5.058	1	5.058	14.09	.002*	.45
Part II	.005	1	.005	.004	.948	.001
Total	4.737	1	4.737	3.355	.085	.16

No significant difference was observed between the two sequences when looking at the results of the immediate posttest, however, a significant difference was yielded for part I in the results of the delayed posttest for the two-part vocabulary test. And this means that, surprisingly, the classroom group with randomized sequence treatment retained more target items over time than the group with simple to complex sequence treatment.

Table 71. *One-way ANOVA: two-part vocabulary test (gains: pretest - immediate posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	6.982	1	6.982	1.822	.193	.09
Part II	7.429	1	7.429	2.022	.171	0.1
Total	28.816	1	28.816	3.253	.087	.15

Table 72. *One-way ANOVA: two-part vocabulary test (gains: pretest - delayed posttest) – classroom groups*

Measure	Sum of squares	df	MS	F	Sig.	η^2
Part I	2.722	1	2.722	1.181	.293	.07
Part II	9.389	1	9.389	4.829	.043*	.23
Total	22.222	1	22.222	6.038	.026*	.27

As far as the gains over time are concerned, a significant difference between groups was observed in the case of the delayed posttest compared to the pretest, where Group 1 (from simple to complex) scored higher (part II and total score), than Group 2 (randomized).

7.2.3 Summary of results

To sum up the results for the second research question, both laboratory and classroom groups learned significantly immediately after treatment and they retained correct target items over a two-week period of time both when measuring by means of the descriptive task and the two-part vocabulary test. As far as the differences in the retention of target information, in the case of the laboratory group, the sequence from simple to complex was statistically more beneficial over time when measured by the descriptive task, whereas in the case of the classroom group, sequence from simple to complex displayed significant results over time, when measured by the two-part vocabulary test.

7.3 Research question on WMC

Does Working Memory Capacity moderate L2 development as affected by task sequencing on the basis of task complexity?

The results of WMC calculated by means of the automated reading span test were generated by the programme at the end of the test. Two WM measures were computed: partial-credit unit scoring and all-or-nothing unit scoring. Here only partial-credit unit scoring is reported, as it is considered by several researchers to be an inclusive and robust scoring of the WMC (Conway et al, 2005). The subsequent analyses were performed by using the results of the two laboratory groups together with no distinction of the type of treatment they were assigned to and then by analysing the results of two laboratory groups separately according to the type sequence which was used in the treatment session (sequence 1: from simple to complex; sequence 2: randomized). The results reported below are given first for the descriptive task and then for the two-part vocabulary test.

7.3.1 Descriptive task

Pearson correlations were performed for the five measures of the descriptive task for the whole pool of participants (see Table 73 – general score and Table 75 – gains) and for the two sequence groups (see Table 74 – general score and Table 76 – gains). As for the general score, no significant correlation was found between WMC and any of the five measures of the descriptive task (see Table 73). The same results were obtained when looking at two sequence groups separately (see Table 74).

Table 73. *WMC as moderator of L2 development in descriptive task*

	Immediate Posttest	Delayed Posttest
All Correct Expressions	.042	.203
All Targets	.196	.244
Correct Targets	.099	.084
Frequency All Targets	.157	.094
Frequency Correct Targets	.021	.021

Table 74. *WMC as moderator of L2 development in descriptive task (two sequences)*

	WM Group	Immediate Posttest	Delayed Posttest
All Correct Expressions	**S-C**	.010	.256
	R	.199	.172
All Targets	**S-C**	.225	.196
	R	.142	.421
Correct Targets	**S-C**	.126	.006
	R	.026	.339
Frequency All Targets	**S-C**	.207	.006
	R	.002	.400
Frequency Correct Targets	**S-C**	.076	.078
	R	.209	.383

No significant correlation was found for a non-split WM group between the measures of the descriptive task and working memory capacity. However, looking at the results for the gains for the five measures of the descriptive task, several significant correlations were displayed for the non-split group of participants: “all targets” and “frequency all targets” in the difference pretest - posttest and “all targets” for pretest - delayed posttest pair (see Table 73). These results point toward a weak correlation between WMC and gains in the retention of all target items. In other words, learners with higher WMC retain more expressions and use them more frequently then lower WMC learners.

Table 75. *WMC as moderator of L2 development in descriptive task - gains*

	Gains pretest - posttest	**Gains pretest - delayed posttest**
All Correct Expressions	.096	.064
All Targets	.356*	.338*
Correct Targets	.235	.202
Frequency All Targets	.287*	.238
Frequency Correct Targets	.201	.182

When splitting the group into simple to complex sequence treatment groups and randomized sequence treatment groups, significant correlations at the immediate posttest were found for both sequences in the case of the following measures: simple to complex sequence – “all correct expressions”, “all targets”, “correct targets” and “frequency all targets”, whereas in the case of randomized sequence only one measure (“all correct expressions”) showed a significant correlation with WMC (see Table 76). As far as the delayed posttest is concerned, a significant correlation was found between WMC“ all targets” in the case of simple to complex sequence, and

WMC and "all correct expressions" and "frequency all targets" in the case of the randomized sequence.

Table 76. *WMC as moderator of L2 development in descriptive task - gains (two sequences)*

	WM Group	Gains pretest - posttest	Gains pretest - delayed posttest
All Correct Expressions	**S-C**	.317*	.245
	R	.434*	.469*
All Targets	**S-C**	.402*	.338*
	R	.213	.364
Correct Targets	**S-C**	.329*	.189
	R	.055	.245
Frequency All Targets	**S-C**	.332*	.191
	R	.099	.442*
Frequency Correct Targets	**S-C**	.309	.157
	R	.099	.292

From Table 76, it can be observed that working memory capacity seems to be more influential in L2 development when it occurs under a simple to complex task sequence. As for the randomized sequence, it is also affected by working memory, but only for two measures.

7.3.2 Two-part vocabulary test

As for the two-part vocabulary test, significant correlations were found in several instances between WMC and the three measures: part I, part II and total score. In the case of the descriptive task, the analyses were run first for both sequence groups together and then separately. Regarding the results of the non-split group, the statistical analysis displayed significant correlations between WMC and part II at the immediate posttest (see Table 77) and between WMC and part I (strong correlation) and total score (moderate correlation) at the delayed posttest. When looking at the results of the split group (see Table 78), in the case of the simple to complex sequence, the results of WMC highly correlated with part I at the immediate posttest and the delayed posttest along with total results of the delayed posttest. As for the randomized sequence, WMC correlated with all the measures at the immediate posttest and with part II at the delayed posttest. These results suggest that WMC seems to be important for the retention of new target expressions both immediately and over a period of time for the students from both groups.

Table 77. *WMC as moderator of L2 development in two-part vocabulary test*

	Immediate Posttest	Delayed Posttest
Part I	.139	.595**
Part II	.275*	.102
Total	.142	.325*

Table 78. *WMC as moderator of L2 development in two-part vocabulary test (two sequences)*

	WM Group	Immediate Posttest	Delayed Posttest
Part I	**S-C**	.452**	.801**
	R	.431*	.026
Part II	**S-C**	.224	-.029
	R	.494*	.413
Total	**S-C**	.049	.517**
	R	.510*	.258

Regarding the correlations of WMC and gains in the two-part vocabulary test, significant correlations were mostly displayed for the non-split group, where WMC significantly correlated with part I at the immediate gains and delayed gains and the total score correlated with WMC with the gains obtained in a two-week interval (see Table 79). When the groups were analyzed separately, only the results of WMC for the simple to complex sequence group correlated with the immediate gains for part I and with the gains over time for part I again and total score (see Table 80).

Table 79. *WMC as moderator of L2 development in two-part vocabulary test - gains*

	Gains **pretest vs. posttest**	**Gains** **pretest – delayed posttest**
Part I	.325*	.514**
Part II	-.083	.129
Total	-172	.423**

Table 80. *WMC as moderator of L2 development in two-part vocabulary test - gains (two sequences)*

	WM Group	**Immediate Posttest**	**Delayed Posttest**
Part I	**S-C**	.401*	.625**
	R	.021	.249
Part II	**S-C**	.050	.186
	R	.134	.036
Total	**S-C**	.278	.539**
	R	.130	.160

7.3.3 Further analyses

In order to further confirm that WMC played a mediating role in L2 development repeated-measures ANOVA with WMC as a covariate was performed. It is seen that statistically for all the measures of the descriptive task WMC was significantly involved in the process of L2 development of new target items as measured by partial-credit unit score (PCUS) in the case of "all correct expressions" F (2,52) = 10.05, p = .001; "all target expressions" F (2,52) = 5.21, p = .001; "correct target expressions" F (2,52) = 4.16, p = .001; "frequency all targets" F (2,52) = 2.02, p = .001 and "frequency of correct targets" F (2,52) = 1.88, p = .001 (see Table 81).

Table 81. *Repeated-measures ANOVA with partial-credit unit scoring (descriptive task)*

Effect	df	MS	F	p	Power
All correct expressions	2	104.16	58.73	.001*	1.00
All target expressions	2	103.91	127.98	.001*	1.00
Correct target expressions	2	65.75	210.76	.001*	1.00
Frequency of all targets	2	43.04	122.81	.001*	1.00
Frequency of correct targets	2	28.87	168.86	.001*	1.00
All correct expressions * PCUS	34	10.05	5.67	.001*	1.00
All targets * PCUS	34	5.21	6.41	.001*	1.00
Correct targets * PCUS	34	4.16	13.34	.001*	1.00
Frequency of all targets * PCUS	34	2.02	5.75	.001*	1.00
Frequency of correct targets * PCUS	34	1.88	10.99	.001*	1.00
Error(All correct expressions)	52	1.75			
Error(All target expressions)	52	.81			
Error(Correct target expressions)	52	.31			
Error(Frequency of all targets)	52	.35			
Error(Frequency of correct targets)	52	.171			

This suggests that learners with higher WMC retained more target expressions of space, remembered their correct form and used them more frequently than their counterparts with lower WMC.

The same pattern is observed when analyzing the results of the WMC in relation to the two-part vocabulary test, where WMC plays a significant role as measured by partial-credit unit scoring (see Table 82): Part I F (2,46) = 5.37, p = .001; Part II F (2,46) = 13.61, p = .001 and Total F (2,48) = 5.91, p = .001. The students with higher WMC conceptualized the new target items more efficiently in terms of their form and use.

Table 82. *Repeated-measures ANOVA with partial-credit unit scoring (two-part vocabulary test)*

Effect	df	MS	F	p	Power
Part I	2	8.21	19.54	.001*	1.00
Part II	2	12.65	93.52	.001*	1.00
Total	2	40.94	50.14	.001*	1.00
Part I * PCUS	32	2.26	5.37	.001*	1.00
Part II * PCUS	32	1.84	13.61	.001*	1.00
Total * PCUS	32	4.83	5.91	.001*	1.00
Error(Part I)	46	.42			
Error(Part II)	46	.135			
Error(Total)	46	.82			

7.4 Summary of the results

To sum up the results presented in this chapter, they will be given according to the research questions previously stated.

1) Does manipulating task complexity from simple to complex affect learners' reported perception of cognitive complexity?

The results showed that task complexity affected learner's perception of task difficulty (Task 1- Task 3; Task 2 – Task 3) as well as their confidence and their motivation (Task 2 - Task 3). As for the time estimation task, a significant difference was found for real time measure (Task 1 – Task 3; Task 2 – Task 3. The results of these two independent measurements suggest that more complex tasks were indeed perceived as more difficult and students had a harder time in keeping track of time for the more complex tasks.

2) *Is there any effect of task sequencing from simple to complex on L2 development of spatial expressions? If so, what is the effect of task sequencing on L2 development of spatial expressions?*

It was found that laboratory groups learned a significant number of target items immediately after treatment along the five measures of the descriptive task and they retained the learned items over time (with the exception of "all correct expressions" and "all targets"). The same pattern was detected with the two-part vocabulary test, where students showed significant gains on the immediate posttest and retained some learned information in all measures (part I, part II, total). When looking at the differences between the two sequences on the immediate posttest and the delayed posttest no significant differences were found between the sequences.

However, when looking at the results of the gains in the descriptive task over time (pretest - delayed posttest) a significant difference was found between the two sequences in two measures: “all targets” and “frequency - correct targets”. No significant differences were found between the two sequences with respect to the results of the two-part vocabulary test.

Similar results were found in the case of the classroom context. Both groups learned over time as measured by the descriptive task and the two-part vocabulary test. They also retained some target items over time in the following measures of the descriptive task: “correct targets” and “frequency – correct targets” and in all the measures of the two-part vocabulary test. As for the sequencing effect on L2 development, it was similar to the laboratory context, but this time only in the case of the two-part vocabulary test, the simple-to-complex sequence showed to play a beneficial role in L2 development over time as seen in the results of the delayed posttest (part I) and the gains (pretest – delayed posttest) part II, total. This suggests that the simple-to-complex sequence helped L2 learners retain target information over time as measured by the two-part vocabulary test.

3) *Does Working Memory Capacity moderate L2 development as affected by task sequencing on the basis of task complexity?*

L2 development as measured by the descriptive task, was moderated by WMC only in the case of the gains immediately after treatment (all targets and frequency all targets) and over time (all targets). When analysing the results of the two-part vocabulary test, WMC played a mediating role both in the general results of the immediate posttest (part II) and delayed posttest (part I and total score) and in the immediate gains (part I) and gains over time (part I and total score). Finally, when looking at the results of the two sequence groups separately, a mediating role was detected on various occasions for both groups immediately after treatment and two weeks later. These results suggest that WMC is both significantly helpful in the L2 development of spatial expressions for the two sequence groups together and WMC similarly beneficial for the students subjected to the different types of treatment (simple to complex sequence / randomized sequence).

CHAPTER VIII

DISCUSSION AND CONCLUSION

8.1 Introduction

This chapter is devoted to the interpretation of the results of the three research questions addressed in the present dissertation. It starts with some issues related to independent measurements of task complexity, followed by the discussion with respect to the role of task sequencing in L2 development in two settings (laboratory and classroom) and then proceeds to the analysis of the mediating role of WMC for L2 development of spatial expressions in the context of task sequencing. It ends with conclusions and suggestions for further research on the main issues of the present dissertation related to task sequencing, task complexity, laboratory versus classroom settings and WMC.

