The Role of Task Complexity and Task Sequencing in L2 Monologic Oral Production

Aleksandra Malicka

THE ROLE OF TASK COMPLEXITY
AND TASK SEQUENCING
IN L2 MONOLOGIC ORAL PRODUCTION

TESI DOCTORAL

presentada per

Aleksandra Malicka

com a requeriment per a l'obtenció del títol de

Doctora en Filologia Anglesa

Directors:

Dr. Roger Gilabert Guerrero

Dr. John Norris

Departament de Filologia Anglesa i Alemanya

Universitat de Barcelona

Dla moich rodziców: Wiesławy i Leszka, i siostry Moniki

ACKNOWLEDGMENTS

Completing this dissertation has been a challenging but rewarding process, and it would not have been possible without the advice, support and help I have received from many people.

First of all, I would like to thank my supervisors, Dr. Roger Gilabert and Dr. John Norris.

It was in one of Dr. Gilabert's classes that I first came across the idea of task-based language teaching. His expertise and passion for this subject sparked my interest in this area and inspired me to pursue it further in my dissertation project. I would like to thank him for believing in my abilities and potential. The knowledge he has very generously shared with me, his commitment to this project, and his guidance, have been invaluable.

During my first research period abroad I had the unique opportunity of working on my dissertation under the supervision of Dr. Norris. I am extremely grateful to him for agreeing to supervise me since then, both in person and from a distance, throughout two continents and three different universities. I am immensely indebted to Dr. Norris for his ideas, his time and his constant interest in my work.

It has been a privilege and an honor to be supervised by such prominent researchers, and I can't express how indebted I am to both of them for the time and effort they have invested in making this project possible. Their enthusiasm, unrelentingly positive attitudes, and constant encouragement have made the whole process a very special and rewarding period in my life. Furthermore, their steadfast rejection of easy solutions, and constant push towards harder work and greater quality have been a great lesson, not just in research but also in life. All in all, I couldn't have asked for a better pair of supervisors.

I would like to extend my gratitude to the members of the GRAL research group, who I had the privilege to work for during my research assistantship period at the University of Barcelona from 2010 to 2013: Dr. Elsa Tragant, Dr. María Luz Celaya, Dr. Joan Carles Mora, Dr. Teresa Navés, Dr. Raquel Serrano, Dr. Imma Miralpeix, Dr. Natàlia Fullana, Dr. Mar Suárez, Dr. Laura Sánchez, Dr. Júlia Barón, and Dr. Ángels Llanes. It was my good fortune to take my first steps as a researcher as part of such an intellectually stimulating group. My special thanks go to the head researcher, Dr. Carmen Muñoz, for taking me on as a research assistant as well as for her kindness, which made my work a real pleasure.

My utmost thanks go to all the students from whom the data was collected, and the teachers who allowed me to do so during their classes: Isabel Gallarde, Raquel Rosa, and Klaus Zilles.

I would like to thank James Pownall for generously agreeing to code the entire data set for accuracy. His help has been invaluable.

Many thanks to the people whose warm welcome made my research stays abroad an unforgettable experience: Lourdes Ortega, Veronika Timpe, Yukiko Watanabe, and Marta González-Lloret.

I would like to thank my dear friend Kasia Paszkiewicz. Our simultaneous pursuit of a PhD meant we could share our joys and our difficulties, making the former greater and the latter easier to deal with.

Thanks also to Eva Cerviño for sharing countless laughs and a few tears, for hearing me out in difficult moments, for her good advice, for motivating me, and for believing in me.

I would also like to thank all those people who have been in my life during this process, and who have given me support and encouragement: Kerry Brennan, Mónica García, Emma Garcia, Antoni Garcia, Maria Grau, Sonia López, Clara Oliveira, Mário Oliveira, Lola Ponce, Marina Ruiz-Tada, Shoko Sasayama, Agata Szarszewska, Brandon Tullock, and Laura Vergés.

My deepest thanks go to Dani Garcia for his constant support, encouragement, and, above all, patience. His day-to-day interest in my work has been invaluable, and his company has very often been an engine pushing me forward in accomplishing my academic objectives. His support during this process has been priceless.

Last of all, I would like to thank my family in Poland: my grandmother Stanisława Socha, my mother Wiesława Malicka, my sister Monika Malicka, and my nephews: Wojtek Janus and Dominik Janus. Spending time with them has been a great source of comfort and energy.

ABSTRACT

In the domain of task-based language teaching (TBLT), researchers have long been interested in exploring the impact of internal task features and conditions on a range of outcomes, such as the occurrence and frequency of conversational episodes (*between-participant interaction*), interlanguage variation at a particular point in time (*performance*), and interlanguage transformation over time (*development*). In the cognitive strand of TBLT explorations, most of the theorizing, and subsequent empirical work, have been guided by the notion of cognitive task complexity, and two particularly influential frameworks have been the Trade-off Hypothesis (Skehan, 1996a, 1998) and the Cognition Hypothesis (Robinson, 2001, 2003). An area which received particular interest from researchers has been determining whether universal task design features exist which systematically influence learners' interlanguage in predictable ways. However, most research carried to date has focused solely on the impact of task complexity by employing a dichotomy of hypothetically simple and complex tasks, rather than a sequence of tasks. Moreover, in the TBLT domain the role of individual differences, for example L2 proficiency, has been a largely underrepresented construct in both conceptual and empirical work.

Given this state of affairs, the objective of the current study was three-fold. First, it aimed to contribute further evidence to the role of task complexity on performance, as measured by general and specific fluency, complexity, and accuracy measures. Second, by employing three tasks of different cognitive complexity levels, rather than a dichotomy, it set out to explore short-term effects of simple-complex task sequencing. Third, it enquired about the role of L2 proficiency by investigating the production of two groups of participants at different stages of competence, as identified through a placement test.

In order to address the aforementioned issues, three tasks of different cognitive complexity levels were developed, identified through Needs Analysis (Long, 2005, 2006), and validated by means of participants' subjective ratings. Cognitive complexity in these tasks was manipulated along two variables form Robinson's (2005, 2007) Triadic Componential Framework: ±number of elements, and ±reasoning demands. The participants in the study (*N*=117), were divided into three groups: 1) simple—complex sequencing (*N*=30), 2) randomized sequencing (*N*=30), and 3) individual task performance, in which different speakers performed

the tasks in its simple, complex, and very complex condition (*N*=18, *N*=19, and *N*=20, respectively). In the sequencing groups, half of the participants were classified as "low proficiency" and half as "high proficiency".

The results of the dissertation have contributed further evidence to the role of cognitive task complexity on performance, with accuracy and lexical complexity being the areas which have shown an increase when task demands were high. The findings revealed a potential role of simple-complex sequencing in promoting more target-like output, but at the same time it was demonstrated that tasks performed in alternative orders presented advantages in other areas of performance: speech rate and lexical complexity. Regarding proficiency, while high proficiency speakers took advantage of increases in cognitive complexity in terms of accuracy, low proficiency speakers did so at the level of structural complexity. The findings obtained were discussed in light of the theoretical task complexity and sequencing models which have guided this work, as well as in light of speech production attention allocation models, and where possible, they were contextualized in light of previous work.

RESUM

En l'àrea de l'ensenyament basat en les tasques pedagògiques (TBLT), els investigadors han explorat l'impacte de les característiques internes de les tasques i les condicions sobre una sèrie de resultats, com ara l'ocurrència i freqüència d'episodis de conversa (*la interacció entre els aprenents*), la variació en la interllengua en un moment particular (*la producció*), i la transformació en la interllengua a llarg termini (*aprenentatge*). En la línia cognitiva de recerca en TBLT, la majoria de les teories, així com les investigacions empíriques, han seguit el concepte de la complexitat cognitiva de les tasques, i en aquest sentit dos marc teòrics de gran influència han estat *Trade-off Hypothesis* (Skehan, 1996a, 1998) i *Cognition Hypothesis* (Robinson, 2001, 2003). L'àrea que ha rebut un especial interès per part dels investigadors ha estat la de determinar si existeixen característiques universals de les tasques pedagògiques que de manera sistemàtica influeixen en la interllengua del parlant. Malgrat això, la majoria de la recerca portada a terme fins ara ha estat enfocada en l'impacte de la complexitat de la tasca mitjançant una dicotomia de dues tasques de nivells cognitius hipotèticament diferents, i no una seqüència de tasques. A més a més, en el camp de TBLT el paper de les diferències individuals, com ara la competència en la L2, és un concepte que no ha rebut suficient atenció, tant a nivell conceptual com en les investigacions empíriques.

Donada aquesta situació, aquesta dissertació té tres objectius. El primer és aportar proves sobre l'impacte de la complexitat sobre la producció mesurada a través de les mesures generals i específiques de fluïdesa, complexitat i precisió lingüística. El segon objectiu és investigar els efectes a curt termini de la seqüenciació de les tasques de simples a complexes, mitjançant tres tasques de nivells cognitius diferents. El tercer objectiu és explorar el paper de la competència en la segona llengua a través de la investigació de la producció de parlants que representen diferents nivells de competència en la L2, prèviament identificats a través d'una prova de nivell.

Per tal d'explorar les qüestions esmentades, es van desenvolupar tres tasques de nivells cognitius diferents, identificades en el procés d'anàlisi de necessitats (anglès *Needs Analysis*; Long, 2005, 2006), i validades a través de les percepcions subjectives dels participants. La complexitat cognitiva de les tasques va ser manipulada segons dues variables del *Triadic Componential Framework* (Robinson, 2005, 2007): ±nombre d'elements i ±raonament. Els participants en l'estudi (*N*=117) van ser

dividits en tres grups: 1) seqüenciació de tasques de simples a complexes (*N*=30), 2) seqüenciació aleatòria (*N*=30), i 3) l'execució de tasques individuals, en la qual diferents participants van fer les tasques en la seva condició simple (*N*=18), complexa (*N*=19), o molt complexa (*N*=20). En cadascun dels grups que van portar a terme seqüències de tasques, la meitat dels participats van ser de "proficiència baixa" i l'altra meitat de "proficiència alta".

Els resultats d'aquesta tesi han aportat proves al paper de la complexitat cognitiva sobre la producció, particularment en les àrees de precisió lingüística i complexitat lèxica. També s'ha mostrat un possible paper de la seqüenciació de tasques en l'ordre de simples a complexes en generar parla caracteritzada per menys errors. Al mateix temps, els participants que van fer les tasques en l'ordre aleatori van mostrar més fluïdesa i complexitat lèxica. Pel que fa a la variable de competència en la L2, s'ha detectat que la complexitat cognitiva creixent presenta avantatges en l'àrea de la precisió en el cas dels participants de proficiència alta, i en l'àrea de complexitat lèxica en el cas dels participants de proficiència baixa. Els resultats obtinguts van ser interpretats en el context dels marcs teòrics pels quals està guiada la present tesi doctoral, així com en el context dels models de producció de la parla i els models d'ubicació d'atenció. Allà on va ser possible, els resultats són contrastats amb recerca prèvia.