8.2 Independent measures of task complexity

The first research question related to the independent measures of task complexity was: *Does manipulating task complexity from simple to complex affect learners' reported perception of cognitive complexity?*

The results showed that task complexity as reported by the affective variable questionnaire was perceived by the participants in the same direction as operationalized from less difficult to more difficult. Moreover, the overall statistical difference for the "difficulty" variable was significant, which was confirmed subsequently by the pairwise comparisons among the tasks (Task 1 vs. Task 3 and Task 2 vs. Task 4). Finally, significant differences in their perception were found between Task 2 and Task 3 with regards to participants' confidence and perception of motivation.

Following the trend of the results from the third pilot, the present results confirmed that the operationalization of task complexity from simple to complex was carried out appropriately, which is important in order to follow with other research questions based on the assumption that tasks in the sequence are organized from simple to complex. Moreover, students also reported significantly less confidence and motivation when completing the third task (the most complex) which is predictable, as the complex task required much higher attention and memory resources and therefore it may have made them feel less confident. As far as motivation is concerned, it is possible that the students felt frustrated while completing the most complex task, which could have been the factor that lowered their motivation.

As for interpretation of time estimation results, it is important to distinguish between two time estimation paradigms - the prospective paradigm and the retrospective paradigm, as they indicate two opposite directions. In the case of the prospective paradigm the person is already aware of his or her task to evaluate the time he or she will spend to complete it; in the retrospective paradigm the task-doer is asked to complete the time estimation questionnaire after he is done with the task. After having conducted a meta-analysis of 117 studies on time estimation, Block et al. (2010:339) came to the conclusion that "in the prospective paradigm, the duration judgment ratio (i.e. the ratio of subjective duration to objective duration) decreases under high-load conditions relative to low-load conditions. In the retrospective paradigm, the duration judgment ratio increases under high-load conditions relative to low-load conditions". In the present study the prospective technique was used where students were subject to a treatment of three tasks. As predicted by Block et al. (2010) for the prospective technique, students were more precise at calculating the time they spent to finish the simple task than the most complex one, which was reflected in the "Time Difference" measure. So, again these results confirmed the right operationalization of task complexity in the present study. It should be stated that only the differences in "real time" between the three tasks were statistically significant, but as Révész et al. (2012) noted, out of three independent measures of task complexity (affective variable questionnaire, time estimation task and dual-task methodology), the affective variable questionnaire was shown to be the most robust measure and the one which best predicts the cognitive load of a task. On the other hand, both measurements can be considered as an indirect measurement of cognitive load. In this sense, it would be interesting to use stimulated recall in order to be able to gain an insight into what mental processes are affected by increased cognitive complexity. Another issue could be related to whether task difficulty perception was influenced by task sequence. For example, was the most complex task perceived as easier when it was the last task in a simple-to-complex sequence as opposed to the first one in the randomized sequence? These are some of the issues to be explored in the future. For the present purposes, after all these analyses it was seen that task complexity was perceived by the students in the same way it was conceived, therefore the task operationalization was valid for further statistical analyses on task sequencing based on task complexity manipulation.

8.3 Task sequencing and L2 development

The second research question was on the effects task sequencing may have on L2 development of the use of spatial expressions: *Is there any effect of task sequencing from simple to complex on L2 development of spatial expressions? If so, what is the effect of task sequencing on L2 development of spatial expressions?*

Two different measures of L2 development of spatial expressions were used (a descriptive task and a two-part vocabulary test), which were different in nature (production and translation, and recognition, respectively). The results are discussed separately for the two measurements.

8.3.1 Descriptive task

When analyzing the results, the focus was put on the three general and two frequency scores as well as and on the gains over time. First, results from both laboratory groups were analyzed (two sequences together) and it was found that participants produced the target items provided to them during the treatment as measured by both general and frequency scores. Moreover, they retained most of the target items they used in the immediate posttest over time, which was detected by the lack of significant differences in the results between the immediate posttest and the delayed posttest (except for the two measures – "all correct expressions" and "all targets"). This interesting finding could be explained by the fact that students became gradually aware of what the focus of the tasks and the treatment was and more importantly of the "mismatch or gap between what they could produce and what they needed to produce" (Schmidt, 2001:6) in order to complete the tasks. For task completion, students had to remember some spatial expressions and more specifically the spatial expressions they would be able to include in task completion. The process involved selection of information available to them from the input materials. At this stage, L2 development was already taking place. Additionally, the depth of L2 development of target expressions depended on the level of the previous knowledge of the students, where, according to Doughty and Williams (1998) L2 learners focus on form of some salient features of the L2, which are distinct from their L1. Later on they incorporate them into their L2 repertoire and finally end up gaining control of them. All the processes described above may have taken place in the L2 development of spatial expressions of the participants of the present study. Moreover, they were exposed to target spatial expressions by means of the input materials, where they had the opportunity to notice and process these expressions of space to be able to complete the tasks of the input (i.e. to follow the instructions and put some furniture items into an empty space). The students remembered, first of all, "simple", easy to understand and to remember (Robinson, 1996; Doughty & Williams, 1998) spatial target expressions ("*along something*", "*on the wall*" or "*near something*") and some of them could also remember some salient forms of the L2 ("*on the right-hand side*", "*on top of*"). The participants also had to retrieve some target items encountered in the input afterwards in order to do the tasks in the treatment sessions, which required utilizing memory and attentional processes.

Next, to answer the second part of the research question a comparative analysis of the two sequences (from simple to complex sequence and randomized sequence) was performed. The differences for the general and frequency scores between the two sequences did not display statistically significant results; however when looking in more detail at the descriptive statistics and the comparative graphs of the two contexts, it was noticed that the simple-to-complex group retained more spatial expressions in a two-week period than the randomized group whose results substantially decreased over time. This trend was confirmed statistically when a comparison of the two groups at the level of gains (pretest – delayed posttest) was run. Students from the simple-to-complex group retained more information over time compared to their counterparts from the randomized group along the two measures: "all targets" and "frequency – correct targets", while "all correct expressions" were at the limit of significance ($p = .54$). These results confirm Robinson's Cognition Hypothesis predictions on the benefits of sequencing tasks from cognitively simple to cognitively complex for L2 development of new linguistic items. By means of the progressive and efficient scheduling of cognitive resources, learners gradually increase their attentional and memory resources when moving from less cognitively demanding tasks to more cognitively demanding tasks, which allowed them, first, to stabilize and automatize their already existing repertoire, and next, to restructure and complexify their model of spatial representations for the given tasks by adding new target items. The simple-to-complex sequence was shown to be especially efficient in the long-term retention of information. It may be due to the fact that by efficiently scheduling attentional and memory resources over micro-processes such as input and focus on form (Doughty, 2001), more noticing and cognitive comparisons takes place and, as a consequence, better integration into long-term memory occurs. Another possible explanation is that upon becoming aware of a gap in their knowledge students tend to retain target information (Swain & Lapkin, 1995). This finding suggests that the gradual and coordinated activation of attentional resources is especially beneficial for further development of the L2 and longer retention of some target items acquired during the treatment. The findings are also in line with the study by Levkina and Gilabert[37] (in press), where the group with the treatment from simple to complex obtained higher learning benefits over time. On the other hand, even if statistical analysis showed significant gains for one group over the other in the delayed posttest, in the descriptive statistics it was clearly seen that the simple-to-complex group retained more target information immediately after treatment as well (see Table 35). However, with the relatively small number of the students involved in the experiment this statistically significant difference was captured only in the delayed posttest. To see how L2 development in the use of spatial expressions took place, below is the example of one of the participants' descriptive task written three times (pretest, immediate posttest, delayed posttest), where all target expressions are highlighted.

Pretest.

Hi Silvia! How are you? Here, everything is ok! You know? I've just visited Henry's new flat and I love the way it's been furnished, especially his living room! I'll describe it to you: it consists on a big white room with a grey carpet in the middle of it, where there are two black armchairs and a crystal table. On the table, there are a laptop and a set of three small white vases, you'd like them! In front of the table, there is a sofa and behind it you can find a wooden chair, a plant and two black-and-white pictures hanging **on the wall**. Next to the sofa, there are wooden stairs and there is a dining table with four white chairs. The stairs lead to the first floor, where the bedrooms are located. And in the centre of the living room handing from the ceiling there is a big red lamp that I'm sure you would like for your own flat. Well Silvia, I have to go, see you next week! Best wishes.

Total: 1 target

Immediate posttest.

Hi Marta! How are you doing? I send you this e-mail to tell you about my friend's furniture! It is so amazing that I want to furnish my flat as he has done it! I'll try to explain it: there is a rug in the middle of his living room, **on top of** it, he has a coffee table with a set of three small vases and a laptop. Around the table, on the rug, there are two black armchairs, one **opposite** the other. Next to the coffee table, **on the back** of the room, there is a sofa and the back wall; there are a plant (**on the left-side corner**), a wooden chair and two black-and-white pictures **above** them, hanging **on the wall**. Next to the sofa, **on the right-hand side of** the room, there are some stairs that separate the living room in two parts: the part that I have been describing and the right-hand part, in

[37] This study is the second pilot described in the present dissertation.

which there is a dining table with four white chairs. Ah! Almost forgot the red lamp hanging on the ceiling, **above** the coffee table, in the middle of the room. I want you to see all this, so maybe I'll ask my friend for some pictures of his flat. Have a nice week-end, Marta!

Total: 8 targets

Delayed posttest.

Hi Marta! How are you doing? As I told you the other day, I wanted to describe the furniture of Miguel's new living room. To begin with, there is a rug in the middle of the living room, and a crystal coffee table put on it. **On top of** the table, there are a set of three small vases and a laptop. And there are two black armchairs, one **opposite** the other, around the coffee table, on the rug as well. There is also a red lamp over the coffee able, hanging on the ceiling. **On the back**, there is a sofa, and, behind it. There are a plant, a wooden chair and two black-and-white pictures handing **on the back wall**, one next to the other. Next to the sofa, there are some stairs, and next to the stairs, there is a dining table with four white chairs. Isn't it an amazing living room? I'll send a picture to you as soon as I can! I have to go, I have an appointment with Marc, so I wish you a very nice weekend! See you!

Total: 4 targets

As observed in the pretest the student used only one target expression, as well as other expressions he/she knew which were not classified as new target items (e.g. "on", "behind"), which were confirmed during the piloting as being well-known by most students, whereas in the immediate posttest the use of target items increased to 8 items, and finally in the delayed posttest the number of items decreased to 4. Along with correctly used target items (one opposite the other), the student also made mistakes with some other expressions (e.g. on the left-side corner) but, on the whole, the example shows L2 development in the use of spatial expressions over time.

Looking at the cognitive processes that also took place in L2 development in the use of new target items, several conclusions may be drawn. Regarding spatial reasoning involved in the type of tasks used in the present study, people normally have to deal with one important issue: to analyze the frame of reference (Levinson, 2003) they use in their L1 and also to introduce some new conceptual linkage from the L2 that does not exist in their mother tongue. In the case of Spanish and Catalan, the two English prepositions "on" and "in" could be expressed with the same Spanish preposition "en", which makes the task linguistically more complex ("on the table / in the corner" versus "en la mesa/en el rincón"). Additionally and independently from the frames of reference, people find it more difficult to deal with a multi-spatial model than with a unitary-spatial model (Byrne & Johnson-Laird, 1989), which means that if the task involves more than one solution (Task 2 or Task 3 in the treatment where students had to decide on where to put the furniture items), it requires more attentional resources which, according to Robinson (2010), promotes better automatization and complexification of the language required by the task. This is what may have happened in the case of the task in the simple-to-complex sequence where students had to resort to the multi-spatial model of task representation and in doing this they were able to activate their cognitive mechanisms of noticing and cognitive comparison.

Another important issue is related to the fact that people tend to create their own single preferred model of spatial reasoning (Zachs & Tversky, 2005) and afterwards apply it to different similar problems they have to solve. In the case of the present study, in Task 3 due to its nature (perspective-taking task with the presence of objects' and space dimensions) students were forced to create a new model to be able to complete the task. Finally, the perspective-taking issue itself (Mainwaring et al., 2003) also had an impact on students' activation of attentional and memory resources, such as in Task 3 where they were forced to explain the placement of the furniture items from the perspective of a delivery man. This task required a shift from the egocentric frame of reference (which does not require much mental effort) to the addressee's frame of reference, which turned out to be a more cognitively demanding task while at the same time it also promoted the use of new target items. So, to sum up, with higher cognitive task demands first enhanced attention to input took place, followed by enhanced use of target items in the immediate posttest and finally better retention of target spatial expressions in the delayed posttest.

As in the case of the laboratory groups, overall results for the two classroom groups showed that all students significantly improved immediately after treatment, however they only retained correct target items over time as measured by "correct targets" and "the number of correct target items", whereas spatial expressions were not remembered correctly and other types were forgotten. When looking at the two groups separately, similarly to what was observed in the laboratory context, for the general scores and the frequency scores no significant differences were found. Furthermore, when looking at gains, no differences between the two sequences in the classroom context were found. This is an interesting finding, as it differs from what occurred in the laboratory setting. Several different explanations may be provided regarding (1) the non-confirmation of the predictions of the SSARC model and the Cognition Hypothesis on the beneficial role of task sequencing on L2 development and (2) on the differences between the two contexts (laboratory and classroom). First of all, it is important to remember that the two contexts were different in their level of proficiency in English, where the two laboratory groups scored statistically higher on all pretests than the two classroom groups. Although an earlier study by Kim (2008) on L2 acquisition of target words by the students with different levels of proficiency did not reveal any significant difference at initial learning and retention, in the present dissertation the students from both settings were not completely unfamiliar with the target expressions of space, which might have led to different paths of L2 development immediately and over time. It could be seen in the use of some spatial expressions, such as "on the right-hand side", where students were all familiar with the separate words which constitute the expression, however it was difficult for them to reproduce the whole expression correctly. The other example is "on top of" which was not a linguistically new item, but at the same time students had difficulties with using the correct preposition "on" and avoiding the use of the definite article in between (e.g. "*on* ***the*** *top of*").