TABLE OF CONTENTS

Chapter 1 Overview of the current dissertation 1

1.1. Introduction 1

1.2 The contribution of the current dissertation 4

1.3 The chapters in the current dissertation 5

Chapter 2 Task and task complexity in cognitive psychology 10

2.1 Introduction 10

2.2 Defining “task” 10

2.2.1 Approaches to the study of tasks 11

2.2.2 Dimensions of “task” 12

2.3 Intrinsic versus extraneous cognitive load 17

2.3.1 Mental load versus mental effort 20

2.3.2 Determining mental load 21

2.4 Scaling down the intrinsic load: simple information first 23

2.5 Summary of the chapter 24

Chapter 3 Task, task complexity, and task sequencing in the domain of task-based language learning and teaching 26

3.1 Introduction 26

3.2 Defining “task” across contexts: research, pedagogy, and testing 27

3.3 Early conceptualizations of task complexity and task sequencing 34

3.4 Skehan’s three-way distinction 38

3.5 Skehan's Trade-off Hypothesis: predictions and psychological underpinnings 42

3.6 Robinson's Triadic Componential Framework 48

3.6.1 Task complexity 49

3.6.2 Task conditions 51

3.6.3 Learner factors 52

3.7 The Cognition Hypothesis: predictions 53

3.7.1 Psychological underpinnings of the Cognition Hypothesis: Multiple Attentional Resource Model 55

3.8 Robinson's and Skehan's models juxtaposed 59

3.9 The SSARC model of pedagogic task sequencing 61

3.10 Summary of the chapter 65

3.11 Implications for the current study 66

Chapter 4 Empirical research into the variables ±elements and ±reasoning demands 68

4.1 Introduction 68

4.2 Defining "reasoning demands" and "elements" 70

4.2.1 Reasoning demands 71

4.2.2 Elements 74

4.3 Research into the variables "elements" and "reasoning demands" in the Triadic Componential Framework 77

4.4 Studies in which task complexity was manipulated along ±reasoning demands 78

4.4.1 Reasoning demands: operationalization of cognitivecomplexity 89

4.4.2 Reasoning demands: independent measurement of cognitive complexity differences 91

4.4.3 Reasoning demands: confounding and mislabeled variables 91

4.4.4 Reasoning demands: emerging patterns in findings and conclusion 92

4.5 Studies in which task complexity was manipulated along ±number of elements 93

4.5.1 Number of elements: patterns in findings 103

4.5.2 Number of elements: degree of difference between a simple and a complex task 104

4.6 Studies in which task complexity was manipulated simultaneously along ±reasoning demands and ±number of elements 106

4.6.1 Simultaneous manipulation of ±reasoning demands and ±number of elements: emerging patterns 113

4.6.2 Degree of difference between the simple and the complex task conditions 115

4.6.3 All categories of studies and all types of manipulation: shortcomings and challenges 115

4.7 Summary of the chapter 123

4.8 Contextualizing the current study in light of previous work 123

4.9 The claims under investigation 125

4.10 Research questions and hypotheses 126

4.10.1 Research questions related to the independent variable of cognitive task complexity 127

4.10.2 Research questions related to the independent variable of task sequencing 128

4.10.3 Research questions related to the independent variable of L2 proficiency 129

Chapter 5 Needs Analysis **130**

5.1 Introduction 130

5.2 Theoretical basis for Needs Analysis 130

5.3 Needs Analysis carried out for this dissertation 132

5.4 Interviews 133

5.5 Observations 136

Chapter 6 Methodology **139**

6.1 Introduction 139

6.2 Participants 139

6.2.1 Measuring participants' proficiency 140

6.3 Development of the pedagogic tasks: two tasks versions and two pilot studies 141

6.3.1 First version of tasks: description, piloting, and findings 141

6.3.2 Final version of tasks: scenario, instructions, and operationalization of cognitive complexity 145

6.4 Scenario 145

6.5 Operationalization of cognitive complexity 146

6.5.1 Theoretical level 147

6.5.2 Observational level 147

6.5.3 Operational level 148

6.5.3.1 Reasoning demands 148

6.5.3.2 Number of elements 149

6.5.4 Operationalization of cognitive complexity in the simple task 151

6.5.5 Operationalization of cognitive complexity in the complex task 152

6.5.6 Operationalization of cognitive complexity in the +complex task 153

6.5.7 Summary of cognitive complexity manipulation 154

6.6 Rationale for the language of input 155

6.7 Piloting of the pedagogic tasks: independent task complexity measurement 156

6.7.1 Procedure 157

6.7.2 Results 157

6.7.2.1 Affective variable questionnaire 157

6.7.2.2 Time Judgment Task 158

6.7.2.3 Stimulated recall 159

6.7.3 Pilot study: summary of results and conclusion 160

6.8 The experiment 161

6.8.1 Study design 161

6.8.2 Data collection procedure 164

6.8.3 Independent and dependent variables 166

6.8.4 Measures 166

6.8.4.1 Fluency measures 168

6.8.4.1.1 Unpruned speech rate (rate A) 168

6.8.4.1.2 Pruned speech rate (rate B) 169

6.8.4.1.3 Dysfluency ratio 169

6.8.4.2 Lexical complexity measures 172

6.8.4.2.1 Guiraud’s Index 172

6.8.4.2.2 D 173

6.8.4.3 Structural complexity measures 174

6.8.4.3.1 Words per AS-unit 174

6.8.4.3.2 Clauses per AS-unit 175

6.8.4.3.3 Words per clause 176

6.8.4.4 Accuracy measures 177

6.8.4.4.1 Errors per AS-unit 178

6.8.4.4.2 Errors per 100 words 178

6.8.4.4.3 Target-like use of prepositions 178

6.9 Data transcription and coding 180

6.10 Data screening 180

6.10.1 Techniques 180

6.10.2 Data sets subject to assessing normality 180

6.10.3 Results of data screening: sequencing conditions and proficiency levels 181

6.10.4 Results: baseline data 185

6.11 Statistical tests 187

6.12 Adjustment of the *p* value 187

6.13 Summary of the chapter 188

Chapter 7 Results 189

7.1 Introduction 189

7.2 Organization of the research questions 189

7.3 Abbreviations 190

7.4 Results of the affective variables questionnaire: difficulty perception 190

7.5 The effect of cognitive complexity on performance 193

7.5.1 The effect of cognitive task complexity on performance in the sequencing conditions (simple-complex vs. randomized) 193

7.5.1.1 Fluency 194

7.5.1.2 Lexical complexity 195

7.5.1.3 Accuracy 196

7.5.1.4 Structural complexity 197

7.5.1.5 Summary of results: the effect of cognitive task complexity on performance in the sequencing conditions 199

7.5.2 The effect of cognitive task complexity on performance in the baseline data 199

7.5.2.1 Fluency 200

7.5.2.2 Lexical complexity 201

7.5.2.3 Accuracy 202

7.5.2.4 Structural complexity 203

7.5.2.5 Summary of results: the effect of cognitive task complexity on performance in individual task performance condition 204

7.5.2.6 Summary of the effect of cognitive task complexity on performance in sequencing conditions and baseline data 205

7.6 The effect of sequencing on performance 208

7.6.1 Sequencing tasks from simple to complex versus randomized sequencing 209

7.6.1.1 Fluency 209

7.6.1.2 Lexical complexity 211

7.6.1.3 Accuracy 212

7.6.1.4 Structural complexity 215

7.6.1.5 Summary of results: simple-complex versus randomized sequencing 217

7.6.2 Effect of sequencing on performance: tasks as performed in a simple-complex sequence versus in isolation 220

7.6.2.1 Task 3 as performed in simple-complex sequence and in isolation 223

7.6.2.2 Task 2 as performed in simple-complex sequence and in isolation 225

7.6.2.3 Task 1 as performed in simple-complex sequence and in isolation 227

7.6.2.4 Summary of results: all tasks performed in a sequence versus in isolation 229

7.7 The effect of proficiency on performance 232

7.7.1 The effect of cognitive task complexity in speakers of different proficiency levels 232

7.7.1.1 Fluency 232

7.7.1.2 Lexical complexity 234

7.7.1.3 Accuracy 235

7.7.1.4 Structural complexity 238

7.7.1.5 Summary of results: the role of cognitive complexity in speakers of different proficiencies in the L2 240

7.7.2 The role of sequencing in low-proficiency participants 242

7.7.2.1 Fluency 243

7.7.2.2 Lexical complexity 245

7.7.2.3 Accuracy 246

7.7.2.4 Structural complexity 248

7.7.2.5 Summary of results: the role of sequencing from simple to complex in low-proficiency speakers 249

7.8 Summary of the chapter 249

Chapter 8 Discussion 252

8.1 Introduction 252

8.2 Research questions 1a and 1b: the effect of cognitive task complexity on performance: sequencing conditions and baseline data 253

8.3 Research question 2a: the effect of task sequencing on performance: simple-complex versus randomized sequencing 267

8.4 Research question 2b: Does performing a task in a sequence lead to qualitatively different output compared to performing a task in isolation? 275

8.5 Research question 3a: How does task complexity affect the performance of speakers of different proficiency levels, as measured by CAF? 284

8.6 Research question 3b: Do low-proficiency learners benefit more from the simple-complex sequence than from the randomized one? 293

8.7 Conclusion 296

8.8 Limitations 298

8.9 Practical implications: textbook writer, syllabus designer, and classroom teacher 300

8.10 Theoretical implications for available task complexity and task sequencing theories 302

8.11 Methodological and study design implications 309

8.12 Concluding remarks 312

References 313

Appendix 1 Pilot study 1 329

Appendix 2 Affective variables questionnaire 341

Appendix 3 Simple task 342

Appendix 4	Complex task	346
Appendix 5	+Complex task	350
Appendix 6	Second pilot study Time judgment task and affective variables questionnaire	354
Appendix 7	Context and general instructions to all tasks	355

LIST OF TABLES

Table 1	Definitions of "task" in TBLT literature	30
Table 2	The Triadic Componential Framework for task classification – categories, criteria, analytic procedures, and design characteristics (Robinson & Gilabert, 2007)	48
Table 3	Summary of studies which manipulated the variable ±reasoning demands	80
Table 4	Dependent variables and their measurement in studies which manipulated the variable ±reasoning demands	83
Table 5	Summary of studies which manipulated the variable ±number of elements	94
Table 6	Dependent variables and their measurement in studies which manipulated the variable ±number of elements	96
Table 7	Summary of studies which simultaneously manipulated the variables ±number of elements and ±reasoning demands	108
Table 8	Dependent variables and their measurement in studies which simultaneously manipulated the variables ±number of elements and ±reasoning demands	110
Table 9	Operationalization of cognitive complexity in the current study	150
Table 10	Descriptive statistics (difficulty and mental effort) Means and standard deviations	158
Table 11	Time judgment task. Descriptive statistics (means and standard deviations)	159
Table 12	Experimental design of the study	163
Table 13	Skewness and kurtosis values: sequencing conditions and proficiency levels	183
Table 14	Skewness and kurtosis values: baseline data	186
Table 15	Difficulty perception: sequencing condition, proficiency level, and individual task performance	191
Table 16	Descriptive statistics of tasks: means and standard deviations (sequencing conditions)	193

Table 17	Repeated measures ANOVA for the effect of cognitive task complexity on performance: fluency	195
Table 18	Pairwise comparisons for fluency (Dysfluency ratio)	195
Table 19	Repeated measures ANOVA for the effect of cognitive task complexity on performance: lexical complexity	196
Table 20	Pairwise comparisons for lexical complexity (D)	196
Table 21	Impact of cognitive complexity on performance (accuracy): Wilcoxon signed-ranks test	197
Table 22	Impact of cognitive complexity on performance (structural complexity): Wilcoxon signed-ranks test	198
Table 23	Descriptive statistics of tasks (individual task performance): means and standard deviations	200
Table 24	Impact of cognitive complexity on individual task performance: fluency (Mann-Whitney U-test)	201
Table 25	Impact of cognitive complexity on individual task performance: lexical complexity (Mann-Whitney U-test)	202
Table 26	Impact of cognitive complexity on individual task performance: accuracy (Mann-Whitney U-test)	203
Table 27	Impact of cognitive complexity on individual task performance: structural complexity (Mann-Whitney U-test)	204
Table 28	Descriptive statistics of tasks (simple-complex versus randomized sequencing): means and standard deviations	210
Table 29	The effect of sequencing on performance (fluency): Repeated-measures ANOVA	211
Table 30	The effect of sequencing on performance (lexical complexity): Repeated-measures ANOVA	212
Table 31	Pairwise comparisons of tasks in the simple-complex sequencing condition (accuracy): Wilcoxon signed-ranks test	213

Table 32	Pairwise comparisons of tasks in the randomized sequencing condition (accuracy): Wilcoxon signed-ranks test	214
Table 33	Pairwise comparisons of tasks in the simple-complex sequencing condition (structural complexity): Wilcoxon signed-ranks test	216
Table 34	Pairwise comparisons of tasks in the randomized sequencing condition (structural complexity): Wilcoxon signed-ranks test	216
Table 35	Descriptive statistics of tasks: simple-complex sequencing and individual task performance	222
Table 36	Effect of simple-complex sequencing compared to individual task performance (task 3): Mann Whitney U-test	224
Table 37	Effect of simple-complex sequencing compared to individual task performance (task 2): Mann Whitney U-test	227
Table 38	Effect of simple-complex sequencing compared to individual task performance (task 1): Mann Whitney U-test	228
Table 39	Descriptive statistics of tasks for low and high proficiency speakers: means and standard deviations	233
Table 40	The role of task complexity and L2 proficiency in performance (fluency): Repeated-measures ANOVA	234
Table 41	The role of task complexity and L2 proficiency in performance (lexical complexity): Repeated-measures ANOVA	235
Table 42	Impact of task complexity on high-proficiency learners (accuracy): Wilcoxon signed-ranks test	236
Table 43	Impact of task complexity on low-proficiency learners (accuracy): Wilcoxon signed-ranks test	237

Table 44	Impact of task complexity on low-proficiency learners (structural complexity): Wilcoxon signed-ranks test	239
Table 45	Impact of task complexity on high-proficiency learners (structural complexity):Wilcoxon signed-ranks test	239
Table 46	Descriptive statistics of tasks: means and standard deviations (low-proficiency participants in the simple-complex versus randomized condition)	244
Table 47	Repeated measures ANOVA for low proficiency: fluency	245
Table 48	Repeated measures ANOVA for low proficiency: lexical complexity	246
Table 49	Mann-Whitney U-test for low proficiency: accuracy	247
Table 50	Mann-Whitney U-test for low proficiency: structural complexity	248

LIST OF FIGURES

Figure 1	Theorizing dimensions of performance (Foster & Skehan, 2001)	43
Figure 2	Rate B in sequencing and baseline condition	205
Figure 3	Dysfluency in sequencing and baseline condition	205
Figure 4	Target-like use of preposition in sequencing and baseline condition	206
Figure 5	Guiraud's Index in sequencing and baseline condition	206
Figure 6	D in sequencing and baseline condition	207
Figure 7	Words/clause in sequencing and baseline condition	207
Figure 8	Clauses/AS-unit in sequencing and baseline condition	208
Figure 9	Errors/AS-unit in simple-complex and randomized sequencing	218
Figure 10	Target-like use of prepositions in simple-complex and randomized sequencing	218
Figure 11	Words/clause in simple-complex and randomized sequencing	219
Figure 12	Clauses/AS-unit in simple-complex and randomized sequencing	219
Figure 13	Rate B in simple-complex sequencing and individual task performance	230
Figure 14	Errors/100 words in simple-complex sequencing and individual task performance	230
Figure 15	Target-like use of prepositions in simple-complex sequencing and individual task performance	230
Figure 16	Words/clause in simple-complex sequencing and individual task performance	231
Figure 17	Words/AS-unit in low and high proficiency participants	241

Figure 18 Words/clause in low and high proficiency participants 242

Figure 19 Clause/AS-unit in low and high proficiency participants 242

LIST OF ABBREVIATIONS

CAF	complexity, accuracy, fluency
CH	Cognition Hypothesis
CLT	Cognitive Load Theory
CMC	computer-mediated communication
FTF	face-to-face communication
NA	Needs Analysis
RAN	randomized sequencing
SC	simple-complex sequencing
TBLT	task-based language learning and teaching
TLU_PREP	target-like use of prepositions

CHAPTER 1

OVERVIEW OF THE CURRENT DISSERTATION

1.1 Introduction

For the last few decades, tasks have occupied an important place in the field of Second Language Acquisition (SLA) in general, and in the domain of Task-based Language Learning and Teaching (TBLT) in particular. This is evidenced in the amount of research carried out to date, the number of publications it produced, and in its presence in conference programs in the field of applied linguistics. The amount of generated conceptual and empirical research inspired the emergence of first systematic research syntheses and meta-analyses dealing with task-based research, and more specifically, with the construct of cognitive task complexity (e.g., Jackson, 2013). Not only does it emphasize the importance of TBLT as a branch of applied linguistics, but also strengthens its position as a domain of scientific enquiry in its own right. As stated by Ortega (2004), "Task-based teaching is a burgeoning research area within instructed second language acquisition" (p. 15).

In the thirty years of existence of TBLT as an arena of theoretical and empirical investigations, researchers have set out to explore a very extensive scope of issues, which have enquired about the effects of tasks in relation to three main broad areas: interaction, production, and development. While early work in the 1980s was characterized by empirical investigations into *interactive* dimensions of tasks, and tasks were conceived of merely as vehicles to elicit production, 1990s witnessed an increasing distancing from the interactive research paradigm towards the *cognitive* end of the spectrum. The distinction between the interactive and the cognitive line of research is crucial insofar as it defines the scope of enquiry of the current dissertation: it falls into the cognitive research agenda.

At the conceptual level, the cognitive strand of research has been associated with the work of two prominent authors: Peter Robinson and Peter Skehan. With such seminal work as "*A cognitive approach to language learning*" (Skehan, 1998), or "*The Cognition Hypothesis, task design, and adult task-based language learning*" (Robinson, 2003), these researchers have contributed greatly to the establishment of task-based language learning and teaching as an independent unit of scientific exploration.

These publications are among the ones which most exhaustively depict these authors' stances on "task", task complexity, and a number of related issues, which are portrayed in two major theories: Skehan's *Trade-off Hypothesis* and Robinson's *Cognition Hypothesis*. These hypotheses, to some complementary and mutually exclusive to others, defined the cognitive line of TBLT empirical research, and gave rise to a considerable amount of scholarly work, including journal articles, book chapters, entire books, and master's and doctoral dissertations.

Since the emergence of the cognitive line of research, numerous studies set out to investigate an array of linguistic phenomena in relation to performance as measured in three broad areas: linguistic fluency, accuracy, and complexity (lexical and syntactic). In a broad sense, the scientific enquiry within the cognitive paradigm, from the performance perspective, can be summarized as exploring whether *universal task design features and conditions exist, that affect speakers in systematic ways*, independently from other task influences or external conditions, such as individual differences. This fundamental issue has been explored across a variety of contexts, with different learner populations, and with a variety of study and task designs.

Despite an accumulated body of research, often times the findings within studies have been inconsequential, and the results across studies have yielded, overall, an inconclusive picture of producing a second language under different conditions. While certain trends have started to emerge, giving way to first generalizations about performance patterns in relation to the different task-internal characteristics and task-external conditions, the accumulated knowledge lacks transparency and is not solid enough in order to be applied to practice. The research carried out to date has answered some conceptual and empirical questions, but it has left many unanswered, with new questions arising as research findings emerge.

Within the broad scope of cognitive investigations, an issue which merits attention is that, despite extensive explorations of the construct of cognitive task complexity, most studies carried out to date have employed a simple dichotomy of tasks, hypothesized to be simple or complex manifestations of cognitive complexity levels, and therefore little is known about the cumulative effect of cognitive complexity on performance. This state of affairs is due to the fact that it was not until a few years ago that the first theoretically driven framework for *task sequencing* in TBLT emerged (and perhaps also within the field of applied linguistics): Robinson's (2010) SSARC model of pedagogic task sequencing. Therefore research into longer sequences of tasks is only starting to emerge, with the upcoming publication "*Task sequencing and instructed second language learnin*g" (Baralt, Gilabert, & Robinson, 2014) being, to my knowledge, the first book-long systematic compilation of studies researching the issue of sequencing.

Finally, in the TBLT domain, the issue of individual differences as potentially mediating performance, has been largely at the periphery of theoretical and empirical work. The existing task complexity models essentially do not go beyond stating the obvious: that individual differences make a difference for whatever outcome task-based research sets out to investigate. However, both the theoretical speculations about their impact and subsequent research have been fragmented and characterized by a lack of a consensus about their exact role or impact.

1.2 The contribution of the current dissertation

Given this state of affairs, the objective of the current study is three-fold. First, by employing three tasks of different task complexity levels, and a battery of general and specific production measures, it aims to provide further evidence in favor of or against two hypotheses which have guided most of the empirical research into task complexity: The Cognition Hypothesis and the Trade-off Hypothesis.

Second, to my knowledge, the current study is among the first ones to investigate the ostensibly facilitative role of sequencing tasks from simple to complex, as postulated in Robinson's (2010) SSARC model of pedagogic task sequencing. The dissertation puts this claim to test by analyzing the production of speakers who performed a sequence of tasks in the order of increasing cognitive complexity, that is, from simple to complex, versus those speakers who performed tasks in a randomized order. It also investigates the performance of tasks of different levels of cognitive complexity as performed individually by different speakers, and in a simple-complex sequence.

The third aim of this dissertation is to investigate the role of an individual difference, L2 proficiency, in task-based performance. More specifically, this dissertation enquires about potentially differential ways in which speakers of different competencies in the L2 are affected by cognitive complexity levels. It is also concerned with the ostensibly facilitative role of simple-complex sequencing in low proficiency participants.

1.3 The chapters in the current dissertation

The study at hand is divided into 8 chapters. Chapters 2 and 3 present the theoretical underpinnings of the current work from a broad and narrow perspective. **Chapter 2** addresses the fundamental notions relevant to the current dissertation: task, task complexity, and cognitive load, in the field of cognitive psychology. This chapter is informed by two major conceptual frameworks: dimensions of "task" as proposed by Woods (1986) in his seminal article "*Task complexity: Definition of the construct*", and the Cognitive Load Theory (Sweller, 1999; Pass & van Merriënboer, 1994a). These two frameworks introduce such theoretical notions relevant to the current dissertation as simple versus complex tasks, cognitive load, mental load, or mental effort. This chapter also introduces the idea of sequencing from the perspective of cognitive psychology. While the major goal of this chapter is to shed light on the difficulty implied in conceptualizing these constructs, it is a first step towards understanding them in the domain of task-based language learning in teaching.

Chapter 3 narrows down the scope of theoretical enquiry, and it discusses the concepts of task, task complexity, and task sequencing in TBLT. It defines "task" by making reference to the different definitions available in relevant literature in the domain, according to the purpose for which a task is created: research, pedagogy, or testing. The discussion of the construct of "task" leads to that of cognitive task complexity. The discussion of task complexity commences with first attempts to conceptualize this construct found in the work of such authors as Prabhu (1987) or

Candlin (1987), which serves as a basis for subsequently introducing the hypotheses and framework which the current dissertation draws on: Skehan's (1996a, 1998) Trade-off Hypothesis and Robinson's (2001, 2003) Cognition Hypothesis. A top-down approach is taken when presenting them: predictions of the two hypotheses for performance and interlanguage development are described first, followed by the psychological underpinnings of both, with a special focus on the issue of attention allocation. On the basis of Skehan's and Robinson's hypotheses and frameworks, a discussion follows on how they conceive of the notion of "task" and task-based learning. Criticism is raised regarding the underrepresentation in both models of individual differences, and more specifically, the role of L2 proficiency in task-based performance and interlanguage development.

The last construct dealt with in this chapter is the issue of task sequencing, and it is discussed in light of Robinsons's (2010) SSARC model of pedagogic task sequencing associated with the Cognition Hypothesis. The focal point of this section are the different functions assigned to tasks with varying degrees of cognitive complexity, and the rationale behind these functions.

Chapter 4 is a synthesis of the state-of-the-art TBLT empirical research relevant to the current dissertation. Given that the body of research carried out to date in the TBLT domain is substantial, the scope of studies selected for a subsequent review is narrowed down to those studies dealing with two variables from Robinson's (2005, 2007) Triadic Componential Framework: ±reasoning demands and ±elements. This choice was motivated by the fact that these two variables are the ones pertinent to the task design in the current dissertation. A pool of thirty studies is divided into three categories, depending on the kind of task design manipulation adopted, and the categories are: studies which manipulated the variable ±elements, studies which manipulated the variable ±reasoning demands, and studies which manipulated both of the aforementioned variables within the same task design.

The overarching aim of this chapter is to critically assess the research methodology encountered in them, identifying their shortcomings and challenges. Special emphasis is placed on crucial features of task design, such as how different researchers in the domain operationalized the construct of cognitive task complexity. The second aim was to interpret the degree of difference between what the researchers hypothesized to be "simple" and "complex" tasks, and to assess the soundness of researchers' criteria in establishing task design differences. Finally, the third aim was to capture the emerging patterns in the data in order to track systematicity (or its lack) in findings across literature. Last but not least, this chapter aims to critically assess the performance measures employed in the studies in terms of their number, variety, and adequacy. An evaluative analysis of relevant literature is followed by a contextualization of the current dissertation in light of previous research. The chapter is concluded with the presentation of claims under investigation, research questions, and hypotheses of the current study.

Chapter 5 presents the process of needs analysis (NA) carried out in order to develop the research tasks used in the current dissertation. A brief introduction to the theoretical rationale behind carrying out a needs analysis, in line with Long's (2005, 2006) rationale, is followed by the description of the objectives of this needs analysis, and the methods used. The objective of this chapter is two-fold: 1) it aims to show how the target tasks identified in the NA can be described in terms of research tasks, and 2) how the different components of the performed tasks can be described in terms of potentially manipulable variables.

Chapter 6 presents the experimental design of the current study and the methodology. First comes a description of the participants in the experiment. Given that one of the issues the current dissertation enquires about is task-based performance as mediated by proficiency, a separate section is devoted to how the participants' proficiency level was measured.

This is followed by the description of two pilot studies which intended to validate the hypothesized designed differences in cognitive complexity. To this end, the results of the subjective ratings on an affective variables questionnaire are analyzed. This is followed by the operationalization of cognitive complexity along the dimensions ±elements and ±reasoning demands, drawing on insights from cognitive psychology and the task complexity theories which inform this work. The experimental conditions in the study are then described. The final part of this chapter is concerned with the justification of the employed production measures, data screening process, statistical tools, and the rationale for adjusting the *p* value for multiple comparisons.

Chapter 7 presents the findings obtained in the experiment. Prior to answering the research questions this dissertation set out to investigate, the results of the independent measurement of task complexity differences, carried out as part of the main study, are presented and briefly interpreted. Subsequently, each of the six research questions is answered, with a particular focus on the trends observed in descriptive statistics, and with references to inferential tests. The chapter concludes with a summary of all results.