Moreover, several studies were carried out regarding the generalizability of the results obtained from laboratory settings (e.g. Gass, Mackey & Ross-Feldman, 2005, 2011 on interaction), which did not find any significant difference between the laboratory and classroom settings. This finding could be also applicable to the present study, where the classroom results confirmed the laboratory results, however the different proficiency levels of the two settings should be still acknowledged.

8.3.2 Two-part vocabulary test

This measure consisted of two parts: the first part was a simple translation from Spanish into English and the second part was the multiple-choice test. The joint results of the two laboratory groups showed similar results as in the case of the descriptive task, where the students significantly improved their knowledge of spatial expressions immediately after treatment and they retained a significant part of the information two weeks after the treatment took place (see Table 46).

When comparing the results of the two laboratory groups, as opposed to what was found in the case of the descriptive task, no significant differences in the case of the two-part vocabulary test were detected. One possible explanation could lie in the nature of the task itself. First, the first part of the task was a translation, which is considered a good technique to measure the knowledge of new vocabulary. However, there is not always an exact and unique word in one language to express the same sense in another language, which makes the analysis more difficult. In the present study, there were some spatial expressions in Spanish, which being translated back into English contained more than one option, as in the case of "encima de la mesa", where the target item was "**on top of** the table", but it could be also translated into English as simply "**on** the table", which had to be considered as a correct answer but not a target one[38]. This fact may have reduced some possible differences between the two sequences. The second part of the task was a multiple-choice test, which is also considered to be a good tool for vocabulary recognition, however when completing this kind of tasks students sometimes use guessing, which is difficult to control for when codifying the data. All these points might have turned out to be influential in the final results of the differences between the two sequences. It should be admitted that in any case, both groups followed the predictable path of developing their knowledge of target spatial expressions where they scored significantly higher at the immediate posttest and forgot very little target information as shown by the delayed posttest.

In the analysis of the overall results from the two classroom groups, all the students significantly increased their knowledge of target expressions of space immediately after treatment and they retained most of the information over time without significant losses (see Table 50 and Table 51). When comparing the benefits in learning from two sequences separately at the immediate posttest and at the delayed posttest a significant difference was found between the sequence from simple to complex and the randomized sequence in the results of "part I" of the delayed

[38] Native speakers of the first pilot only used "on top of" and not "on".

posttest, where the group under alternative treatment (randomized sequence) completed the translation significantly better than the target group. However, the results of gains over time showed that the target group remembered more target expressions after a two-week break than the group in the randomized sequence. With these results, the Cognition Hypothesis regarding L2 development (2005, 2007a, 2010) is once again confirmed for the benefits from learning through task sequencing over time, as it happened with the laboratory context. When students found themselves in a cognitively more demanding context after dealing with simpler tasks, they efficiently allocated their attentional and memory resources which led to a better retrieval, practice and retention of new target forms. Additionally, learning may have taken place because learners may have noticed their gap in the knowledge of the spatial expressions they needed to complete the task during the pretest and the input sessions, so they "noticed" them in the input materials and subsequently used some of them in the treatment session. In the case of the simple to complex sequence, students were not cognitively overloaded by task complexity from the very beginning, but were gradually given more complex tasks to deal with. This fact allowed them to allocate their attention first to the adequate use of the new target items of space and after that get prepared to use them in a cognitively more demanding task. Therefore, the simple to complex sequence turned out to be more beneficial for L2 development of spatial expressions.

Apparently, in the present study the productive task was a better measure and indicator of L2 progress for laboratory groups where their overall L2 proficiency and experience with the use of English was significantly higher than in the case of the classroom groups. Translation and recognition of target items were better predictors of L2 development of the two classroom groups, where the participants were not as experienced and as proficient in English as their counterparts. The distinction between the two settings and the influence different types of assessment may have in the results on L2 development deserves more empirical studies.

8.4 The mediating role of WMC on L2 development

The third research question concerned the role that WMC plays as a mediator in L2 development of the use of spatial expressions in the laboratory setting: *Does working memory capacity moderate L2 development as affected by task sequencing on the basis of task complexity?*

The statistical analysis was performed first with regards to the overall scores of both sequence groups and WMC and then separately for each sequence. The results are discussed first in relation to the descriptive task and afterwards in relation to the two-part vocabulary test.

8.4.1 Descriptive task and WMC

In the analysis of all the students from the laboratory setting WMC moderately correlated with the gains only for one measure – "all targets" (see Table 75). However, in the statistical analysis of the two sequence groups separately, several moderate correlations were detected for both the simple-to-complex group and the group with randomized sequencing. Very interestingly, in the case of the group with the randomized task sequencing WMC only correlated with one measure "all correct expressions" in the gains immediately after treatment and two weeks after it, whereas in the case of the simple-to-complex sequencing moderate correlations of the immediate gains, as well as gains at the delayed posttest were observed with several measures of the descriptive task focused to specifically detect the retention of target expressions of space: "all targets", "correct targets", "frequency all targets". This finding confirms the predictions of the Cognition Hypothesis according to which individual differences such as WMC contribute to L2 production and L2 acquisition in general (e.g. Goo, 2012; Mackey et al., 2012; Révész, 2012), and it also highlights the fact that WMC mediates L2 development in a more significant way when the whole learning process is organized taking into account the cognitive and attentional resources of the students. In these kinds of treatment settings, both WMC and sequence were beneficial for the retention of new target items. In the case of the randomized sequence group, the students were at a disadvantage in terms of the optimal task sequencing conditions, however some of them could still retain a significant number of target expressions of space immediately after treatment as well as over time, which might be explained by their higher WMC which could distinguish them as far as the retention of target items is concerned, even under disadvantageous treatment conditions.

Another important issue related to the role of WMC in L2 development and L2 production is the fact that in order to notice, process and retain new target lemmas, students first have to conceptualize them and then be able to formulate some new spatial expressions (Levelt, 1989; Kormos, 2006). When it comes to cognitively more demanding conceptualization, statistically more attentional and memory resources are required from the students or the task-doers. In this context, WMC may be considered an important ability for L2 production and development. This was, in any case, observed in the present study.

8.4.2 Two-part vocabulary test and WMC

The results with respect to the two-part vocabulary test and WMC are even more revealing as far as L2 development along task sequencing is concerned. In the overall results for the two laboratory groups moderate correlations between WMC and the results of Part I and Total were already displayed. In the separate analysis of the two sequence groups it was clearly observed that WMC was equally decisive at the immediate posttest for the two groups; however WMC moderately correlated mostly with the results of the simple-to-complex group at the delayed posttest. As previously mentioned, when two conditions (high WMC and simple-to-complex sequence) are put together, attentional and memory resources are simultaneously and optimally involved into task completion, which leads to a greater amount of retained new target items over time. Students are first given simple tasks, so they can use their memory resources to retrieve spatial expressions they have become familiar with in the input sessions. Later on, they can move to a more cognitively demanding task, but as it requires moderate levels of attention and memory, students are still capable of dealing with both new target items and cognitively more demanding tasks. In the last, more demanding task, students have already practiced and automatized some of the target expressions of space; therefore, they can allocate all their attentional and memory resources to the completion of the most complex task without any detriment to formulation of new target items.

When looking at the correlation between WMC results and the gains immediately after treatment and two weeks later, moderate correlations are already displayed at the level of overall results, especially for the gains from treatment obtained over time. Once the pool of participants was split into two sequence groups, it became clear that the benefits from WMC are attributed to the simple-to-complex sequence group with strong correlations between WMC and the gains at the delayed posttest. These results again confirmed the finding in the case of the descriptive task and WMC where higher WMC, i.e. higher ability to process and to retain new information simultaneously along with performing tasks under the simple-to-complex sequence became crucial for obtaining higher results over time.

8.5 Conclusion

The main objective of the study was two-fold: first, to explore whether task sequencing from simple to complex based on the SSARC model elaborated by Robinson (2005, 2007a, 2010) turns out to be more beneficial for L2 learning and, if so, in what way. Second, it aimed to examine the mediating role of WMC on L2 development by means of task sequencing. From the data of the present dissertation, all research questions were answered, where several assumptions forwarded at the beginning were totally or partially confirmed and several factors related to the nature of participants and the types of assessment tasks to take into account were distinguished. After having carried out the whole experiment and analyzing the results, the following findings and subsequent conclusions have been drawn:

1. Robinson's SSARC model has been shown to be a feasible theoretical basis for empirical studies in L2 development, when looking at +/- spatial reasoning demands. For the present study it was crucial to have a clear framework for task classification (i.e. triadic componential framework) and a step-by-step model of task sequencing elaborated on the basis of cognitive load to design and carry out the experiment whose primary goal was to analyze L2 development of spatial

expressions. With the increasing interest in task sequencing within TBLT, it is especially important to have a testable and empirically valid model of task sequencing, and Robinsons' SSARC model offers a clear and concise model suitable for research in laboratory settings and for teaching in classroom, since it is becoming a common approach worldwide.

2. The results in this dissertation support the theory that task sequencing from cognitively simple to cognitively complex may be beneficial in classroom contexts and not just in laboratory contexts, as it was demonstrated in the present study where students from classroom setting successfully retained target expressions of space over time. However, it is important to highlight first the significant effect from task sequencing was detected over time (from the results of the delayed posttest), so from the present study no empirical evidence showed any immediate benefits from using treatments based on tasks sequenced from cognitively simple to cognitively complex. Benefits from task sequencing over time were not detected by the same measures either (the descriptive task and the two-part vocabulary test), but by the productive task in the case of the laboratory setting and by the translation and recognition task in the classroom setting, which also shows its possible liaison with the level of proficiency of the students. However the latter assumption should be further investigated in the future.

3. Results in this experiment demonstrated that working memory capacity might mediate L2 development over time. Following the predictions of the Cognition Hypothesis, a higher WMC turned out to be beneficial for the students who were subject to the simple-to-complex sequence. And as in the overall results, WMC played a mediating role in the retention of targeted expressions of space in a period of two weeks by the students.

8.6 Implications

The present research has various implications at the theoretical, methodological and pedagogical levels. To my knowledge, the research of this dissertation stands as one of the first studies aimed at testing Robinson's SSARC model on task sequencing. The findings of the study provide empirical evidence for the Cognition Hypothesis, regarding the following aspects: (1) it confirms that task sequencing on the basis of task complexity may be beneficial for L2 development under the conditions presented in this research; (2) L2 learners' individual differences in cognitive abilities (here, WMC) affect L2 development, with the beneficial effect gradually increasing along with an increase of the cognitive and attentional resources required to complete a task.

This research should hopefully stand as a useful methodological reference for future research as it contains a thorough analysis of how cognitive task complexity was defined over a series of pilots previous to the main study and also how target items were selected on the basis of a series of pilots. Furthermore, it gives a sample of a theoretical overview involving studies from different fields of science such as cognitive psychology, a field upon which initial assumptions of how task complexity should be operationalized were based. As for the assessment tools, it was noted that the results also depend on the nature of the task (productive task, translation or recognition), the choice of which can in the end be influential on the final results of a study.

Finally, the findings of the present research are also applicable for classrooms and for the design of syllabi based on tasks, as it provides the rationale and an empirical example of how tasks could be designed and organized in order to be included in the classroom setting. The study also makes some predictions about the effects that task sequencing may have on L2 development over time and overall it demonstrated that the theoretical model of task sequencing, the SSARC model, is feasible and applicable to the conditions of a real classroom as well as to laboratory settings. Undoubtedly, much more research is needed to further analyze different aspects of task sequencing and its effects on L2 development.

8.7 Limitations and further research

As any other study, the present dissertation also has a number of limitations. In the end, participants from the laboratory and the classroom study were not comparable as it was originally intended due to their differences in L2 proficiency, L2 knowledge and their overall linguistic background. Participants from the laboratory groups were students of English Studies, so they dealt with many aspects of the language as a subject all the time, whereas the classroom groups, although placed at a similar level of proficiency, did not have the same advantage as they came from different degrees unrelated to the linguistic sciences. In future research, it may be interesting to compare the two settings, with the students being very similar in all these contexts. On the other hand, it may be also interesting to specifically examine the effects L2 proficiency and study abroad, among others, may play in L2 development through task sequencing. As in many other studies, the number of students involved in the project also represents a limitation. A higher number of participants would be desirable to perform further statistical analyses for several aspects of the dataset. Finally, future research could also take into account a researcher versus a teacher as an additional factor in L2 development, as in the present research the laboratory groups were instructed by the author of the study, whereas the two classroom groups were taught by their usual teacher, which sometimes could be an additional factor affecting the final results.

Regarding the methodology, in the present study among several resource-directing variables, proposed by Robinson (2001a) and Robinson & Gilabert (2007) only one variable, +/- spatial reasoning demands, was selected to its further operationalization, so the findings cannot be generalized to other variables. Additionally, treatment tasks were decision-making tasks; however, it has already been shown that the results can vary upon different task types and modes (Gilabert et al., 2011), so studies on the effects of the task sequenced along other resource-directing variables may have on L2 development as well as the tasks sequenced following the principles of Robinson's SSARC model along both resource-directing and resource-dispersing variables on L2 development would be needed. Finally, it is important to carry out more studies on task sequencing and L2 development using various types of tasks, such as narrative tasks, map tasks, etc.

The present study includes only one treatment session in order to control for external input and home rehearsal. However, in real classroom conditions, most of the time students do not receive all the input and the treatment of the same topic in one session, but rather the teacher splits it into several days. In this way, more classroom and laboratory studies would be needed with treatment organized into several sessions to see whether the effects from task sequencing will still persist and to what degree.

The assessment tools were not general but were specifically designed for the purposes of the present research, which makes the results less subject to comparison. For the future research in this area, it would be probably good to elaborate a series of assessment tasks valid for different studies in order to make them more comparable.

Finally, the present study focuses only on L2 development in the use of spatial expressions, which corresponds to a tiny part of the huge linguistic repertoire; therefore much more research on L2 acquisition and L2 development of grammatical and lexical aspects of English (and other languages) is needed.

The final remark is on the limitations regarding WMC selected for inclusion in the main experiment. First, only one test of WMC, the automated reading span, was used to measure WMC of the students. Although it has been proved to be a robust measure of WMC, more spans would likely be a good addition to measuring the construct. Second, while WMC, constitutes an important part of humans' individual differences, there are several other constructs, such as attention, which should not be ignored and should be considered for further research in the field of SLA in relation to task sequencing.