In **Chapter 8** the obtained findings are interpreted in light of the main theoretical frameworks which guided the current work: the Cognition Hypothesis, the Triadic Componential Framework, and the SSARC model of pedagogic task sequencing. Other work referred to are production models, as well as models of attention allocation during task performance, among others. Interpretations to obtained results are followed by the implications for available task complexity models, methodological implications for task design in the studies in TBLT domain, and practical implications. The chapter concludes with a number of observations and reflections on the current state of TBLT conceptual and empirical investigations, and makes recommendations for its advancement as a realm of scientific enquiry.

CHAPTER 2

TASK AND TASK COMPLEXITY IN COGNITIVE PSYCHOLOGY

2.1 Introduction

In the first introductory chapter it was briefly shown that in the domain of task-based language teaching, researchers have been interested in the notion of cognitive task complexity as a way of investigating, describing and explaining a range of language phenomena. The objective of the general theoretical chapter at hand is two-fold. It defines the notions of "task" and "task complexity", and related concepts: cognitive complexity, cognitive load, and mental effort. It does so from a broad, cognitive psychology perspective. Throughout the chapter the challenges and shortcomings associated with defining each of these notions are discussed. It is a first conceptual step towards understanding their definition and operationalization in a narrower realm of TBLT theoretical and empirical investigations.

2.2 Defining "task"

This section of the chapter defines the concept of "task" as has been advanced in cognitive psychology, making reference to the seminal work by Woods (1986). His article in the area of cognitive psychology: "*Task complexity: Definition of the construct*", extensively referenced in the fields of cognitive psychology, educational psychology, and education, importantly distinguished two approaches to the study of tasks. Secondly, it was an attempt to analyze and describe the theoretical dimensions of "task" as a first step to developing a general *theory of tasks*. This, in turn, had implications for theorizing about a related notion, task complexity, in neighboring disciplines.

2.2.1 Approaches to the study of tasks

Woods (1986) makes an important distinction when it comes to choosing the approach to the study, analysis and description of tasks and their characteristics. Two broad perspectives he distinguishes are the *empirical approach* and the *theoretical approach*. The basic assumption of the empirical approach is that task characteristics are derived from a set of empirical data. A pool of data serves as a first step in determining task features, in the absence of a formal theoretical definition. Tasks are therefore described *inductively*.

On the opposite end of the spectrum is the theoretical approach, in which task characteristics are identified and described in advance of designing a task. Only once their internal features have been thoroughly depicted, are their characteristics measured, and they are subject to subsequent empirical testing. The main difference between the two approaches consists, then, in an *a posteriori* (empirically based) versus an *a priori* (theoretically driven) approach to investigating tasks, their internal features, and conditions.

Woods notes that a fundamental caveat to the empirical perspective is that "task characteristics identified frequently confound task and non-task elements, particularly interactions between task attributes and individual attributes" (p. 61). The empirical approach to researching features of tasks therefore carries the risk of the conflation of different factors, some of them being unique to the internal characteristics of tasks (task effects), and others being external, additional sources of influence (individual effects). This does not allow a description of tasks independently from other sources of influence, or from other external conditions. Woods lists three consequences when the different sources of variation are not disentangled: 1) spurious potential for *construct validity*, 2) a lack of sound basis for the *predictions of task effects*, and 3) low feasibility of *operationalization*. Conflating different sources of influence is the main reason why, as Woods postulates, the study of tasks should be carried out from the theoretical perspective: it keeps task internal features separate from task external ones.

2.2.2 Dimensions of "task"

With the theoretical paradigm in mind, Woods embarked on distinguishing three parameters any task is made up of, or can be broken into: *products, required acts, and information cues*. Let us look closely into the characteristics of each of these building blocks.

Product is the abstract outcome of any task. It can be described in terms of a task goal or objective, such as "building a house", "landing a plane", or "choosing a university degree". In designing any task, the product should be specified prior to identifying the required acts or information cues. There are two dimensions to every product: an object (e.g., a glass) or an event (e.g., serving a customer), and a defining attribute, which can be manifest in quantity, quality, timeliness, or cost.

(Required) acts is a pattern or patterns of behavior "with some identifiable purpose of direction" (p. 64), which are indispensable for task completion. They are actions one must take in order to complete a task, and they can differ in number from few to many, depending on the nature of the defined product.

Finally, *information cues* are "pieces of information about the attributes of stimulus objects upon which an individual can base the judgments he or she is required to make during the performance of a task" (p. 65). Information cues are circumstances or conditions which have an influence on required acts.

On the basis of these three building blocks of tasks, it can be stated that, at an abstract level, a task can be conceived of as an interaction between three factors: its goal, or outcome, actions required in order to accomplish it (a single action or multiple actions), and the conditions which accompany these actions, which can also be individual or multiple. A task's outcome is identified prior to defining the required acts and conditions.

Let us now see how this conceptual approach translates itself into a practical example. In his 1986 article, Woods exemplified the concept of task, and the different abstract factors, with a task performed by air traffic controllers: landing planes. The upcoming paragraphs show how the different components of this task translate themselves into the different theoretical dimensions of task presented above.

In the context of the aforementioned abstract notions, at the product stage, one can conceive of the product of such a task as "landing planes safely", which is, essentially, the outcome or objective of this task. There is a series of *required acts* an air traffic controller must take in order to successfully complete this task. Woods lists a total of six acts, performed in this order: (1) choice of hold pattern, (2) order of landing, (3) instructions to (re)locate in holding pattern, (4) choice of runway, (5) landing instructions to the pilot, and (6) taxiing instructions to pilot. These are the required acts, or "basic units of behavioral requirement" (p. 65), necessary for the completion of this task.

Each of these required acts is associated with possible sources of influence, which take the form of *information cues*. Woods associated multiple cues with each act. Let us focus on one of them, "giving landing instructions to the pilot". It is dependent upon the following possible attributes, or conditions: weather, visibility on ground, ground conditions, and quality of approach. These cues, rather than being unique to this particular required act, can be used to describe some of the other previously mentioned acts. In total, Woods lists six actions needed to be taken (required acts) and seventeen possible sources of influence (information cues), which are associated with the task of landing a plane safely.

Woods' taxonomy assumes that any task can be described in terms of the three components described above, and that such three-way distinction can be a basis for developing a theory of tasks. More importantly, it is believed that the three abstract building blocks of tasks can be used as a way to build a theory of *task complexity,* which, in turn, may serve as a device to describing features which *distinguish a simple task from a complex one*.

Researchers in disciplines such as cognitive psychology and education have long been interested in describing simple versus complex tasks, and the notion of task complexity, at an abstract, conceptual level. These endeavors have shown that, paradoxically, defining the notions of "simple" versus "complex", is, in itself, far from a simple task. As Quesada et al. (2005) stated, "Although complexity is a term used pervasively in psychology and is operationalized in different ways, there are no psychological theories of complexity … the definitions of problem solving, complexity, and complex problem-solving are not well specified. It is difficult to build good theories in a field where the basic definitions are blurry" (p. 5). A sound conceptual framework within which to situate or debate the construct of complexity is therefore lacking. Nevertheless, many disciplines have witnessed attempts to capture the theoretical dimensions of complexity.

Perhaps a good starting point to further debate is the following example by Halford, Wilson, and Phillips (1998). They illustrated the notion of complexity with an example from everyday life in which one must decide on the choice of a restaurant:

> "The more important the occasion, the more expensive our choice of restaurant, though importance might have more influence when we have plenty of money than when we have little. Here we have an interaction between two determining factors. This situation can be represented as a ternary relation, comprising a set of ordered triples in which each amount of money and each level of importance is associated with a restaurant" (p. 805).

What this example cogently illustrates is that a task such as choosing a restaurant is constrained, or determined, by a number of factors. It is suggested that the relative difficulty implied in a task such as the one described above depends upon *the number of intervening factors*, which in this particular task manifest themselves as "importance of the occasion" and "budget".

The complexity of a problem can therefore be conceived of as the function of the number of factors which interact with each other, and they represent a hierarchy in which a binary relation is more complex than a unary relation, but less complex than a ternary relation; a ternary relation is more complex than a binary relation, but less complex than a quaternary relation, et cetera. Similar tasks to the one described above spring to mind easily; so much so that it can be stated that virtually all everyday activities, choices, and problems that an individual or groups of individuals are involved in, are constrained by few to multiple factors, each of them contributing to the overall inherent complexity or difficulty of a situation. Some examples include choosing a holiday destination (potential factors: budget, people in a party, place, means of transport, etc.), enrolling on a language course (potential factors: a center's years of existence, its teaching staff, location, price, size of groups, etc.), or driving to a location (potential factors: weather conditions, familiarity with the route, traffic, road conditions, etc.). In the context of conducting research, one can conceive of an experimental design as a set of related factors: independent and dependent variables. A one-way design is a binary relation, a two-way design is a ternary relation, et cetera. “Experimental designs with more factors permit more complex interactions, but at the cost of more observations (participants)” (Halford et al., 1998, p. 805). In Halford et al.’s (1998) terms, the internal complexity of all these, and multiple other situations, can be described in terms of the number of factors these situations are bound by. Overall, the more factors and relations between them, the more inherent difficulty presented by an activity, task, or problem.

Halford et al.’s (1998) everyday life example of complexity is a first step to analyzing this concept from a more formal perspective. Let us analyze how these and other authors have approached the notion of complexity. According to Halford et al. (1998), “Complexity is defined as the number of related dimensions or sources of variation (…) we have defined cognitive complexity intuitively in terms of the number of interacting variables represented in parallel and have conceptualized it in terms of the number of arguments in a relation” (p. 803). In Sweller et al.’s (1998) terms, “cognitive load is generally considered a construct representing the load that performing a particular task imposes on the cognitive system” (p. 266). While the first of the above definitions is partially overlapping with what was already stated before – namely, that complexity is about “related dimensions” and “sources of variation”, what merits attention in both of the above definitions is that they narrow down the scope of, or define complexity, in terms of its *cognitive* dimension.

In cognitive psychology, the notion of cognitive complexity is typically associated with one of the most influential, robust and heavily researched theories, the Cognitive Load Theory (CLT; Sweller, 1999; Sweller, van Merriënboer, & Paas, 1998; Pass & van Merriënboer, 1994a). It was the first attempt to give a profound understanding of, and a systematic definition of the construct of cognitive complexity. As noted by Pollock et al. (2002), "The Cognitive Load Theory uses some aspects of human cognitive architecture and of the structure of information to provide instructional designs that facilitate understanding, learning and problem solving" (p. 62). The main concern of the Cognitive Load Theory is the design of such instructional interventions, which optimally use people's limited cognitive processing capacity to transfer the acquired knowledge to new settings (Paas, Touvinen, Tabbers, & Van Gerven, 2003). Crucial to the understanding the construct of cognitive complexity in this theory is the distinction between two types of load: intrinsic load and extraneous load. The next section defines each of them.

2.3 Intrinsic versus extraneous cognitive load

CLT distinguishes two broad sources of influence of cognitive complexity in a task, or two types of cognitive load: intrinsic and extraneous.

Intrinsic load is the inherent complexity of instructional material. It can be exemplified, for example, with procedures required in order to solve a mathematical problem. *Extraneous load*, on the other hand, is the load inherent in the instructional format, such as the way in which information is organized in instructions, or practical demonstrations. In the light of CLT, these two sources of influence, rather than operating in separation, interact and contribute to the overall complexity of a problem. The fundamental difference between the two kinds of cognitive load is that whereas the nature of intrinsic load does not lend itself to changes in the way of instructional intervention, extraneous load "is unnecessary cognitive load and can be altered by instructional interventions" (Sweller et al., 1998, p. 259). Extraneous load is, therefore, undesirable, and, as far as possible, it should be maintained at a low level.

While clearly affecting overall complexity in different ways, in cognitive psychology literature these two kinds of load have typically been discussed jointly. However, particularly relevant to the current dissertation is the notion of intrinsic cognitive load, and the remainder of this chapter is devoted to its careful treatment.

As was mentioned before, intrinsic load is related with the inherent complexity of a task itself, and it can fall anywhere on the range of possibilities from "low" intrinsic load, characterized by few demands placed on the task performer, and therefore consuming few resources, and requiring little effort, or "high" intrinsic load, characterized by the exact opposite: heavy demands, many resources required, and a lot of effort. The following definition of intrinsic load by Pollock et al. (2002) helps to understand the difference between a task imposing "high intrinsic load" versus "low intrinsic load":

> "The intrinsic cognitive load of information is determined by the extent to which various elements interact. An element is the information that can be processed by a particular learner as a single unit in working memory" (p. 8).

The key concepts this definition introduces is are "element" and "element interactivity". They are parameters according to which "low" or "high" intrinsic load can be established. Let us commence by defining "element". An element is "anything that has been or needs to be learned, most frequently a schema" (Sweller et al., 1998, p. 259). An example of elements often recalled in CLT literature are vocabulary items and grammar properties in a foreign language.