REFERENCES

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*, *Vol. II* (pp. 89-195). New York: Academic Press.

Atkinson, R.C. & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American, 224,* 82-89.

Aula, A., Khan, R. & Guan, Z. (2010). How does search behavior change as search becomes more difficult? *Proceedings of CHI '10*, 35-44.

Baddeley, A.D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

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

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

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Baddeley, A.D., Bressi, S., Della Sala, S., Logie, R.H. & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease: A longitudinal study. *Brain, 114*, 2521-2542.

Baddeley, A.D., Gathercole, S.E. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A. D. & Hitch, G. J. (1974). In G.A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-89). New York: Academic Press.

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

Baralt, M. (2010). *Task complexity, the Cognition Hypothesis, and Interaction in CMC and FTF environments*. Georgetown University, USA.

Baralt, M. (in press). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning*. London: Bloomsbury Publishing.

Baralt, M., Gilabert, R. & Robinson, P. (Eds.) (in press). *Task sequencing and instructed second language learning*. London: Bloomsbury Publishing.

Becker, A., & Carroll, M. (1997). *The Acquisition of spatial relations in a second language.* Amsterdam: John Benjamins.

Bergman, R.A. & Slobin, D.I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.

Bergsleithner, J.M. (2010). Working memory capacity and L2 writing performance. *Ciências and Cognição*, *15 (2)*, 2-20.

Bialystok, E. (1991). *Language processing in bilingual children.* Cambridge: Cambridge University Press.

Block, R. A., Hancock, P. A. & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*, 330-343.

Bonner, S. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society, 19(3)*, 213-234.

Bowerman, M. & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 387-427). Cambridge: MIT Press.

Breen, M. (1989). The evaluation cycle for language learning. In R. K. Johnson (Ed.), *The Second Language Curriculum*. (pp. 187-206). Cambridge: Cambridge University Press.

Brindley, G. (1987). Factors affecting task difficulty. In D.Nunan (Ed.), *Guidelines for the development of curriculum resources* (pp. 45-56). Adelaide: National Curriculum Resource Centre.

Bronchart, J. P. & Sinclair H. (1973). Time, tense and aspect. *Cognition, 2*, 107–30.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10(1)*, 12-21.

Brünken, R., Plass, J.L. & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: auditory load and modality effects. *Instructional Science, 32*, 115-132.

Brusilovsky, P. & Vassileva, J. (2003). Course sequencing techniques for large-scale web-based education. *International Journal of Continuing Engineering Education and Life Long Learning, 13(1-2)*, 75-94.

Bygate, M., Skehan, P. & Swain, E. (Eds.) (2001). *Researching pedagogic tasks: Second language learning, teaching and testing.* London: Pearson Education Limited.

Byrne, R.M.J. & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language, 28*, 564-575.

Byström, K. (1999). Information seekers in context: an analysis of the 'doer' in INSU studies. In T.D. Wilson & D.K. Allen (Eds.), *Exploring the contexts of information behaviour* (pp. 82-95). London: Taylor Graham.

Cadierno, T. & Robinson, P. (2009). Language typology, task complexity, and the development of L2 lexicalization patterns for describing motion events. *Annual Review of Cognitive Linguistics, 7*, 245-276.

Cambell, D.J. (1988). Task complexity: a review and analysis. *Academy of management review, 13 (1)*, 40-52.

Candlin, C. (1987). Towards task-based language learning. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (pp. 5-22). London: Prentice Hall.

Conway, A.R.A., Cowan, N., Bunting, M.F., Therriault, D.J. & Minkoff, S.R.B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

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

Conway, A.R.A, Kane, M.J. & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7(12)*, 547-552.

Caplan, D. & Waters, G.S. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp.337-389). NewYork, NY: Cambridge University Press.

Carey, S. & Spelke, E. S. (1994). Domain-specific knowledge and conceptual change. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169-200). Cambridge, UK: Cambridge University Press.

Carlson, R.A. (1997). *Experienced cognition*. Mahwah, NJ: Lawrence Erlbaum.

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

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

Crookes, G. (1986). *Task classification: A cross-disciplinary review.* Center for Second Language Classroom Research, Technical Report # 4, University of Hawaii.

Dalby, M.A., Elbro, C. & Stødkilde-Jørgensen, H. (1998). Temporal lobe asymmetry and dyslexia: An in Vivo study using MRI. *Brain and Language, 62(1)*, 51-69.

Daneman M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*, 445–464.

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

Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

DeKeyser, R.M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-533.

DeKeyser, R.M., Salaberry, R., Robinson, P. & Harrington, M. (2002). What gets processed in processing Instruction? A commentary on Bill VanPatten's "processing Instruction: An Update". *Language Learning, 52, (4)*, 805-823.

Della Sala, S., Baddeley, A., Papagno, C. & Spinnler, H. (1995). *Annals of the New York Academy of Sciences*, *769*, 161-172.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition.* Cambridge: Cambridge University Press.

Ellis, N. (1996). Sequencing in SLA. *Studies in Second Language Acquisition*, *18 (1),* 91-126.

Ellis, N. & Schmidt, R. (1997). Morphology and longer distance dependencies. *Studies in Second Language Acquisition*, *19 (2),* 145-171.

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

Ellis, R. (Ed.) (2005). *Planning and task performance in a second language*. Amsterdam: John Benjamins.

Emerson, M. & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48(1)*, 148-168.

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, and 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* (pp.102-134). London: Cambridge Press.

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

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Fortkamp, M.B.M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition, 32*, 259-296.

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

Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J. & Viallet, F. (2005). Articulation in early and late bilinguals' two languages: Evidence from functional magnetic resonance imaging. *NeuroReport, 16*, 761–765.

García Mayo, M.P. (Ed.) (2007). *Investigating tasks in formal language learning.* Clevedon: Multilingual Matters.

Garey, M.R. (1973). Optimal task sequencing with precedence constraints. *Discrete Mathematics, 4(1)*, 37-56.

Gass, S., Mackey, A. & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning, 55 (4)*, 575-611.

Gass, S., Mackey, A. & Ross-Feldman, L. (2011). Task-based interactions in classroom and laboratory settings. *Language Learning, 61 (s1)*, 189-220.

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

Gilabert, R. (2007a). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, *45,* 215-40.

Gilabert, R. (2007b). The simultaneous manipulation of task complexity along planning time and [+/-Here-and-Now]: Effects on L2 oral production. In M.P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp.136-156). Clevedon: Multilingual Matters.

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

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

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

Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (pp. 1008-1025). Hillsdale, NJ: Lawrence Erlbaum.

Givon, T. (1995). *Functionalism and Grammar.* Amsterdam: John Benjamins.

Givon, T. 2002. *Bio-Linguistics: The Santa Barbara lectures*. Amsterdam: John Benjamins.

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

Hackman, J. R. (1969). Toward understanding the role of tasks in behavioral research. *Acta Psychologica*, *31*, 97- 128.

Hackman, J.R. & Lawler, E.E. (1971). Employee reaction to job characteristics. *Journal of Applied Psychology Monograph, 55*, 259-286.

Harley, B. & Hart, D. (2002). Age, aptitude, and second language and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 302-330). Amsterdam: John Benjamins.

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R.J. Harris (Ed.), *Cognitive processing in bilinguals*. Amsterdam: Elsevier.

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

Hendy, K. C., Liao, J. & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load, *Human Factors, 39*, 30-47.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.

Herskovits, A. (1997). Language, spatial cognition and vision. In O. Stock (Ed.), *Spatial and temporal reasoning* (pp. 155-202). Dordrecht: Kluwer.

Hicks, R., Miller, G.W. & Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology, 89*, 719-730.

Huang, H.-J. & Mercer, V.S. (2001). Dual-task methodology: Application in studies of cognitive and motor performance in adults and children. *Pediatric Physical Therapy, 13(3)*, 133-140.

Hyltenstam, K. & Abrahamsson, N. (2003). Maturational constraints in SLA. In: C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 539–588). Oxford: Blackwell.

Ishikawa, T. (2007). The effect of increasing task complexity along the [+/-Here-and-Now] dimension on L2 written narrative discourse. In M.P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp.136-156). Clevedon: Multilingual Matters.

Ishikawa, T. (2008). *Investigating the Effect of Intentional Reasoning Demands on L2 Speech Production*. Unpublished Dissertation. Aoyama Gakuin University, Japan.

James, W. (1890). *The principles of psychology*. NewYork: Holt, Rinehart & Winston.

Johnson, D. S. & Kanfer, R. (1992). Goal-performance relations: The effects of initial task complexity and task practice. *Motivation and Emotion, 16*, 117-141.

Johnson-Laird, P. (1983). *Mental Models. Towards a cognitive science of language, inference, and consciousness.* Harvard University Press. Cambridge.

Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum.

Jones, D.M., Macken, W.J. & Nicholls, A.P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology, 30*, 656-674.

Just, M.A. & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kaushanskaya, M. & Marian, V. (2007). Bilingual language processing and interference in bilinguals: evidence from eye-tracking and picture naming. *Language Learning, 57(1)*, 119-163.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58 (2)*, 285-325.

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

King, J. & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Klauer, K.C. & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133*, 355-381.

Klein, W. & Perdue, C. (1992). *Utterance structure. Developing grammars again.* Amsterdam: John Benjamins.

Klein, W., & Perdue, C. (1997). The basic variety or: Couldn't natural languages be much simpler? *Second Language Research, 13*, 301-347.

Kovelman, I., Baler, S. A. & Petitto, L.-A. (2008). Bilingual and monolingual brains compared: A functional magnetic resonance imaging investigation of syntactic processing and a possible neural signature of bilingualism. *Journal of Cognitive Neuroscience, 20*, 153–160.

Kormos, J. (2006). *Speech production and second language acquisition.* (Cognitive sciences and second language acquisition). Mahwah, N.J.: Lawrence Erlbaum.

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

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

Krashen, S. (1985). *The Input Hypothesis: Issues and implications.* California: Laredo Publishing Co.

Krashen, S. & Terrell. T. (1983). *The natural approach: Language acquisition in the classroom.* Oxford: Pergamon Press.

Kuiken, F., Mos, M. & Vedder, I. (2005). Cognitive task complexity and second language written performance. In S. Foster-Cohen, M. P. García-Mayo & J. Cenoz (Eds.), *EUROSLA Yearbook* (Vol. 5) (pp. 195-222). Amsterdam: John Benjamins.

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

Kuiken, F. & Vedder, I. (2007b). Cognitive task complexity and linguistics performance in French L2 writing. In M.P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp.117-135). Clevedon: Multilingual Matters.

Kuiken, F. & Vedder, I. (2007c). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17(1)*, 48-60.

Kuiken, F. & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *18*, 48-60.

Lambert, C. & Robinson, P. (in press). Learning to perform narrative tasks in the L2 classroom: A longitudinal intact-group, semester long study of task sequencing effects. In M. Baralt, R. Gilabert & P. Robinson (Eds.), *Task sequencing and instructed second language learning*. London: Bloomsbury Publishing.

Landau, B. & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217-265.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass.: MIT Press.

Levinson, S. C. (1996). Relativity in spatial conception and description. In J. J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 177-202). Cambridge: Cambridge University Press.

Levinson, S.C. (2003). *Space in language and cognition*. Cambridge: Cambridge University Press.

Levkina, M. & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, I. Vedder & F. Kiuken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 171–198). Amsterdam: John Benjamins.

Levkina, M. & Gilabert, R. (in press). Task sequencing in the L2 development of spatial expressions. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing* *and instructed second language learning*. London: Bloomsbury Publishing.

Liu, J., Gwizdka, J., Liu C. & Belkin, N.J. (2010). Predicting task difficulty for different task types. *Proceedings of ASIS&T '10*.

Liu, P. & Li, Z. (2012). Task complexity: a review and conceptualization framework. *International Journal of Industrial Ergonomics, 42*, 553-568.

Logie, R.H. (1995). *Visuo-Spatial Working Memory*. Hove, UK: Lawrence Erlbaum.

Long, M.H. (1985). A role for instruction in second language acquisition: task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 77-99). Clevedon: Multilingual Matters.

Long, M.H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*, 251-285.

Long, M.H. (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.

Long, M.H. & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26 (1), 27-55.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-63). Cambridge: Cambridge University Press.

Mackey, A., Adams, R., Stafford, C. & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60,* 501–533.

Mackey, A., Philp, J., Egi, T., Fujii, A. & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). Amsterdam: John Benjamins.

Mainwaring, S. D., Tversky, B., Ohgishi, M. & Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation, 3*, 3-42.

Malicka, A. & Levkina, M. (2012). Measuring task complexity. Does EFL proficiency matter? In A. Shehadeh & C. Coombe (Eds.) *Task-based language teaching in foregin language contexts. Research and implementation*. Amsterdam: John Benjamins.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

McCormick, E.J. (1976). Job and task analysis. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.

McDonough, L., Choi, S. & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259.

Meara, P., & Milton J., 2005. X-Lex: *The Swansea levels test* [CD-ROM]. UK: Express Publishing.

Meara, P. & Miralpeix, I., 2006. *Y_Lex: the Swansea advanced vocabulary levels test*. v2.05. Swansea: Lognostics.

Michel, M.C., Kuiken, F. & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics*, *45(3)*, 241-259.

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

Miyake, A. & Friedman, N.P. (1998). Individual differences in second language proficiency: working memory as language aptitude. In A. F. Healy & L.E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (p. 339-364). Mahawah, N. J: Lawrence Erlbaum.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal love” tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Miyake, A., & Shah, P. (Eds.) (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Mizera, G. J. (2006). *Working memory and L2 oral fluency.* PhD Dissertation. University of Pittsburgh.

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

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

Nori, R., Iachini, T. & Giusberti, F. (2004). Object localisation and frames of reference. *Cognitive Processing, 5*, 45-53.

Norman, D. & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. Center for human information processing. In R.J. Davidson, G.E.Schwartz & D. Shapiro, *Consciousness and Self-Regulation, Vol.4* (pp.1-18). New York: Plenum Press.

Norris, J. M. & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 716-761). London: Blackwell.

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

Nunan, D. (2004). *Task-based language teaching: A comprehensively revised edition of designing tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Olive, T. (2004). Working memory in writing: Empirical evidences from the dual-task technique. *European Psychologist, 9*, 32-42.

Orstein, R.E. (1969). *On the experience of time*. Harmondsworth, UK: Penguin.

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

Perdue, C. (Ed.) (1993a). *Adult language acquisition: Cross-linguistic perspectives. Volume 1, Field methods.* Cambridge: Cambridge University Press.

Perdue, C. (Ed.) (1993b). *Adult language acquisition: Cross-linguistic perspectives. Volume 2. The results.* Cambridge: Cambridge University Press.

Peterson, N. G. & Bowners, D. A. (1982). Skill, task structure and performance acquisition. In M. D. Dunnette & E. A. Fleishman (Eds.), *Human performance and productivity (Vol.1).* Hillsdale, NJ: Lawrence Erlbaum.

Peterson, L. & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58(3*), 193-198.

Pienemann, M. & Johnston, M. (1987). Factors influencing the development of language proficiency. In D.Nunan (Ed.), *Applying second language acquisition research.* (p. 45-142). Adelaide: National Curriculum Resource Centre.

Pienemann, M., Johnston, M. & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*, 217-243.

Prabhu, N. S. (1987). *Second language pedagogy.* Oxford: Oxford University Press.

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

Rapp, D. N. & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 988-1001.

Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C. & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition and Computation, 5 (2-3)*, 239-269

Recio, M. (2011). *The effects of task complexity on L2 oral production as mediated by differences in working memory capacity*. Unpublished Master dissertation, University of Barcelona.

Révész, A. (2008). *Task complexity, focus on form-meaning connections, and individual differences: A classroom-based study*. Paper given at the AILA 2008 conference.

Révész, A. (2012). Working memory and the observed effectiveness of recasts in different L2 outcome measures. *Language Learning, 62 (1)*, 93-132.

Révész, A., Michel, M. & Gilabert, R. (2012). *Measuring cognitive task demands: A step forward in task complexity research.* The 22d edition of the EuroSLA conference.

Révész, A., Michel, M. & Gilabert, R. (2013). *Measuring cognitive task demands using dual-task methodology, subjective self-ratings, time estimation and estimated recall.* The AAAL conference, Dallas (USA).

Richards, J., Platt, J. & Weber, H. (1985). *Longman dictionary of applied linguistics.* Harlow, Essex, England: Longman.

Robinson, P. (1995a). Task complexity and second language narrative discourse. *Language Learning*, *45*, 99-140.

Robinson, P. (1995b). Attention, memory and the 'noticing' hypothesis. *Language Learning*, *45*, 283-331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27-67.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45-99.

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

Robinson, P. (2001b). 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.

Robinson, P. (2001c). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*, 368-392.

Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*, 45-105.

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

Robinson, P. (2007a). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, *45*, 191-213.

Robinson, P. (2007b). Criteria for classifying and sequencing pedagogic tasks. In María del Pilar García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon: Multilingual Matters.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz and L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind.* (pp. 243-268). Amsterdam: John Benjamins.

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

Robinson, P. & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and secondlanguage learning and performance. *International Review of Applied Linguistics in Language Teaching, 45*, 161-176.

Rock, I., Wheeler, D. & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology, 21*, 185-210.

Rodriguez-Fornells, A., van der Lugt, A., Rotte, M., Britti, B., Heinze, H.-J. & Münte, T.F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience, 17(3)*, 422-433.

Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics, 7*, 149-182.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 229–248). Oxford, UK: Oxford University Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction,* (pp. 1-32). Cambridge: Cambridge University Press.

Shah, P. & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General, 125(1)*, 4-27.

Shehadeh,A. & Coombe, C.A. (2012). *Task-based language teaching in foreign language contexts*. Amesterdam: John Benjamins.

Shirai, Y. & Andersen R. W. (1995). The acquisition of tense-aspect morphology: prototype account. *Language, 71*, 743–762.

Simnett, R. (1996). The effect of information selection, information processing and task complexity on predictive accuracy of auditors. *Accounting, Organizations and Society, 21(7-8)*, 699-719.

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

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

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

Slobin, D. (1993). Adult language acquisition: A view from child language study. In C. Perdue (Ed.), *Adult language acquisition: Cross-linguistic perspectives*, Vol. 2: The results, (pp. 239-252). Cambridge: Cambridge University Press.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donder's method. *Acta Psychologia, 30*, 276–315.

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

Talmy, L. (1983). How language structures space. In L. Herbert, J. Pick & P. L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum Press.

Tanenhaus, M.K., Magnuson, J.S., Dahan, D. & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistics processing. *Journal of Psycholinguistic Research, 29(6),* 557-580.

Tanenhaus, M.K. & Spivey-Knowlton, M.J. (1996). Eye-tracking. *Language and Cognitive Processes, 11(6)*, 583-588.

Taylor, H.A. & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language, 35,* 371-391.

Thomas, E.A.C. & Weaver, W.B. (1975). Cognitive processing and time perception. *Perception and Psychophysics, 17*, 363-367.

Tomasello, M. (2003). *Constructing a language*. Harvard, MA: Harvard University Press.

Topi, H., Valacich J.S. & Hoffer, J.A. (2005). The effects of task complexity and time availability limitations on human performance in database query tasks. *International journal of Human-computer Studies, 62(3)*, 349-379.

Traxler, M.J. & Pickering, M.J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35(3)*, 454-475.

Trebits, A. & Kormos, J. (2008). Working Memory Capacity and Narrative Task Performance. *Proceedings form the 33rd International LAUD Symposium*, Landau/Pfalz, Germany.

Turner, M. & Engle, R.W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

Underwood, G. & Swain, R. A. (1973). Selectivity of attention and the perception of duration,*Perception, 2(1)*, 101 – 105.

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

Unsworth, N., Redick, T.S., Heitz, R.P., Broadway, J.M. & Engle, R.W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17(6)*, 635-654.

Van den Branden, K. (2006). *Task-based language education*. Cambridge: Cambridge University Press.

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

VanPatten, B. & Cadierno, T. (1993a). SLA as input processing: A role for instruction, *The Modern Language Journal*, *77 (1)*, 45-57.

VanPatten, B. & Cadierno, T. (1993b). Explicit instruction and input processing, *Studies in Second Language Acquisition*, *15*, 225-243.

Weissheimer, J. & Mota, M.B. (2009). Individual differences in working memory capacity and the development of L2 speech production. *Issues in Applied Linguistics, 17*, 34-52.

Weist, R.M. (1989). Time concepts in language and thought. Filling the Piagetian void from three to five years. In L .Levin & O. Zakay (Eds.), *Time and human cognition: A life span perspective* (pp. 63-118). Amsterdam: North-Holland.

White, R. (1988). *The ELT curriculum: Design, innovation and management*. Cambridge, MA: Basil Blackwell.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Widdowson, H.G. (1979). *Explorations in applied linguistics*. Oxford: Oxford University Press.

Willis, D. (1990). *The lexical syllabus*. London: Harper Collins.

Wood, R.E. (1986). Task complexity: definition of the construct. *Organizational Behavior and Human Decision Processes*, *37(1)*, 60-82.

Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24(1)*, 1-27.

Zacks, J. M., Mires, J., Tversky, B. & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation, 2*, 315–332.

Zacks, J. M. & Tversky, B. (2005). Multiple systems for spatial imagery: Transformations of objects and perspective. *Spatial Cognition and Computation, 5,* 271-306.

Appendix A. *Affective variable questionnaire and time estimation task (Pilot 1)*

Anxiety and Perceived Difficulty Questionnaire

Please, tell me how much you agree or disagree with the following statements by circling a number from 1 to 6:

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
1	**2**	**3**	**4**	**5**	**6**

1.	This task was difficult.	1 2 3 4 5 6
2.	I felt like I didn't have enough time before I respond.	1 2 3 4 5 6
3.	I felt rushed during the task.	1 2 3 4 5 6
4.	I was relaxed and comfortable completing the task.	1 2 3 4 5 6
5.	This task was stressful for me.	1 2 3 4 5 6
6.	I enjoyed doing the task.	1 2 3 4 5 6
7.	The instructions of the task were clear.	1 2 3 4 5 6
8.	The task made me feel anxious.	1 2 3 4 5 6
9.	I could complete the task easily.	1 2 3 4 5 6
10.	I had no problem in completing the task.	1 2 3 4 5 6

TIME ESTIMATION QUESTIONNAIRE

How much time did the task take you to complete?

SET 1

TASK 1	**TASK 2**	**TASK 3**

SET 2

TASK 1	**TASK 2**	**TASK 3**	**TASK 4**

Appendix B. *Retrospective Protocol*

Retrospective Protocol

I. General questions

1. How did you find the tasks?

2. What are the main problems that you had in completing them?

3. Did you have enough time to prepare?

4. Did you find the instructions clear enough or there is something that you would probably miss in there?

5. Did you find this task realistic enough?

II. Set 1 and Set 2

1. Which task did you find the most difficult /the easiest one? Why ? Please, give me your reasons?

2. Which task did you find more interesting or less interesting ? why?

III. Based on the AVQ

Compare answers for sets 1 and 2 and ask the appropriate questions.

Appendix C. *Simple Task (pilot 1, version 1)*

THREE COLOURS: WHITE

You've just moved to a new flat.

Now you have to furnish it.

You decided to start with the kitchen

(on the right):

You've got a list of the items you would like to put in it (in the left column) and their photos (in the right column):

1. A table

   length x diameter x height

174cm x 126cm x 74cm

2. Chairs

   width x depth x height

46cm x 64cm x 76cm

3. A carpet

   diameter

150 cm

4. A picture

   width x height

70cm x 100cm

5. A clock

   diameter

65cm

6. An armchair

   width x depth x height

80cm x 70cm x 77cm

7. Two stools

   width x depth x height

50cm x 40cm x 98cm

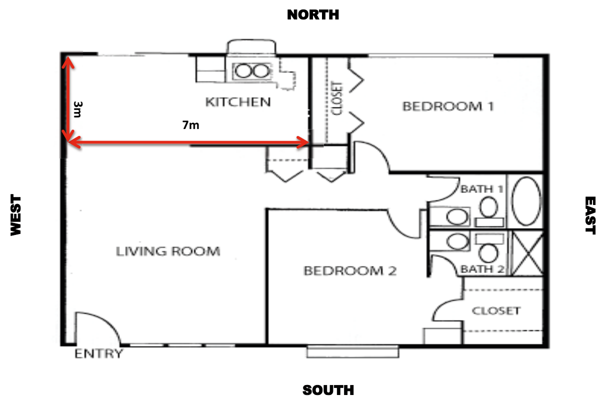
z

YOUR TASK

Now you are at a furniture shop.

Please, as precisely as possible explain to a furniture delivery man where you want each of the objects in your kitchen.

You have **1 minute** to think it over before you start.



Appendix D. *Complex task (pilot 1, version 2)*

THREE COLOURS: RED

This is the last space to furnish, your living room on the right ...

Once again you've got a list of the items you would like to be in your living room in the left column. Here, on the contrary, you still doubt which object you want in there. So, you've got three options to choose from:

	44¨	37¨	46¨
1. A TV set			
	130cm x 25cm	140cm x 32cm	135cm x 34cm
2. A floor lamp *height x shade diameter*			
	215cm x 90cm x 75cm	210cm x 85cm x 79cm	203cm x 95cm x 85cm
3. A sofa *width x depth x height*			
	118cm x 75cm x 46cm	110cm x 70cm x 44cm	100cm x 100cm x 45cm
4. A coffee table *length x width x height*			
	95cm x 95cm	95cm x 45cm	50cm x 90cm
5. A wall photo *width x height*			
	205cm x 40cm x 205cm	160cm x 35cm x 195cm	170cm x 35cm x 170cm
6. A bookshelf *width x depth x height*			
	150 cm	145 cm	135 cm
7. A floor vase *height*			

YOUR TASK

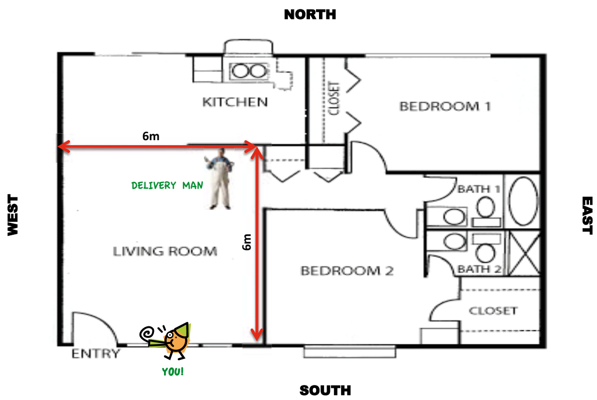
You are still with the delivery man to explain to him how to furnish your living room.

But, first, you have to choose one of the three items given above. Please, do **NOT** refer to the size of the objects, which as previously, is given to help you place them in the living room.

Second, give the instructions to the furniture delivery man on where you want each of the objects to be placed in your living room.

Here again, please, talk from the position of the delivery man, as he is on the opposite side of the room.

As previously, you've got **2 minutes** for preparation.



Appendix E. Simple task (pilot 1, version 1)

LET'S FURNISH YOUR FLAT!

TASK 1

You've just moved to a new flat.

Now you have to furnish it.

You decided to start with the kitchen

(on the right):

You've got a list of the items you would like to put in it (in the left column) and their photos (in the right column):

1. A table
   length x diameter x height

174cm x 126cm x 74cm

2. Chairs
   width x depth x height

46cm x 64cm x 76cm

3. A carpet
   diameter

150 cm

4. A clock
   diameter

65cm

5. An armchair
   width x depth x height

80cm x 70cm x 77cm

6. Two stools
   width x depth x height

50cm x 40cm x 98cm

YOUR TASK

Imagine you are in the furniture shop.

With as much precision as possible explain to the furniture delivery man where you want each of the selected objects in your kitchen.

During your explanation for each item, please, make a reference to the two objects, which are already in the room (marked with a cross).

You have **1 minute** to think it over before you start.

Appendix F. Complex task (pilot 1, version 2)

TASK 3

This is the last space to furnish, your living room on the right …

Once again you've got a list of the items you would like to place in it (in the left column) and their photos (in the right column).