In light of the CLT, vocabulary items are isolated, separate units, which is the reason why in the learning process they do not interact with each other. In other words, their *element interactivity is low*. The opposite extreme is represented by elements high in interactivity, such as, for example, a syntax of a language, which the CLT conceptualizes as high in element interactivity on the grounds that "we can learn vocabulary items individually but we cannot learn grammatical syntax without considering several vocabulary items" (Pollock et al., 2002, p. 62). Along similar lines, one cannot understand how an electrical circuit works without considering simultaneously multiple components and the relations between them: it is also an example of a task characterized by high element interactivity.

In cognitive psychology, the fundamental reason for why tasks with high element interactivity are considered *complex* is because many different elements have to be processed in a parallel fashion in a learner's mind. Such simultaneous processing of multiple elements poses substantial demands on the task performer, because the different components must be held in memory and attended to at the same time for task execution. On the other hand, tasks characterized by low element interactivity are considered *simple* because the different elements can be processed in a gradual, step-by-step fashion. Such processing poses fewer demands on the speaker because elements are attended to one at a time; that is, an element is taken care of only *after* the previous one has been dealt with. Simple tasks therefore do not imply holding many different components in memory simultaneously, which is what fundamentally distinguishes them from complex tasks.

It can be broadly stated that in cognitive psychology, the parameter of element interactivity is employed to differentiate between a simple and a complex task, and it is one of the fundamental claims of the Cognitive Load Theory. These insights, however compelling, leave a fundamental question unanswered: how much difference in element interactivity should two (or more) tasks exhibit in order to fall into the "simple" versus "complex" category. This issue boils down to the following question: *how can we objectively measure the complexity of a task*? While the cognitive psychology literature does not provide a straightforward answer to this question, researchers in this discipline have set out to investigate this issue using a variety of methods and criteria. Before presenting these, a crucial distinction must be made between two constructs: mental load and mental effort.

2.3.1 Mental load versus mental effort

The Cognitive Load theory states that intrinsic load is made up of two sources of influence: a task-based dimension (mental *load*) and a learner-based dimension (mental *effort*). As was the case with intrinsic and extraneous load, these two dimensions of cognitive load also interact and affect the overall complexity of a problem or task. However, given that they fall into different conceptual categories (task effects and non-task effects), they must be distinguished from each other.

Mental load "refers to the load that is imposed by task (environmental) demands" (Sweller et al., 1998, p. 266). This kind of load stems from the interaction between two factors: task and subject characteristics. Following Paas et al. (2003), "it provides an indication of the expected cognitive capacity demands and can be considered an a priori estimate of the cognitive load" (p. 64). It is therefore concerned with *the demands that a task imposes on a task performer*. The latter, *mental effort*, refers to "the amount of cognitive capacity or resources that are actually allocated to accommodate the task demands" (Sweller et al., 1998, p. 266). It *therefore reflects the actual amount of cognitive load* that a task imposes on a speaker.

Linking the concept of mental load to the previous explanation of element interactivity, it can be stated that tasks characterized by high element interactivity impose a high mental load, and on the other hand, tasks low in element interactivity impose a low mental load. This relationship is a deceptively straightforward one in that the mental effort is often times not commensurate with mental load. As stated by

Sweller et al., (1998), "The question of how to determine cognitive load is difficult for researchers, because of its multidimensional character and the complex interrelationships between performance, mental load, and mental effort" (p. 266).

2.3.2 Determining mental load

Cognitive psychologists have long been interested in how to determine the amount of cognitive load, and they have suggested a number of ways in which this can be accomplished. Wierwille and Eggemeier (1993), among others, listed three types of methods identified in cognitive psychology research, and these are: subjective, physiological, and performance-based indices.

Subjective methods spring from the idea that humans have the capacity of retrospectively reflecting about cognitive processes, and assessing the amount of mental effort invested in performing a task. Such assessment in cognitive psychology has been carried out by means of assigning a numerical value to the mental effort represented by a task. This method is typically carried out using a rating scale, and the task performer is asked to mark the numerical value which best reflects the effort they invested in task execution. It is therefore a subjective post task-performance technique, in which the participant judges the perceived difficulty of a task. Within the group of subjective techniques, a different method is that of stimulated recall. In this method the participant reflects verbally on the difficulty of a task.

The second group of methods are *physiological metrics*. They include methods such as eye-tracking, respiration rate, changes in heartbeat, or changes in pupil diameter. The underlying assumption is that the variation at physiological level reflects changes at the level of cognitive functioning (Sweller et al., 1998). It is on these grounds that these methods are believed to be good predictors of the demands a task poses on a learner.

Performance-based measures are concerned with the measurement of the main, primary task a participant is exposed two. Within performance-based measures a distinction can be made between measuring the primary task by means of, for example, the overall time invested in performing it, the speed of task delivery, or the number of errors. A different technique is that of measuring the primary task *in relation to an additional task*. Following Cegarra and Chevalier (2008), “The idea underlying measurement of an additional task is that the capacity that is not being used to perform the primary task can be used to perform another task” (p. 989). The fundamental idea behind this technique is that two tasks are performed within the same time frame, one being the primary task and the other one the secondary task, and they are performed in a parallel fashion. A real-life example is driving a car at the same time as holding a conversation. Research into this dual-task paradigm has typically required a participant to perform the primary, or main task, and their performance was interrupted by some sort of signal, such as a tone or a flash of light. Reacting to this interruption is what constitutes the secondary, or additional task, and a response can take the form of, for example, hitting a key on the keyboard or clicking with a mouse. The time it takes to react to the secondary task is reflective of the amount of cognitive resources invested in the performance of the main task (Gwizdka, 2010). Unlike subjective methods associated with assigning a numerical value to the mental effort, the latter techniques (physiological and performance-based) are *objective* ways of measuring mental load.

2.4 Scaling down the intrinsic load: simple information first

Given that some tasks are inherently characterized by high element interactivity, and therefore pose a substantial mental load, an area of scientific enquiry in cognitive psychology has been that of *how the intrinsic load of complex tasks can be reduced*. As Bannert (2002) noted, “sequencing theories … deal with appropriate information sequencing in instruction so as to optimize learning” (p. 143). In other words, there has been a concern regarding how cognitive load can be managed during tasks and what scaffolding techniques can be employed in order to accomplish it. One approach identified in cognitive psychology literature is that of *sequencing information bits from simple to complex*. Following Van Merriënboer et al. (2003),

> “The first approach identifies simple-to-complex versions of the whole task to decrease *intrinsic cognitive load;* novice learners start to practice on the simplest version of the whole task encountered by experts in the real world and progress toward increasingly more complex versions (…). It is clearly impossible to use highly complex learning tasks from the start of a course or training program because this would yield excessive cognitive load for the learners, with negative effects on learning, performance, and motivation (…). The common solution is to let learners start their work on relatively simple learning tasks and progress toward more complex tasks” (p. 6).

Two important implications follow from the above quotation: 1) intrinsic cognitive load must be decreased for the learners, and 2) given a high cognitive load of complex tasks, any instructional design must commence with simple tasks before moving on to the complex ones. One of the suggested approaches to scaling down the cognitive complexity is offered by elaboration theory (Reigeluth 1999b). It proposes sequencing on the basis of simplifying conditions. Van Merriënboer et al. (2003) provide an example of how the complexity of the “literature search task” can be scaled down by keeping the complexifying factors low in beginning stages, and progressively scaling them up. This task can be made more or less complex along such factors as 1) clearness of the concept definitions within a domain or across domains (*clear vs. unclear*), 2) scope of literature available on a topic (*small vs. large*), 3) number of relevant domains and databases, and how familiar these are to the task performer (*one familiar domain and/or database vs. many unfamiliar domains/ databases*), 4) search type (*keywords vs. abstracts*), and 5) number of search terms (*few vs. many*).

Taking into consideration the above factors, a *simple* “literature search” task is one performed in a familiar domain, with clearly defined concepts, and using few search terms which return few publications. A complex task represents the opposite extreme. In between the two extremes of complexity, there are several tasks with differing levels of cognitive complexity which are developed by modifying one or a few of the simplifying conditions, with simple tasks representing a low intrinsic load, and complex ones representing a high intrinsic load.

In general terms, scaling down the cognitive complexity of a task with a high intrinsic load is a scaffolding strategy employed in order to assist the learner in progressing from cognitively simple to cognitively complex task versions.

2.5 Summary of the chapter

In this chapter, the notions of “task” and “task complexity” were defined in the broad context of cognitive psychology, a neighboring discipline to Applied Linguistics, and to the domain of task-based language learning and teaching. It was shown that the main concerns of cognitive psychologists have been to define the notion of task itself, to draw the distinction between what is “simple” versus “complex”, and the taxing issue of defining task complexity. Finally, it could be observed how cognitive complexity of a task can be scaled down by means of presenting simple information bits prior to the complex ones, and how a real-life academic task (“literature search”) falls into the simple-complex distinction. With all of the above concepts, it was made clear that these notions present different degrees of challenge and shortcomings at the conceptual level.

This chapter was a first theoretical step towards understanding the above mentioned phenomena: task, task complexity, cognitive load, and partially also sequencing, in TBLT. The objective of the next chapter is to present them in this domain.

CHAPTER 3

TASK, TASK COMPLEXITY, AND TASK SEQUENCING IN THE DOMAIN OF TASK-BASED LANGUAGE LEARNING AND TEACHING

3.1 Introduction

The previous chapter took a broad perspective on the issue of task, task complexity, and sequencing, given that these notions were presented in the neighboring discipline of cognitive psychology. The chapter at hand takes a more specific perspective, and narrows down the different approaches to theorizing about the notions of task, task complexity, and task sequencing, to the domain of task-based language teaching. The information presented in this chapter is divided into two parts. The first one is concerned with early attempts to arrive at theoretical definitions of task complexity, and the criteria suggested for subsequent sequencing decisions. This part presents and critically assesses the work of Prabhu, Brown et al., Candlin, Brindley, and Nunan. The work of these authors was the first theoretical step in the development of more systematic approaches to the construct of complexity.

The second part of this chapter deals with two hypotheses which directly inform the current dissertation, and which were briefly advanced in the first chapter: Skehan's (1996a, 1998) Trade-off-Hypothesis, and an associated three-way schematic distinction between different sources of influence, and Robinson's Cognition Hypothesis (CH; 2001a, 2003). Associated with the latter hypothesis, this chapter also presents the Triadic Componential Framework (2005, 2007), and Robinson's (2010) proposal for sequencing, the SSARC model of pedagogic task sequencing. While this chapter argues that the last two authors' work, and particularly that of Robinson, presents a conceptual advancement over previous proposals in TBLT, several unresolved issues and challenges are reflected upon.

3.2 Defining "task" across contexts: research, pedagogy, and testing

Before moving on to the past and current theorizing about task complexity and sequencing, the notion of "task" in the TBLT domain is presented. While in the previous chapter the notion of "task" was presented from a very broad cognitive psychology perspective, this chapter presents it in a narrower context: as a construct employed in task-based research, pedagogy and testing.

The notion of "task" is one of the most widespread conceptual ideas in task-based approaches to instruction. Over the years, it has sparked tens of operational definitions, which depend on three main factors: who designs a task, the purpose it is created for, and who performs it. Given these considerations, any attempts to theoretically classify tasks are incomplete without limiting the scope of debate to a specific context. In conceptual and empirical TBLT research alike, authors frequently draw a distinction between tasks as used for instructional purposes (*pedagogic tasks*) and tasks as used for research purposes (*research tasks*). This distinction was explained by R. Ellis (2000) in the following way: "Task is both a means of clinically eliciting samples of learner language for purposes of research (Corder, 1981), and a device for organizing the content and methodology of language teaching (Prabhu, 1987)" (p. 194). A more detailed account of this distinction is offered by Bygate, Skehan, and Swain (2000a):

> "task is viewed differently depending on whether the perspective is that of research of pedagogy. Researchers, for example, may view a task in terms of a

set of variables that impact on the performance and language acquisition whereas teachers see it as a unit of work in an overall scheme of work" (p. 110).

A task can therefore be conceived of differently depending on the general purpose it serves. In both contexts – research and pedagogy – it is viewed *as a device to elicit speech*, but the latter context is characterized by using task as a unit of classroom procedure, possibly implemented to provide opportunities for oral practice; thus, it is learner-oriented and has some high-level pedagogic goal. In contrast, research tasks do not go beyond eliciting speech, and particularly from the cognitive standpoint, they are often conceived of as a vehicle for manipulating different task design characteristics and features.