1. A TV set

44¨

2. A sofa

width x depth x height

215cm x 90cm x 75cm

3. A coffee table

length x width x height

118cm x 75cm x 46cm

4. A wall photo

width x height

95cm x 95cm

5. A bookshelf

width x depth x height

205cm x 40cm x 205cm

6. A floor vase

height

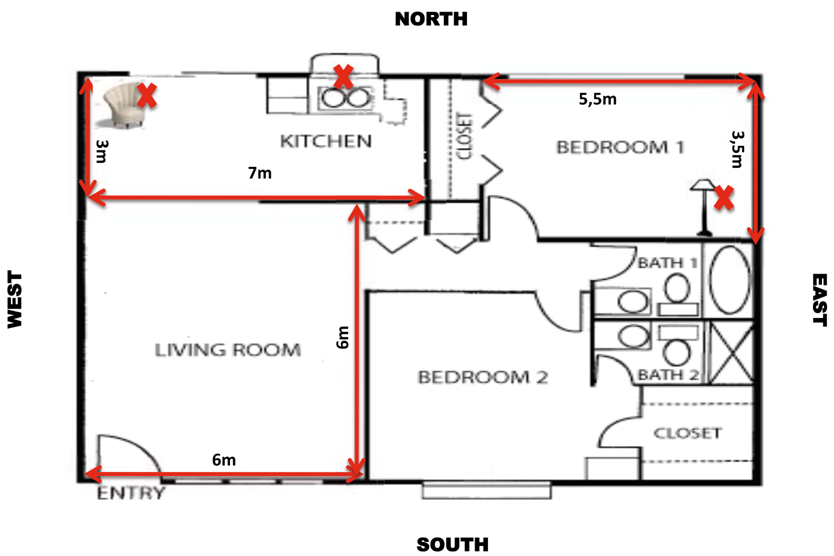
150 cm

YOUR TASK

Last minutes in the furniture shop.

For your last space, as precisely as possible explain to a furniture delivery man where you want each of the objects in your living room.

You have **1 minute** to think it over before you start.


NORTH
KITCHEN
3m
7m
CLOSET
5,5m
BEDROOM 1
3,5m
WEST
EAST
BATH 1
BATH 2
LIVING ROOM
6m
BEDROOM 2
CLOSET
6m
ENTRY
SOUTH

Appendix G. *Input (student A)*

Using prepositions of Place

(Student A)

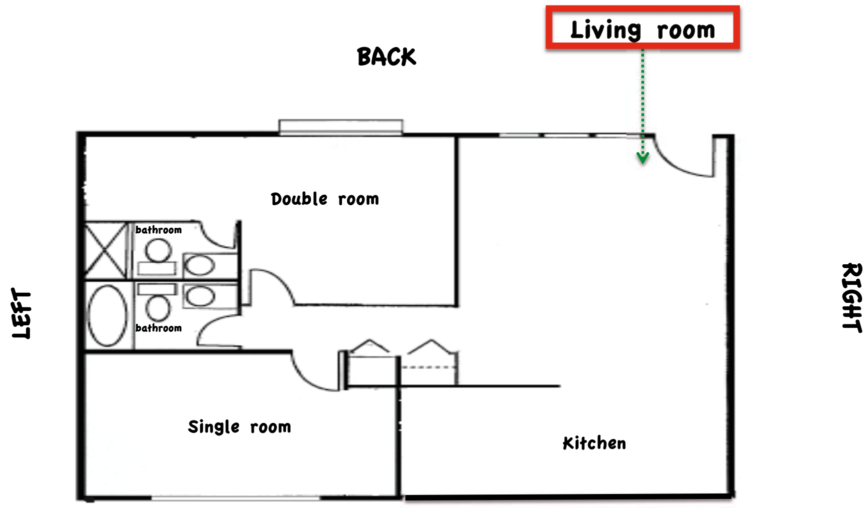
In this packet are three tasks.

The <u>FIRST task</u> you will do with your teacher.

The <u>SECOND</u> and the <u>THIRD</u> tasks you will do with your partner.

Follow the teacher's instructions and draw the pieces of furniture on the living room map.

TEXT 1



TEXT 1

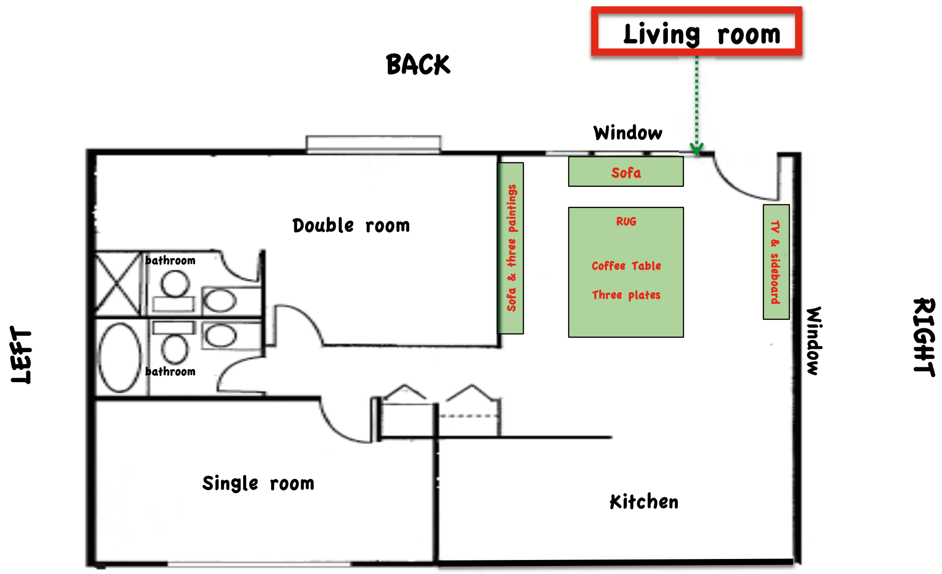
There are two windows in the living room. The first one is **at the back**. The second one is **on the right-hand side** of the room.

On the right-hand side of the room there is also a sideboard and a TV. The sideboard is **near** the window on the left. The TV is **on** the wall **above** the sideboard.

There are two sofas in the room. One of them is **along** the left-hand side wall. It's **opposite** the sideboard. Three paintings are **on** the wall **above** the sofa.

The other sofa is **along** the back window.

A big rug is right **in the middle of** the room. There is a coffee table on the rug. Three plates are **on top of** the coffee table.



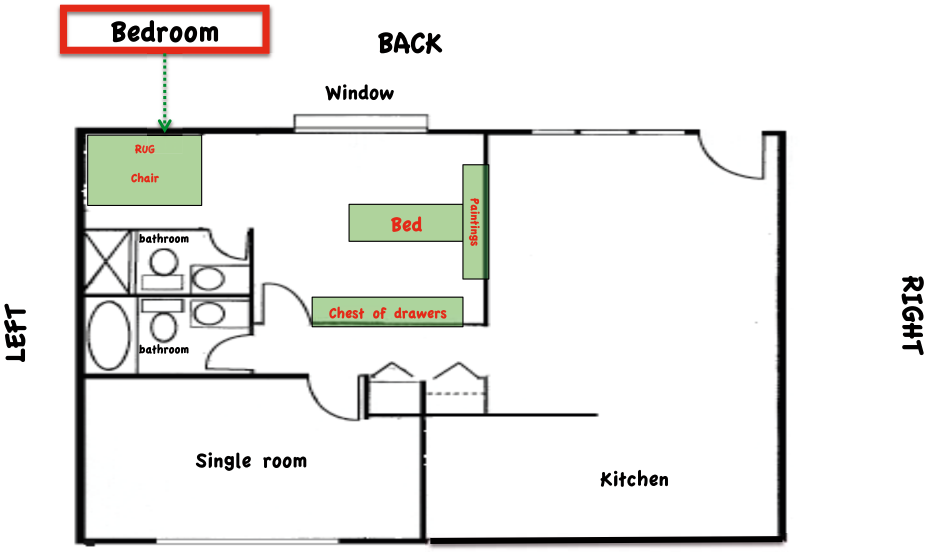
Now, please, give the instructions to your partner.

TEXT 2

This bedroom has got one big window **at the back**.

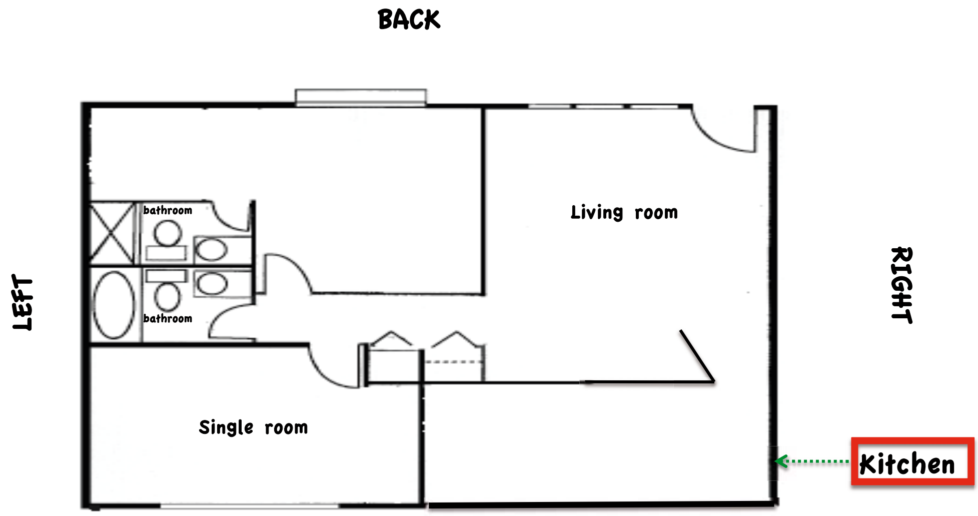
There is a bed **along** the right wall. There are two paintings **on** the same wall **above** the bed. The chest of drawers is **along** the wall **opposite** the window **near** the door.

There is a small round rug **in the** left-hand corner **at the back** of the room. There is a chair **on top of** the rug.



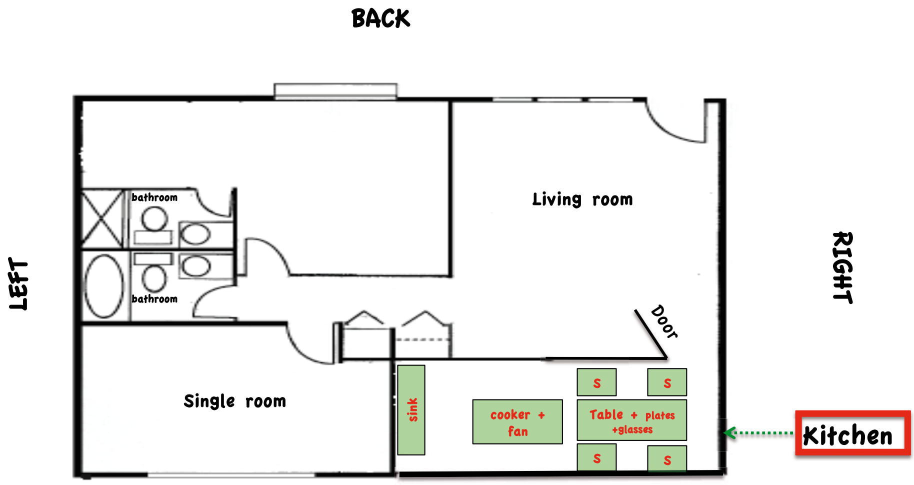
Now, please, follow the instructions of your partner.

TEXT 3



TEXT 3

There is a door **at the back** in the right corner. The sink unit is **along** the left wall. The cooker is **in the middle of** the kitchen **opposite** the sink. The extractor fan is **above** the cooker. There is a table **near** the cooker, **on the right-hand side**. Some plates and glasses are **on top of it**. Four stools are **around** the table.



Appendix H. *Input (student B)*

Using prepositions of Place

(student B)

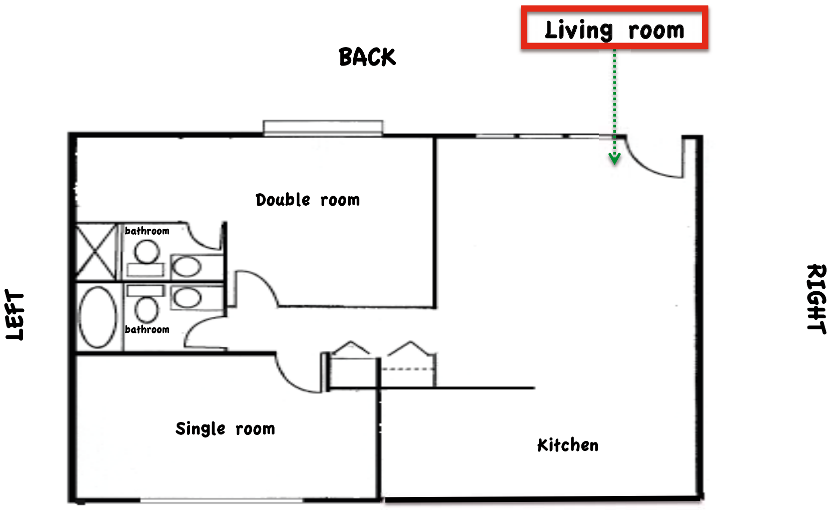
In this packet are three tasks.

The FIRST task you will do with your teacher.

The SECOND and the THIRD tasks you will do with your partner.

Follow the teacher's instructions and draw the pieces of furniture on the living room map.

TEXT 1



TEXT 1

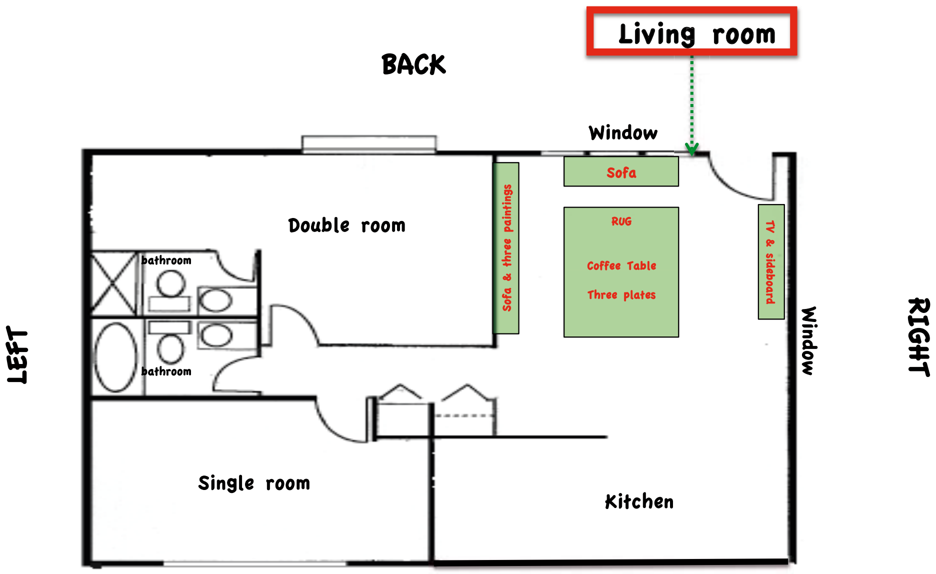
There are two windows in the living room. The first one is **at the back**. The second one is **on the right-hand side** of the room.

On the right-hand side of the room there is also a sideboard and a TV. The sideboard is **near** the window on the left. The TV is **on** the wall **above** the sideboard.

There are two sofas in the room. One of them is **along** the left-hand side wall. It's **opposite** the sideboard. Three paintings are **on** the wall **above** the sofa.

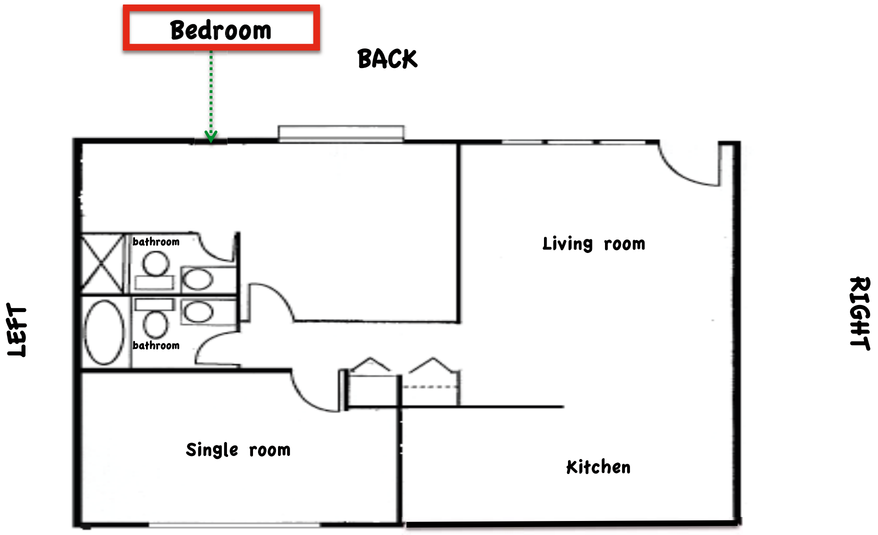
The other sofa is **along** the back window.

A big rug is right **in the middle of** the room. There is a coffee table on the rug. Three plates are **on top of** the coffee table.



Now, please, follow the instructions of your partner.

TEXT 2

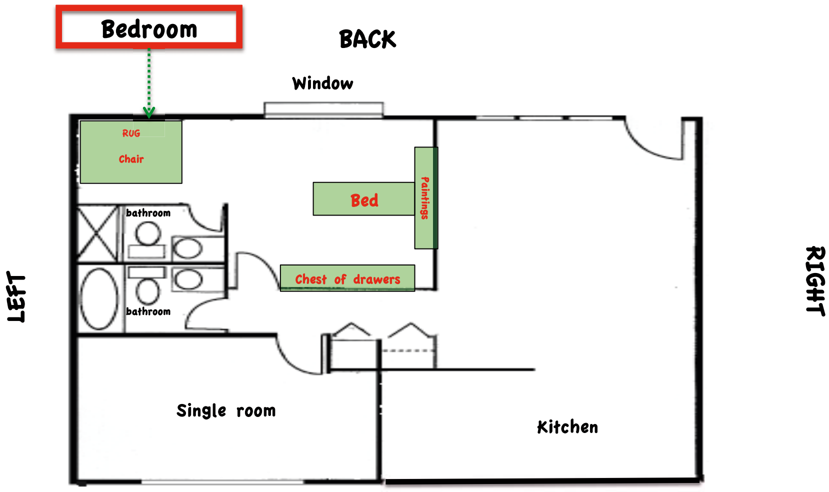


TEXT 2

This bedroom has got one big window **at the back**.

There is a bed **along** the right wall. There are two paintings **on** the same wall **above** the bed. The chest of drawers is **along** the wall **opposite** the window **near** the door.

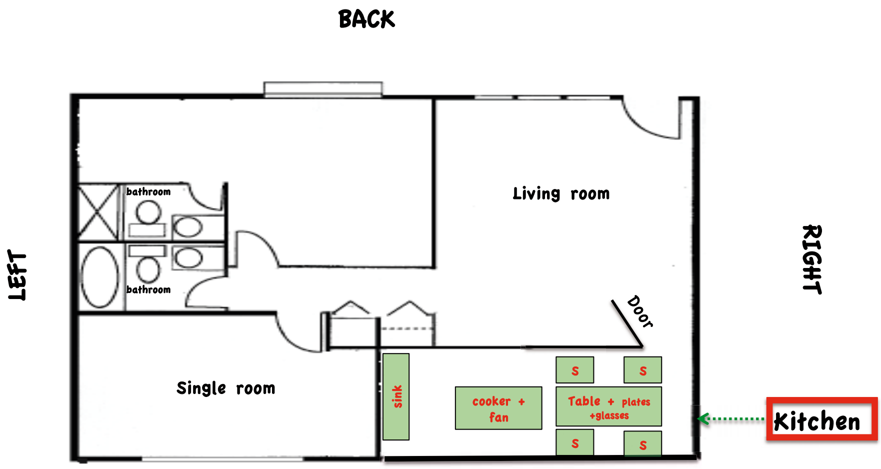
There is a small round rug **in the** left-hand corner **at the back** of the room. There is a chair **on top of** the rug.



Now, please, give the instructions to your partner.

TEXT 3

There is a door **<u>at the back</u>** in the right corner. The sink unit is **<u>along</u>** the left wall. The cooker is **<u>in the middle of</u>** the kitchen **<u>opposite</u>** the sink. The extractor fan is **<u>above</u>** the cooker. There is a table **<u>near</u>** the cooker, **<u>on the right-hand side</u>**. Some plates and glasses are **<u>on top of it</u>**. Four stools are **<u>around</u>** the table.



Appendix I. *Treatment (sequence from simple to complex)*

Sequence 1

A new flat in …

London

Your Name & Surname: ______________________________________

TASK 1

"Double Bedroom"

You've just moved to a new flat in London. Now you have to furnish it.
You decided to start with **the double room**.
You've got a list of the items you would like to put in your bedroom:

Lamp ONE

Side Table ONE

Lamp TWO

Side Table TWO

A TV set

A bed

A sideboard

A wardrobe

A painting

A rug

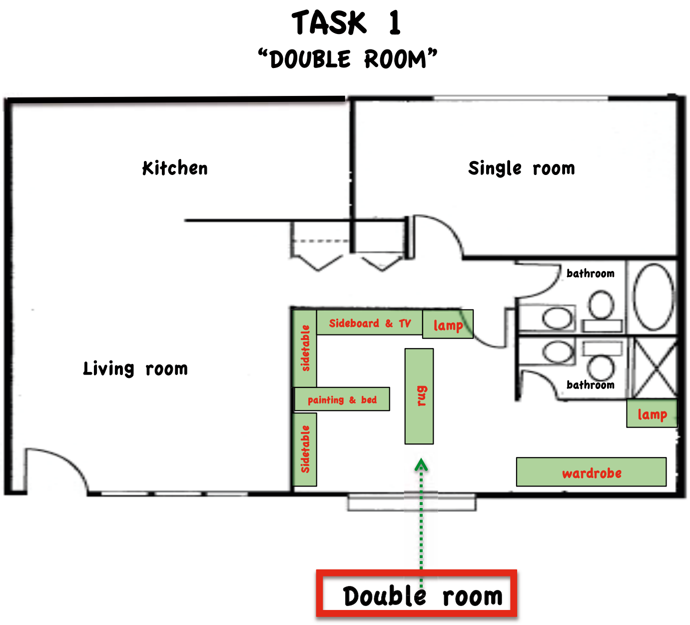
What do you have to do ...

Imagine that delivery service came to your flat.

You have to explain to the delivery man as precisely as possible where you want **each of the objects** given above in the room.

Below you've got the plan of your flat where you've already got the places of these objects. So, the only thing to do is to explain it.

Please, start immediately.



TASK I

I. What are your impressions?

Evaluate the **FIRST TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0	1	2	3	4	5	6	7	8	9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0	1	2	3	4	5	6	7	8	9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0	1	2	3	4	5	6	7	8	9	This task was **INTERESTING,** **so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer:

TASK 2

"Kitchen"

You continue with **the kitchen**. Here again you've got a list of the items you would like to put in your kitchen:

A table

Kitchen furniture set

Two stools

Floor lamp ONE

Floor lamp TWO

Four chairs

A fridge

A bin

A clock

A picture

r

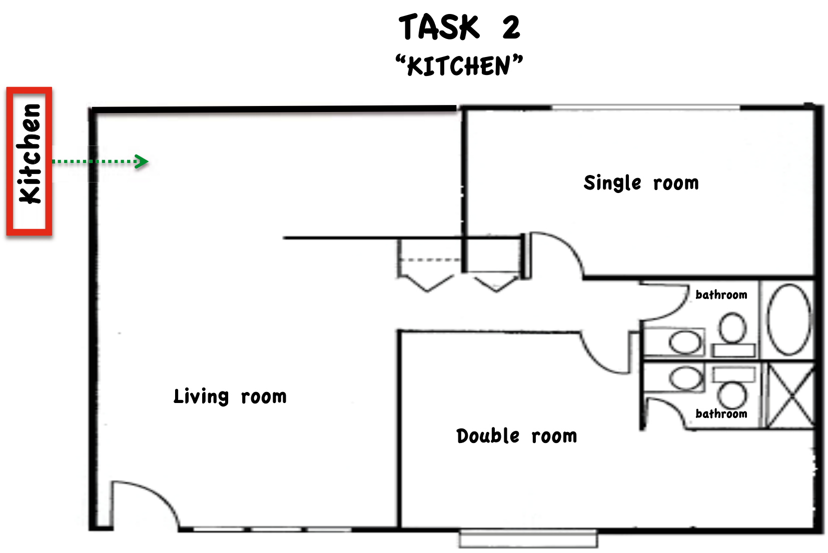
What do you have to do ...

Once again, you have to explain to the delivery man as precisely as possible where you want **EACH** of these objects in the room.

This time the objects are **NOT** placed.

Please, start your explanation immediately.

DO NOT DRAW ON THE PLAN!



TASK II

I. What are your impressions?

Evaluate the **SECOND TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0 1 2 3 4 5 6 7 8 9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0 1 2 3 4 5 6 7 8 9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0 1 2 3 4 5 6 7 8 9	This task was **INTERESTING, so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

TASK 3

"Living room"

You've almost done. The only space to furnish is **your living room**. Here again you've got a list of the items you would like to put in it:

A TV set

50"inches

Sofa ONE

width x depth x height
200cm xv 90cm x 70cm

Sofa TWO

width x depth x height
200cm xv 90cm x 80cm

Sofa THREE

width x depth x height
200cm xv 90cm x 70cm

A rug

length x width
200cm x 150cm

A painting

width x height
70cm x 100cm

A bookshelf

width x depth x height
200cm x 40cm x 200cm

A coffee table

length x width x height
110cm x 70cm x 40cm

A bathtub

length x width x height
200cm x 70cm x 120 cm

Drums

length x width x height
150cm x 50 cm x 140 cm

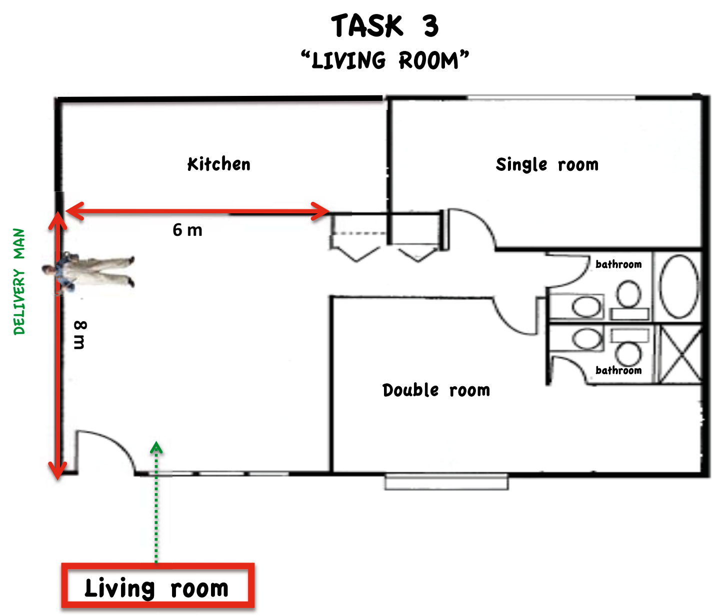
What do you have to do …

For the last time you have to explain to the delivery man as precisely as possible where you want **EACH** of these objects in the room.

You have to **(I)** place all the objects, **(II)** take into account the size of the rooms and the objects and also **(III)** the positions of the delivery man in the room.

Please, start your explanation immediately.

DO NOT DRAW ON THE PLAN!