From the pedagogical perspective, a further distinction can be made between "task" as a free-standing unit of classroom activity, administered and performed within a specific time period, possibly being complementary to the main textbook material. From this perspective, a task is one of the tools of pedagogic intervention, incorporated into an otherwise non task-based curriculum. From a wider pedagogical perspective, a task can be a unit around which a language syllabus is organized, with tasks being the main focal point of classroom proceedings. In both senses, a task is a means to achieving some sort of communicative goal, in the narrow sense providing a stimulus for meaningful language practice, and in the broad sense (task-based syllabus or language program), it is possibly a means to achieving a specific threshold of L2 competence.

While the above distinction between pedagogic and research tasks is a frequently referenced one in conceptual and empirical literature, none of the definitions presented above contemplates the third general purpose of tasks: their use in the contexts of task-based assessment. In the testing context, "task" has been defined in the following way by Long and Norris (2000):

> "Task-based language assessment takes the task itself as the fundamental unit of analysis motivating item selection, test instrument construction, and the rating of task performance. Task-based assessment does not simply utilize the real-world task as a means for eliciting particular components of the language system, which are then measured or evaluated; instead, the construct of interest is performance of the task itself" (p. 60).

In the testing context, "task" is conceived of as a means to elicit production, and its outcome (which can take the form of oral or written production, reading or listening comprehension) is juxtaposed with standards according to which a learner at a given stage of L2 competence is expected to perform. In other words, a "task" in the testing context is a way of verifying learner's linguistic and other skills at a particular point in time against a set of abstract criteria. In a wider sense, a task is the main building block of a larger-scale assessment program.

The above distinction contextualized "task" in three broad areas: research, pedagogy and testing, but it has not defined "task" in the domain. Years of research into TBLT literature have generated multiple definitions of this construct, offering something of the order of tens of definitions of this construct. Selected definitions available in TBLT literature are summarized in Table 1 (pp. 31-32).

Such multiplicity of definitions of a single construct presents advantages in some senses, and handicaps in others. On one hand, the existence of multiple definitions reflects the multifaceted nature of "task", and does not limit its scope to a particular situation or context, and very importantly, defines it in a broad and narrow sense. On the other hand, the more conceptual approaches to this construct, the less unanimity as to what a "task" is can be generated, making it potentially challenging to compare one task with another in research, pedagogy, and testing terms. From the latter perspective, drawing the line between a task and a different unit may become a blurry conceptual area.

Table 1. Definitions of "task" in TBLT literature

Long (1985)	A piece of work undertaken for oneself or for others, freely or for some reward. Thus, examples of tasks include painting a fence, dressing a child, filling out a form, buying a pair of shoes (…). In other words, by "task" is meant the hundred and one things people *do* in everyday life, at work, at play, and in between. Tasks are the things people will tell you they do if you ask them and they are not applied linguists.
Richards, Platt, and Weber (1985)	An activity or action which is carried out as the result of processing or understanding language, i.e. as a response (…) Tasks may or may not involve the production of language. A task usually requires the teacher to specify what will be regarded as successful completion of the task.
Crookes (1986)	A piece of work or an activity, usually with a specified objective, undertaken as part of an educational course, or at work.
Candlin (1987)	One of a set of differentiated, sequeanceable problem-posing activities involving learners and teachers in some joint selection from a range of varied cognitive and communicative procedures applied to existing and new knowledge in the collective exploration and pursuance of foreseen or emergent goals within a social milieu.
Breen (1987)	Any structured language learning endeavor which has a particular objective, appropriate content, a specified working procedure, and a range of outcomes for those who undertake the task. "Task" is therefore assumed to refer to a range of workplans which have the overall purpose of facilitative language learning from the simple and brief exercise type, to more complex and lengthy activities such as group problem-solving or simulations and decision-making.
Prabhu (1987)	An activity which required learners to arrive at an outcome from given information through some process of thought and which allowed teachers to control and regulate that process was regarded as a task.
Skehan (1998)	A task is an activity in which meaning is primary, there is some communication problem to solve, there is some relationship to real-world activities, task completion has some priority, and the assessment of the task is in terms of outcome.
Carroll (1993)	Any activity in which a person engages, given an appropriate setting, in order to achieve a specifiable class of objectives.
Bachman and Palmer (1996)	An activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation.
Willis (1996)	Activities where the target language is used by the learner for a communicative purpose (goal) in order to achieve an outcome.

Table 1 continued.

Bygate et al. (2001)	An activity which requires learners to use language, with emphasis on meaning, to attain an objective.
Ellis (2003)	[task] involves a primary focus on meaning (…), it involves real-world processes of language use (…), it can involve any of the four language skills (…), it engages cognitive processes (…), and it has a clearly defined communicative outcome.
Nunan (2004)	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 form.
Littlewood (2007)	An array of learning activities from the simple and brief exercise type to more complex and lengthy activities such as group problem solving or stimulations and decision-making.
Swales (2009)	One of a set of differentiated, sequenceable goal-directed activities drawing upon a range of cognitive and communicative procedures relatable to the acquisition of pre-genre and genre skills appropriate to a foreseen or emerging socio-rhetorical situation.

Several general observations can be made about the definitions of “task” offered in table 1: 1) Some definitions are largely overlapping; 2) some definitions distinguish between tasks which require the production of speech, whereas other do not; 3) still others explore this construct from the point of view of a learning goal, while others focus on task as a vehicle for communication. Regardless of the scope of definitions, their goal, and other considerations, what these definitions have in common is the idea that *a task is a goal-oriented meaningful activity*. In other words, tasks are not created or performed “for the sake of it”; on the contrary: every task has a clearly identifiable objective. In what follows, the different conceptualizations of “task” are presented in a top-down manner.

The very general and oft-cited definition by Long (1985) includes under the term “task” even those communication situations in which communication is, in fact, very scarce or inexistent. This situates tasks within the realm of real-word activities or actions, which one gets involved in. In Long’s stance, using language to transact a task can be obligatory for task transaction or optional; still other tasks do not require it whatsoever. A fundamental idea in Long's definition is that *tasks should resemble real-life situations*.

A narrower and more specific definition of task is that provided by Ellis (2003) who labels task as a "work plan" and he points its underlying features: "it [a task] involves a primary focus on meaning (…), it involves real-world processes of language use (…), it can involve any of the four language skills (…), it engages cognitive processes (…), and it has a clearly defined communicative outcome" (p. 10). This definition, compared with Long's, is limited to tasks in which language use is obligatory, which suggests that communicativeness is one of the prerequisites for a unit of activity to be labeled as "task".

To Skehan (1998), a task is an activity in which "meaning is primary, there is some communication problem to solve, there is some relationship to real-world activities, task completion has some priority, and the assessment of the task is in terms of outcome" (p. 85). The main idea behind performing and completing a task, then, does not lie in how *accurately* the task is accomplished, but in how successful the performer is in conveying the message. In other words, the main focus is on *meaning*, not on *form,* the idea shared also by Nunan (2004). He defines a task as "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 form" (p.76). Along the same lines, Van de Branden (2006) remarks that task-based syllabuses do not divide language into small parts, smaller units of analysis. This kind of syllabus takes "holistic, functional, and communicative tasks, rather than any specific linguistic item, as the basic unit for the design of educational activity" (p. 34).

The aforementioned ideas concerning *task* have explored it from the point of view of the criterion of communicativeness. Littlewood (2007), however, suggests that communicativeness is only one of the possible criteria for classifying tasks. He refers to the concepts of task put forward by Breen (1987), who depicts a task as an array of learning activities which range from simple and brief exercises to complex and long activities, as exemplified by group problem solving or decision-making tasks. The two authors make a distinction between "communication tasks", in which meaning is prioritized over form, and "enabling tasks", whose primary focus is on linguistic aspects. Although, as Littlewood (2004) suggests, communicativeness must not necessarily be the ultimate criterion for tasks, tasks are units which provide a link between outside-classroom reality and inside-classroom pedagogy.

A common ground for all the conceptualizations is that they *emphasize meaning*. In this sense, in the previously described contexts: pedagogy, research, and testing, a task is fundamentally different from an "exercise" or an "activity". However, the above discussion also showed that the concept of "task" is not very clear-cut. It may encompass any human activity and it defies clear definition, which is evidenced in it being subject to multiple interpretations.

As was advanced in the previous chapter, one of the challenges associated with the notion of tasks, beyond defining them, is that of establishing a range of difficulty, on which some tasks fall into the "simple" category, whereas others into the "complex" one. This categorization is the object of the next section: it presents how the notion of "simple" versus "complex" tasks was defined in the TBLT domain, which, in turn, leads to the concept of task complexity and task sequencing, that is, establishing the criteria according to which tasks can be presented in a certain order.

The focal point of this chapter is not to present previous models in an exhaustive way, but rather to serve as a starting point in the discussion of sequencing, and to present the evolution of theoretical criteria which guided task selection and sequencing decisions. The main objective of including them here was to juxtapose the proposals which directly informed the current study with early work in this area.

3.3 Early conceptualizations of task complexity and task sequencing

The theorizing about the notions of “simple” versus “complex” dates back to early 1980s, and one of the first classifications was that offered by Brown et al. (1984). Brown et al. suggested basing sequencing criteria on the level of abstractness inherent in a task. In this proposal, abstract notions (such as argumentation or justification) were thought to be “more difficult” than dynamic relationships, and dynamic relationships were more so than static relationships. Static relationships referred to describing relationships among objects, and it was suggested that the number of objects involved and the difficulty of the relationship between them were factors determining relative difficulty. Describing dynamic events, such as referring to actions, activities or processes, was categorized as more difficult. Finally, “tasks which require the speaker to communicate abstract notions, for instance in argument or justifications, are more difficult again” (Brown et al., 1984, p. 51).

Along similar lines, one of the early criteria was based on task types. Prabhu (1987) in the Bangalore Project assessed different types of tasks potentially representing different levels of challenge for the speaker. A hierarchy was suggested in which opinion gap tasks presented a greater challenge than reasoning gap tasks, and reasoning gap tasks presented a greater challenge than information gap tasks. The lowest level in this hierarchy was attributed to simple information transmission, followed by information transmission requiring inferencing, deduction, and potentially other mental processes. The greatest challenge was attributed to those tasks which dealt with preferences, attitudes, feeling, and beliefs. As pointed out by Robinson (2001), this early distinction was essentially based on open versus closed tasks.

Late 1980s witnessed a conceptual development in theorizing about the notion of task complexity and sequencing, given that somewhat more tangible criteria started to emerge. Candlin (1987) identified six parameters along which sequencing decisions can be taken. Each will be introduced in turn. *Cognitive load* referred to clear, logical sequence versus lack of it, individual versus multiple characters, and individual versus multiple actions. *Communicative stress* involved such factors as topic familiarity (known vs. unknown) and number of interlocutors (one or few vs. many), and a series of interlocutors' characteristics: their level of communicative competence, their familiarity with the subject matter, and the extent to which a task followed a clear, structured organization. *Code complexity and interpretive density* was concerned with the complexity of texts used in tasks, in linguistic and argumentation terms. *Content continuity* referred to how much a pedagogic tasks resembled a real world target task. *Process continuity* was related to learners' choice about sequencing tasks. Finally, *particularity* and *generalizability* referred to the relative familiarity or unfamiliarity of a situation.

In a similar vein, Brindley (1987) proposed a three-way distinction between the following groups of factors: learner factors, task factors, and text factors. *Learner factors* encompassed such elements as confidence and motivation, prior learning experience, ability to learn at the pace required, possession of necessary language skills, and relevant culture knowledge. *Task factors* included the degree of cognitive complexity, number of steps, amount of context support, and amount of time provided. Finally, *text factors* were concerned with the characteristics of texts used in tasks, such as length, clarity, and familiarity.

Similar to Brindley's account is that of Nunan (1989), who put forward a set of criteria including *input factors*, which encompassed such aspects as grammatical complexity of the input (simple sentences vs. sentences containing subordination), length of a text, propositional density, the amount of low-frequency vocabulary, the speed of spoken texts, the number of speakers involved, the explicitness of the information, the discourse structure, and the clarity with which it is signaled. *Learner factors* were related to the reality which the learner brought to the task, and it encompassed such components as background knowledge, linguistic knowledge, confidence, interest, motivation, and observed ability in language skills. Finally, *procedural factors* were "operations that the learners are required to perform on input data" (Nunan, 2004, p. 122), and they were made up of such components as relevance (relative relevance of the task to the learner), complexity (steps involved, complexity of instructions, cognitive demands, amount of information), context provided (required prior knowledge, preliminary activity), processibility of language of the task (processing capacity required by the task, whether a task requires the use of specific forms), available help (assistance from teacher, peers, or materials, and interlocutor solidarity), degree of grammatical complexity (required complexity of structures), available time (time pressure, planning time), and follow-up (debriefing and feedback).

As can be observed, the last three of the presented proposals (those by Candlin, Brindley and Nunan) contained a lot of overlapping elements, and in several aspects they simply employed different nomenclature to denote the same concept. A case in point are the constructs of cognitive load, task factors, and input factors present in these authors' proposals: they all refer to the amount of mental effort needed to perform a task. The three of them also contemplated the factors which learners bring to a task, as well as a battery of other unrelated considerations potentially influencing overall difficulty, which defy a clear categorization as a single group of factors, especially those in Nunan's proposal (e.g., the concept of available help). Also, some groups of factors seem to have conflated manipulable and non-manipulable elements (e.g., topic familiarity and number of interlocutors versus interlocutors' characteristics, referred to as *communicative stress* in Candlin's proposal).

On the basis of the above analysis it can be concluded that these first attempts at establishing criteria to sequencing tasks conflated factors clearly belonging to different categories, some perhaps being well suited to deliberate manipulations (such as cognitive load - changing the number of characters or parallel actions, code complexity, text factors, or procedural factors), and other being external sources relatively alleviating or adding to the overall cognitive complexity of tasks. For example, while cognitive load (task factors or input factors) potentially lent itself to being manipulated, in the case of several other factors it is challenging to conceive of exactly how they could inform sequencing decisions, given that they were not predictable in advance of instruction (such as the factors having to do with what the learner brings to a task). All these groups of factors influence performance in unique ways, but if sequencing decisions were to be made upon them, they should be disentangled and considered in separation. This obvious caveat was one of the main points of Robinson's (2001a) criticism of these frameworks: they did not make a distinction between *task complexity* and *task difficulty*. Beyond being an issue of nomenclature, this distinction is of utmost importance given that in current approaches to TBLT these designate different concepts, with task difficulty being an area of investigation in its own right.

Despite clear shortcomings and somewhat fragmented ideas, these first conceptualizations led to the emergence of one of the proposals which informed large portions of current TBLT investigations: Skehan's (1996a, 1998) three-way distinction and the Trade-off Hypothesis. One of the elements which clearly distinguishes these proposals from the early ones is that, aside from merely establishing criteria, they link them to *performance, understood as the interplay of fluency, accuracy, and complexity, and to interlanguage development*. The next section describes them in detail, starting with predictions, through psychological underpinnings, and the role of "task" and task-based instruction.

3.4 Skehan's three-way distinction

Skehan's stance partially draws on the proposals of Nunan (1989) and Candlin (1987). Similarly to these authors, he proposed a distinction between three groups of factors which contribute to the overall complexity of a task: *code complexity* (language required), *cognitive complexity* (thinking required), and *communicative stress* (performance conditions). Taken together, these factors affect the overall difficulty of a task.

The first group of factors, *code complexity*, is concerned with the linguistic demands the task places on the learner, and it deals with the linguistic resources, and other knowledge types, needed for task completion. The basic assumption is that the more "advanced" structures a task requires in order to be completed, and the more linguistic resources are necessary for task transaction, the more complex a task is.

An area in which this group of factors presents a shortcoming is the notion of "advanced" or "complex" structures, which are largely unspecified. It is not articulated what differentiates an advanced structure from a non-advanced one, or on what theoretical grounds such a distinction can, or should, be made. It is therefore subject to discussion to what extent a task can be classified as "simple" or "complex" on the grounds of linguistic complexity: for example, is narrating in past progressive more complex than narrating in present progressive, or are they equally complex because they imply the use of the progressive rather than the simple verb form. As Skehan (1998) noted, "Language is simply seen as less-to-more complex in fairly traditional ways since linguistic complexity is interpretable as constrained by structural syllabus considerations, or developmental sequences" (p. 99).

The second group of factors in his framework – *cognitive complexity* – is concerned with the degree of cognitive complexity of a task's content and ways in which complexity can be manipulated. Cognitive complexity can take on different representations, which in Skehan's framework are grouped into two broad categories: *cognitive* familiarity, including familiarity of topic, familiarity of discourse genre, and familiarity of task, and *cognitive processing*, including information organization, amount of computation, clarity of information given, and sufficiency of information. These different considerations can manifest themselves in a variety of ways in the design of a task. In general terms, *cognitive processing* is associated with the amount of online computation a task requires (little vs. a lot), and cognitive familiarity is concerned with ready-made solutions, or at least recognizable transactional patterns that scaffold the access to relevant knowledge under performance conditions, available during task performance.

Following Skehan's categorization, a "simple" task is one which, for example, requires the speaker to talk about specific people or events (as opposed to abstract ones), requires the speaker to narrate a story with a clear storyline (as opposed to dealing with multiple narrative components), a task which the speaker is familiar with (as opposed to a new, unfamiliar task, or task type), and one which makes available all the information necessary for task completion (as opposed to unavailable information).

A closer look at "cognitive complexity" reveals that perhaps the factors it embraces belong to different conceptual categories. Whereas storyline complexity, or a task's level of abstractness, can be classified under the generic term "cognitive complexity", it is highly debatable whether certain task types are more cognitively demanding than others, and if so, on what grounds (Skehan & Foster, 2001, provide an example of a "riddle" and "jigsaw" as examples of tasks which consume more attentional resources than narrating "a simple story").

Finally, the third set of factors, *communicative stress*, is related to the pressure of communication caused by a task, and it includes factors such as *time pressure* (how quickly the task has to be done), *modality* (tasks which involve either speaking and writing or listening and reading), *scale* (e.g., number of participants, number of relationships, etc.), *stakes* (the importance of performing the task and doing it correctly), and finally, *control* (the degree of freedom given to the learner regarding task execution).

A task simple along the aforementioned variables is one performed with no time limit (as opposed to the pressure of performing a task within a specific time frame), there is one holder of information (as opposed to information split between the different task participants), performing the task has little importance (as opposed to tasks with high stakes), and the learner has an influence on how the task is implemented (as opposed to imposed task input or procedures accompanying its execution). To my mind, some factors which belong to this last category miss specification. For example, it is underspecified whether they add to a task's difficulty or alleviate it, for example the factor "scale": it is unclear what impact having two interlocutors versus multiple interlocutors has on overall complexity. Skehan also seems to be making the point that tasks which combine two input modalities (reading and listening), elsewhere referred to as "integrated tasks"[1], place heavier performative demands than those which require one input channel (either visual or auditory). However, the rationale for such distinction is lacking in his framework.

As can be observed, many issues in Skehan's proposal, as in the previous ones, are subject to debate. Nonetheless, it is claimed that, taking all of the above factors into consideration, a task can be designed in such a way that it can be classified as either "simple" or "complex". Within the cognitive line of research into TBLT, of which Skehan's approach is an example, it is claimed that tasks of different levels of cognitive complexity have differential effects on learners' speech. The predictions about the effect of tasks on speech are the object of the next section.

[1] Skehan does not refer to them as "integrated tasks", but at least in some language assessment studies tasks which require a simultaneous operation of two or more skills are labeled as "integrated" (see e.g., Cumming et al., 2005).

3.5 Skehan's Trade-off Hypothesis: predictions and psychological underpinnings

Skehan's predictions are related to speakers' performance in three areas: fluency, accuracy, and complexity. Before moving on to presenting his predictions, the definitions of these constructs are presented, following Skehan and Foster (1999), p. 96-97:

Fluency: the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems.

Accuracy: the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error.

Complexity/range: the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system.

In general terms, Skehan's hypothesis is concerned with the ways in which complex tasks affect the areas mentioned above. The fundamental claim behind his hypothesis is that the three areas of performance described above cannot be attended to at the same time, because attending to one area of performance takes attention away from the others. Therefore, a tension is bound to occur between *meaning and form*. This idea is summarized in Figure 1.

More specifically, Skehan claims that under a complex task condition, a speaker's production will be more fluent, and there will be a trade-off effect between accuracy and complexity. What follows is his rationale for this prediction.

Skehan's proposal draws on a single-resource model developed by Van Patten (1985). Central to this model are the notions of meaning (content) and form (linguistic accuracy). Following Van Patten (1999), "while humans may indeed direct conscious attention to form in and of itself, the question is not whether they can do this; the question is *whether or not they can do this while they process input for meaning"* (p. 288, italics in the original). According to this theory, a single finite volume of attention is available for allocation to competing task demands, and so there exists a tension, or competition, between these two components.

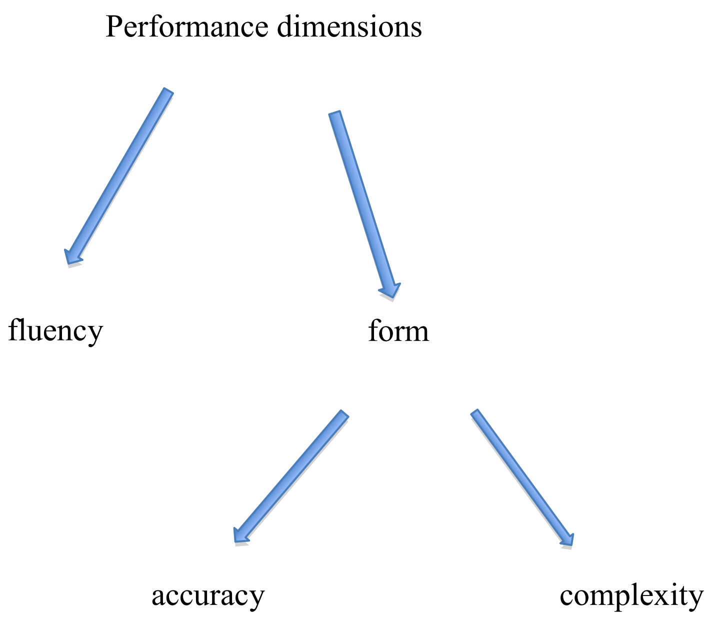


Figure 1. Theorizing dimensions of performance (Skehan & Foster, 2001)

When a learner's attentional limits are reached, they must prioritize one of the areas, either meaning or form, and a learner's "default priority" is to attend to the former rather than to the latter. He further argues that attention to form essentially takes places under two circumstances; when attention to form is a prerequisite for the recovery of meaning, and when "comprehension as a skill is automatized" (p. 288), because it frees attentional resources to focus on the linguistic forms.

Drawing on the idea of a single resource model, the central claim in Skehan's approach is that human information processing capacity is limited and this constraint "has far-reaching effects on second language processing and use" (Skehan 1998, p. 73). Following VanPatten, Skehan claims that when adult L2 learners engage in production, meaning is the primary objective rather than form. Since attending to meaning (understood as dealing with the content of the task or task completion) takes up most of the available attentional resources, there are scarce resources left for attending to form. As a consequence, complexity or accuracy to some extent will inevitably have to be sacrificed.

Therefore, when engaging in performance, there is an unavoidable tension between form (complexity and accuracy) on one hand, and fluency on the other hand. Skehan (2009) further explained this tension: "(…) within form, one can contrast attention directed to using challenging language (complexity) relative to conservative, less advanced language, but greater accuracy" (p. 511). In other words, learners choose between *meaning*, simply transacting a task and "getting the job done", and *form*, which can manifest itself by using the task as a resource to push the boundaries of the interlanguage through employing more difficult language (complexity) and error-free behavior (accuracy). Learners allocate attention in the way they consider most adequate: they prioritize the areas of performance they wish to attend to, as attentional resources are insufficient to attend to all areas of performance simultaneously, with a likely important role played by individual differences. Therefore, a competition occurs between accuracy and complexity: a decision to attend to one of these dimensions of performance has an inevitable effect of degrading the other areas. Although fluency, complexity and accuracy are all ultimate long-term goals when producing a second language, they “compete with one another during ongoing communication” (Skehan, 1998, p. 73).

Skehan and Foster (2001) suggest two mutually exclusive strategies that learners can adopt, faced with limited resources: “accuracy last” approach and “safety first” approach. In the former, the choice to stretch the interlanguage in terms of syntax and lexis happens at the expense of number of errors committed. In the case of the latter, an attempt to make fewer errors results in a poorer range of syntactic structures and lexical items, such that priority is given to the precision of expression, at the expense of complexity of production. From here follows the main assumption of Skehan’s hypothesis, as mentioned above, that as tasks become more complex, fluency will be boosted, at the expense of accuracy and/or complexity.

To sum up, there were two main focal points to the two previous sections. The first one was Skehan’s three-way distinction between code complexity, cognitive complexity, and learner factors which all affects the performance of a task. It was also shown that, by manipulating factors, tasks can be made ostensibly more “simple” or “complex”. Along these lines, the second focal point was to show how complex tasks affect performance, and possible interlanguage development, and the rationale for the specific predictions.