TASK III

I. What are your impressions?

Evaluate the **THIRD TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0	1	2	3	4	5	6	7	8	9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0	1	2	3	4	5	6	7	8	9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0	1	2	3	4	5	6	7	8	9	This task was **INTERESTING,** **so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

Appendix J. *Treatment (random sequence)*

Sequence 2

A new flat in …

London

Your Name & Surname: ______________________________________

TASK 1

"Kitchen"

You've just moved to a new flat in London. Now you have to furnish it. You decided to start with **THE KITCHEN**. You've got a list of the items you would like to put in it:

A table

Four chairs

Kitchen furniture set

A fridge

Two stools

A bin

Floor lamp ONE

A clock

Floor lamp TWO

A picture

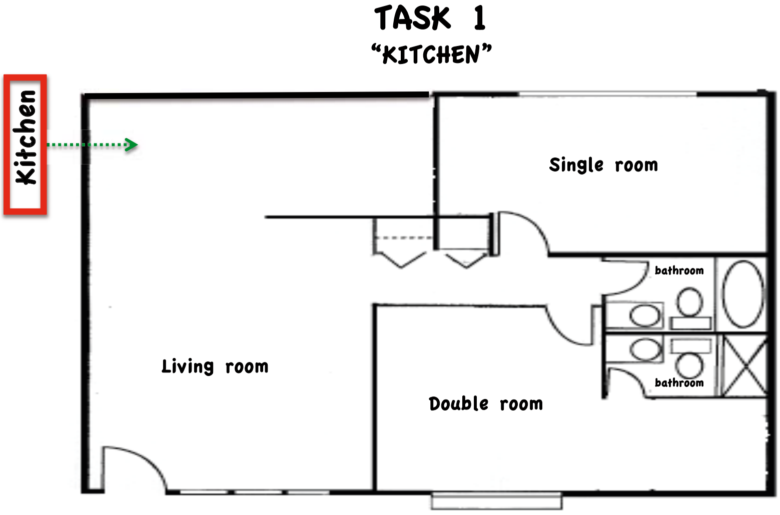
What do you have to do ...

Imagine that delivery service came to your flat.

You have to explain to the delivery man as precisely as possible where you want **EACH** of these objects in the room. You have to place **ALL** the objects!

Please, start your explanation immediately.

DO NOT DRAW ON THE PLAN!



TASK I

I. What are your impressions?

Evaluate the **FIRST TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0	1	2	3	4	5	6	7	8	9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0	1	2	3	4	5	6	7	8	9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0	1	2	3	4	5	6	7	8	9	This task was **INTERESTING, so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

TASK 2

"Double Bedroom"

You continue with **THE DOUBLE ROOM**. Here again you've got a list of the items you would like to put in your kitchen:

Lamp ONE

Side Table ONE

Lamp TWO

Side Table TWO

A TV set

A bed

A sideboard

A wardrobe

A painting

A rug

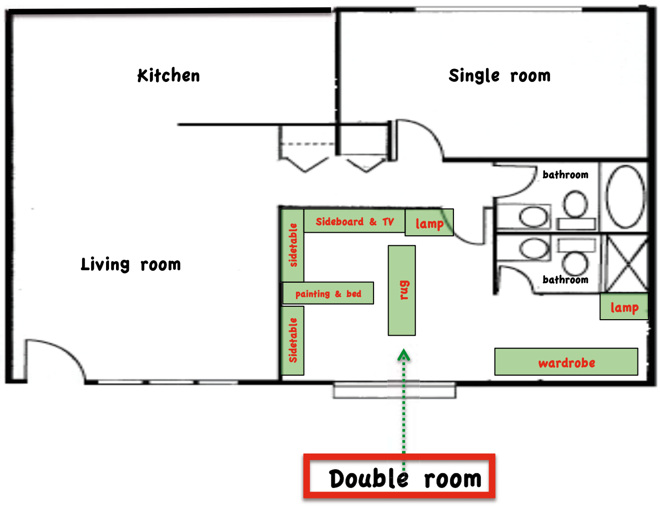
What do you have to do ...

You have to explain to the delivery man as precisely as possible where you want **<u>each of the objects</u>** given above in the room.

In the plan of your flat you've already got the places of these objects. So, the only thing to do is to explain it.

Please, start immediately.

TASK 2
"DOUBLE ROOM"



TASK II

I. What are your impressions?

Evaluate the **SECOND TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0	1	2	3	4	5	6	7	8	9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0	1	2	3	4	5	6	7	8	9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0	1	2	3	4	5	6	7	8	9	This task was **INTERESTING, so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

TASK 3

"Living room"

You've almost done. The only space to furnish is **YOUR LIVING ROOM**. Here again you've got a list of the items you would like to put in it:

A TV set

50''inches

Sofa ONE

width x depth x height
200cm xv 90cm x 70cm

Sofa TWO

width x depth x height
200cm xv 90cm x 80cm

Sofa THREE

width x depth x height
200cm xv 90cm x 70cm

A rug

length x width
200cm x 150cm

A painting

width x height
70cm x 100cm

A bookshelf

width x depth x height
200cm x 40cm x 200cm

A coffee table

length x width x height
110cm x 70cm x 40cm

A bathtub

length x width x height
200cm x 70cm x 120 cm

Drums

length x width x height
150cm x 50 cm x 140 cm

What do you have to do ...

For the last time you have to explain to the delivery man as precisely as possible where you want **EACH** of these objects in the room.

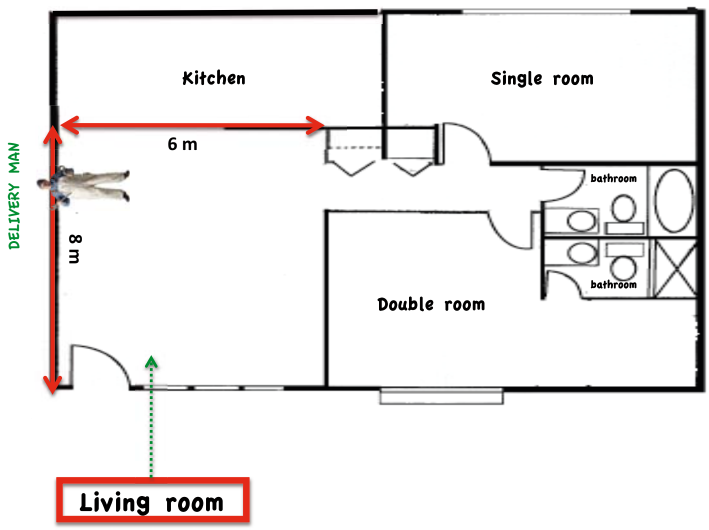
You have to **(I)** place all the objects, **(II)** take into account the size of the rooms and the objects and also **(III)** the positions of the delivery man in the room.

Please, start your explanation immediately.

DO NOT DRAW ON THE PLAN!

TASK 3

"LIVING ROOM"



TASK III

I. What are your impressions?

Evaluate the **THIRD TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0 1 2 3 4 5 6 7 8 9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0 1 2 3 4 5 6 7 8 9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** so I **DON'T want** to do **more** tasks like this	0 1 2 3 4 5 6 7 8 9	This task was **INTERESTING, so** I **WANT TO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

Appendix K. *Descriptive task*

TASK

"LIVING ROOM"

Imagine that you've just moved to London and now you have to furnish your new flat. Today you went to visit a friend of yours and you loved her /his living room.

Here is the living room.

WHAT DO YOU HAVE TO DO? ...

Now you're already at home and you want to share your impressions with ***your friend*** *who is still in Spain.* ***Write him/her an e-mail*** *describing the living room you've just seen. Remember that you have to mention where each of the following objects was in the living room (as you see them in the photo):*

1. a sofa
2. two black armchairs
3. a rug
4. a set of three small white vases
5. a plant
6. a wooden chair
7. two black-and-white pictures
8. a coffee table
9. a dining table with four white chairs
10. stairs
11. a red lamp
12. a laptop

You've got **15** ***minutes*** *to write your e-mail.*

Appendix L. *Two-part vocabulary test*

Pre-test

Name:__________________________________

VOCABULARY TEST

Part I.

Translate the following groups of words into English.

Group of words	Your translation
encima de la mesa	
en el lado derecho de la cama	
en el centro de la habitación	
cerca de la pared	
al fondo	
delante de la ventana	
enfrente de la cama	
en la pared	
en la esquina derecha	
a lo largo de la pared	
una lámpara (colgada)encima de la mesa	
alrededor de la mesa	

Part II.

Read the following sentences and circle the correct answer:

1. I'm going to put the TV __________ top of the coffee table.

   a) at the b) on the c) at d) on

2. The armchair can go __________ right-hand side of the bed.

   a) at the b) on the c) in the d) to the

3. I want to put the chairs __________ counter ____________ the table.

   a) near to the /around b) near to the / at c) near the /around d) near the / at

4. The stools will go just __________ stove.

   a) opposite to the b) opposite c) opposite the d) opposite of the

5. I would like the table __________ corner of the room __________ back.

   a) in the / in the b) at the / in the c) at the / at the d) in the / at the

6. I'll put the next big piece of furniture on the right __________ wall.

   a) along to the b) along the c) along of the d) along

7. I will hang the lamp __________ middle of the room__________ table

   a) at the / above the b) at the / above c) in the / above the d) in the / above

8. I would hang the pictures __________ wall facing the sliding doors.

   a) at the b) on the c) above the d) to the

9. I will have two chairs __________ the table.

   a) at front of b) at front to c) in front of d) in front to

Appendix M. Affective variable questionnaire and time estimation task

TASK I

I. What are your impressions?

Evaluate the **FIRST TASK** by circling the appropriate answer in each case:

I thought this task was **EASY**	0 1 2 3 4 5 6 7 8 9	I thought this task was **DIFFICULT**
I felt **FRUSTRATED** doing this task and did **NOT** do it **WELL**	0 1 2 3 4 5 6 7 8 9	I felt **relaxed** doing this task and did it **well**
This task was **NOT INTERESTING,** soI **DON'T want** to do **more** tasks like this	0 1 2 3 4 5 6 7 8 9	This task was **INTERESTING, so** I **WANTTO DO** more tasks like this

II. How much time did it take you to complete the task?

Your answer: ________________

CUESTIONARIO

El OBJETIVO de esta encuesta es recoger información sobre tu **EXPERIENCIA** aprendiendo lenguas, en especial **INGLÉS**. Es **IMPORTANTÍSIMO** para nosotros que contestes todas las preguntas con la máxima sinceridad y precisión. La información que te pidamos se tratará de manera **TOTALMENTE CONFIDENCIAL**.

1. Nombre y Apellidos: ______________________________
2. Correo electrónico de contacto: ______________________
3. Fecha de nacimiento (DD/MM/AAAA): ________________
4. UNIVERSIDAD donde estás estudiando: ________________
5. ESTUDIOS universitarios que estás cursando: ____________
6. AÑO de inicio de los estudios universitarios: ____________
7. ¿Cuántos años tenías cuando empezaste a aprender inglés ? ________
8. ¿Has hecho clases extraescolares de inglés ?

   Sí ❑ No ❑

   En caso afirmativo, ¿durante cuantos años? _______ i ¿con qué frecuencia ________ ?

9. ¿Has hecho estancias en un país de habla inglesa ?

   Sí ❑ No ❑

   En caso afirmativo, en la tabla de abajo indica, ¿en qué país has estado y durante cuánto tiempo ?

	País	Duración
1.		
2.		
3.		

10. Has estudiado otros idiomas aparte de inglés ?

    Sí ❑ No ❑

    En caso afirmativo en la tabla de abajo indica, ¿qué idiomas y durante cuánto tiempo?

	Idioma	Duración
1.		
2.		

3.

11. ¿Ves programas de TV y/o películas en versión original?

Sí ❑ No ❑

En caso afirmativo, ¿con qué frecuencia ?

Algunas veces al año	❑	Algunas veces a la semana	❑
Algunas veces al trimestre	❑	Cada día	❑
Algunas veces al mes	❑		

12. ¿Lees textos largos en inglés (libros, revistas, páginas de internet)?

Sí ❑ No ❑

En caso afirmativo, ¿con qué frecuencia ?

Algunas veces al año	❑	Algunas veces a la semana	❑
Algunas veces al trimestre	❑	Cada día	❑
Algunas veces al mes	❑		

13. ¿Intercambias correspondencia (e-mails, cartas, chats) en inglés?

Sí ❑ No ❑

En caso afirmativo, ¿con qué frecuencia ?

Algunas veces al año	❑	Algunas veces a la semana	❑
Algunas veces al trimestre	❑	Cada día	❑
Algunas veces al mes	❑		

14. ¿Hablas inglés fuera de la Universidad ?

Sí ❑ No ❑

En caso afirmativo, ¿con qué frecuencia ?

Algunas veces al año	❑	Algunas veces a la semana	❑
Algunas veces al trimestre	❑	Cada día	❑
Algunas veces al mes	❑		

Especifica, por favor, con quién (colegas, familia, amigos etc.) y con qué finalidad (trabajo, comunicación etc.):

___

___

___

15. ¿Qué factores, crees que han influenciado en tu aprendizaje de inglés ? Puedes marcar más de una casilla.

La edad en que empecé a aprender idiomas	❏
Mi familia	❏
Mis amigos y compañeros	❏
Mi propio esfuerzo y persistencia	❏
Mi facilidad para aprender idiomas	❏
La importancia que tenían los resultados para mí	❏
Mi motivación por aprender idiomas y/o inglés	❏
Mis profesores	❏
Las clases extraescolares de inglés	❏
La estancia en el extranjero	❏

Otros (especifícalo, por favor):

____

____

16. ¿Cuál fue el momento clave en tu aprendizaje de inglés cuando te diste cuenta de que estabas dominando muy bien este idioma ?

____

____

____

____

17. En la escala de **diez** evalúa tu dominio de todos los aspectos de cada lengua que hablas, incluida(s) tu(s) lengua(s) materna(s).

Castellano

Comprensión de lectura	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprensión auditiva	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión escrita	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión oral	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Catalán

Comprensión de lectura	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprensión auditiva	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión escrita	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión oral	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Inglés

Comprensión de lectura	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprensión auditiva	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión escrita	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión oral	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

____

Comprensión de lectura	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprensión auditiva	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión escrita	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión oral	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

____

Comprensión de lectura	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprensión auditiva	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión escrita	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Expresión oral	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Entiendo que la CONFIDENCIALIDAD de mi identidad y información personal como participante en esta investigación queda garantizada, y doy mi consentimiento para el uso de la información de esta investigación EXCLUSIVAMENTE para los propósitos de investigación.

Firma ______________________

MUCHÍSIMAS GRACIAS POR TU COLABORACIÓN!