What merits attention in Skehan’s stance on tasks is that he does not really go beyond proposing the three-way distinction and predictions, and several fundamental issues are left unaddressed in Skehan’s proposal: where do the tasks come from (*task selection*) and how should they be sequenced in a syllabus (*task sequencing*). In other words, Skehan offers a typology of different factors potentially exerting an impact on performance, and possibly also on interlanguage development, but he does not specify their source. He partially agrees with Long’s (2005, 2006) idea of Needs Analysis, which posits the identification of the needs of learner communities, but the criticism he raises is that carrying it out is restricted to certain language learning contexts, and therefore is not always possible. The unaddressed issue of "where the tasks come from" is directly related to that of how tasks should be sequenced in a syllabus. Skehan and Foster (2001)

pointed out the following:

> "The individual task has to be located, in a principled way, in longer-term instructional sequences which seek to promote balanced development, such that improvement in one area will be consolidated by improvements in others" (pp. 193-194).

Along similar lines, Skehan (1996a) stated that "It is imperative that tasks are sequenceable on some principled criterion, since the basis on which tasks are ordered will be a reflection of what attentional resources they require" (p. 51). However, beyond stating that tasks can be placed in longer-term sequences, and that this operation must be guided by informed decisions, Skehan does not offer any fine-grained proposal regarding actual criteria according to which such sequencing can take place. It is fundamentally unclear whether sequencing decisions should be taken on the basis of a sole group of factors (such as cognitive ones), or whether more than one group is, or should be, at play.

Leaving the issues of task selection and sequencing unaddressed, Skehan's perspective on tasks and task-based instruction can perhaps be classified as a "weak" approach. He speaks from the classroom perspective, and conceives of tasks as communicative classroom activities, potential vehicles of communication, perhaps employed in order to practice different language forms and thus representing the "focus on formS" spectrum of syllabus options. Along these lines, the three-way distinction he proposes, more than being a model of task complexity, is a battery of possibilities to be tried out by a teacher or other language professional, with a community of learners.

Relating these issues to the psychological underpinnings presented earlier, he understands a "task" as a device by means of which *balance can be achieved between meaning and form*, leading to the development of the three areas of performance: fluency, accuracy, and complexity. In his view, this is done by exposing learners to tasks which do not exceed learner's ability in terms of linguistic and cognitive demands. They should pose only a reasonable challenge, and be "doable" using the current linguistic resources, and other knowledge types, that the learner is in possession of when performing a task. Such "reasonable challenge", according to Skehan, empowers the learner because they feel that what is expected of them is within their cognitive and linguistic limits, and so their motivation to perform a task increases. As Skehan (1998) stated, " (…) it is important that tasks are chosen which are of the appropriate level of difficulty, are focused in their aims between fluency, accuracy, and complexity, and have some basis in task-based research." (p. 131)

The ideas presented in the last few sections summarize Skehan's understanding of the effect of tasks on performance and development, and the way he conceives of task and task-based instruction. The objective of the upcoming section is to present the proposals developed by Peter Robinson, which directly inform the current dissertation. Similarly to Skehan, Robinson also suggested a triad of factors affecting performance, which, together with his Cognition Hypothesis and SSARC model of pedagogic task sequencing, constitute his major work. Each of them is introduced and discussed in the upcoming sections.

3.6 Robinson's Triadic Componential Framework

Similarly to Skehan, Robinson drew a distinction between three groups of factors potentially affecting performance, and these are: task complexity, task condition, and task difficulty, which act together and influence overall task complexity. Robinson's Triadic Componential Framework is summarized in table 2.

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

As can be observed, in this framework the three groups of factors represent different sources of influence: task complexity is related with cognitive factors, task condition is related with interactive factors, and task difficulty are learner factors, "in order that each dimension can be studied separately, and also that complex interactions among these dimensions and factors can be studied for their multiplicative effects on L2 learning and performance" (Robinson & Gilabert 2007, pp. 161-162). Each of them is divided into different subgroups, and the subgroups are further broken down into individual variables.

It can be observed that in this framework a clear distinction is drawn between potential different sources of influence: cognitive factors, interactive factors, and learner factors. These three categories place different kinds of demands during task performance: the first group is concerned with the attention and memory resources required for task execution, the second group concerns interactional and interactant demands, and the third group encompasses ability and affective variables. This framework therefore clearly distinguishes between potential sources of complexity, and, unlike other proposals, separates the different ways in which performance can be affected. Let us know analyze the different building blocks of this framework in more detail.

3.6.1 Task complexity

In the first group of factors – *cognitive factors* – an important distinction is made between two types of dimensions: resource-directing and resource-dispersing dimensions, which, within the overall "cognitive" category, place different kinds of demands on the speaker.

Resource-directing dimensions place conceptual demands on the speaker and they draw their attention to specific L2 forms and structures, such as lexical items or syntactic structures. This is achieved by manipulating the specific variables this group of factors is comprised of: ±here-and-now, ±few elements, ±reasoning (spatial, intentional, and causal), and ±perspective-taking. Following this categorization, a learner may for example be required to narrate a story in the present versus in the past, take the first person or third person perspective on an event, or describe a location. A task designed along these variables subsequently makes the learner pay attention to specific aspects of the language code. Narrating in the past draws their attentional resources towards past forms, taking the perspective of a different person draws attention to the third person "s" marker, and explaining a location may gear attention towards the use of prepositions of place. Similarly, complex reasoning about relationships between objects or characters, compared with simple information transmission, may push the emergence of cognitive state verbs. It can be stated overall that the resource-directing dimensions force the learner to notice certain aspects of the language code, and they have the potential to "cause a shift from the pragmatic to syntactic mode or to push development beyond the basic learner variety" (Robinson 2005, p. 8). When these linguistic demands are met, it may lead to the analysis of the interlanguage and the emergence of new form-function mappings.

Resource-dispersing dimensions make increased performative demands on a learner's attentional and memory resources. In contrast to the previous category, these dimensions do not direct the attentional resources to the linguistic aspects of a task, but disperse learners' attention over multiple non-linguistic dimensions of the task. They can manifest themselves as such variables as ±planning time, ±single task, ±task structure, ±few steps, ±independency of steps, or ±prior knowledge. A task designed along resource-dispersing variables can therefore give the learner time to plan the message or not, it can be done in one single step or require two or multiple steps in order to be completed, and it can follow a clear organizational structure, or it can be incoherent. Tasks manipulated along resource-dispersing variables are predicted as placing procedural demands on the learner, and they are believed to foster automaticity and control over the linguistic resources available in the interlanguage system. Following Bialystok (1994), Robinson and Gilabert (2007) claim that

> "Meeting these demands during pedagogic task performance therefore should facilitate not *analysis,* and development of new L2 form-concept mappings but rather automatic *access* to, and control of, an already established interlanguage system" (p. 166).

3.6.2 Task conditions

The second group of factors – *task conditions*, is concerned with two types of demands: interactive characteristics of tasks, and interlocutor characteristics. The first subcategory includes such dimensions as the direction of information flow: each participant holds part of the information necessary for task completion, and therefore information sharing is crucial to executing the task (*a two-way task*), or one participant is in the possession of all the required information (*a one-way task*). Task can have a clearly defined "correct" solution (*a clozed task*) or several multiple solutions (*an open task*). Furthermore, the task participants may be required to work towards a common goal (*convergent task*) or there might have conflicting interests (*divergent task*).

The variables present in the second category (interlocutor characteristics) are believed to influence the dynamics of participation and between-participant collaboration during task performance. They include such factors as, for example, gender, participant familiarity, or status. Rather than concerning task design *per se*, these variables reflect particular characteristics of the speakers involved in performing a task.

3.6.3 Learner factors

Finally, *learner factors* are concerned with the important distinction between how *difficult*, as opposed to how *complex*, a particular task is. In Robinson's framework, the perception of task difficulty is dependent upon individual differences, between any two learners, in affective factors such as aptitude, confidence, and motivation, among others. In other words, task difficulty accounts for *between-learner variation* when performing a simple and a complex task, whereas task complexity explains *within-learner variation* when performing any two tasks.

As can be observed, although all three groups of factors interact to influence performance, Robinsons clearly separates task factors from other types of influences, which is what distinguishes his framework from the previously described task complexity proposals. In his framework, task complexity factors, and partially also interactive factors, are amenable to deliberate task design manipulations. The variables, which belong to these factors, are the potential components of complexity in individual tasks and are subject to manipulation in multiple ways in task design, in order to develop a "simple" and a "complex" task. Ideally, having information about the learners can help designers match instruction to individual differences. Tasks designed in line with these criteria have a potential of channeling attention to some aspect of the discourse, or to one of the three dimensions of performance: fluency, accuracy or complexity.

From the different variables and groups of variables described above, Robinson proposes that tasks should be sequenced on the basis of their cognitive complexity, and, similarly to Skehan, he makes predictions about how engaging in complex tasks affects performance. These are thoroughly described in his Cognition Hypothesis, which is the object of the next section.

3.7 The Cognition Hypothesis: predictions

The fundamental claim of the Cognition Hypothesis is that tasks should be sequenced in an order of increasing cognitive complexity to enable a balanced development of the four dimensions of production: fluency, accuracy, and structural and lexical complexity. In the previous section we saw that within cognitive factors, Robinson draws a distinction between resource-directing and resource-dispersing factors. This section presents his specific predictions for performance of these two dimensions.

Robinson predicts that on *resource-directing dimensions* (those which draw the speaker's attention to linguistic aspects of a task), both complexity and accuracy will be boosted, to a possible detriment of fluency. Robinson justifies his claim arguing that complex tasks, given that they require a wider range of linguistic resources than their simple counterparts, force the learner to stretch their interlanguage in order to meet the increased cognitive demands. As Robinson (2001a, 2003) claims, compared to simple tasks, complex tasks are more attention-demanding and require a greater analysis of interlanguage on the part of the learner. This, in turn, leads to "more noticing of relevant forms in the input and problematic forms in the output leading to more incorporation and modification" (Robinson, 2001a, p. 305). Faced with high demands, learners need to stretch their linguistic repertoire and search for the lexis, syntax and precision of expression, which are desirable, if not necessary, for successful task completion.

Regarding the predictions of CH on *resource-dispersing dimensions* (e.g., planning time), which divide attention rather than draw it to any specific linguistic features of output, complex tasks will result in less accuracy and complexity. When tasks are complex along both types of dimensions simultaneously, there is a high probability of synergetic effects (Robinson & Gilabert, 2007). Where the two kinds of dimensions are mixed, the positive effects of increasing only resource-directing dimensions are less significant, if not nonexistent. These general predictions of the CH on production are complemented by further claims about the effects of tasks. These predictions involve effects of tasks on interaction and learning opportunities, and on individual difference-task performance and learning interactions (Robinson & Gilabert, 2007). Regarding the effects on interaction, the more interactive tasks are, the more between-participant interaction and negotiation of meaning they will generate. The latter, in turn, will provide space for dealing with challenging forms in both input and output. Finally, the third group of effects involves individual differences and learning interactions. Here CH claims that there is a relationship between learner factors (task difficulty) and task factors (task complexity). The assumption is that (the quality of) task-based performance is susceptible to individual differences and affective factors. For instance, the expected effect of boosted complexity and accuracy in the case of complex tasks may be expected to be found in learners with low output anxiety, an effect reported in a study by Robinson (2007).

3.7.1 Psychological underpinnings of the Cognition Hypothesis: Multiple Attentional Resource Model

Robinson's Cognition Hypothesis takes as its starting point Wickens' (2002, 2007) Multiple Attentional Resource Model, which draws on multiple resource theory (Navon & Gopher, 1979). Broadly speaking, the latter is concerned with the relative interference which occurs between two tasks if these are done in a simultaneous fashion in terms of the attention devoted to each of these tasks. Three elements are critical in determining the extent to which one of these tasks will be compromised: resource demands (further subdivided into effort competition, effort demands, and effort investment), resource similarity, and allocation policy.

There are three components to the model developed by Wickens (2007): processing modalities, processing codes, and processing stages. Specifically in terms of language, modalities can be of two types: hearing (auditory) versus reading (visual), and so can be codes: verbal/linguistic material versus spatial/analog non-verbal material. While attending to any two tasks carries an intrinsic difficulty, Wickens (2007) argues that when division occurs *across***,** rather than *within*, modalities and codes, managing two tasks is easier. Along these lines, performing a task, which, for instance, requires the division of attention between visual and auditory stimuli, poses less of a burden than when task input is of the same type, for example only auditory. If the latter is the case, the two tasks interfere with each other, which likely leads to a competition between them. As noted by Wickens (2007), "there will be relatively high competition between two perceptual tasks, two working memory tasks or a perception and working memory task (e.g., reading while rehearsing), while this competition will be greatly reduced if the same perception or working memory task is time shared with a response task (speaking, or manipulating)" (p.187). Essentially then, the model provides rationale for “good” versus “bad” performance, depending on whether two tasks draw on the same dimension of attention, or different dimensions. Finally, attention policy is concerned with the task performer’s priority given to one task over the other.

The last of the mentioned dimensions, *processing stages*, distinguishes between perception, cognition and responding. According to this model, these three components represent different stages when processing information; different pools of resources are responsible for different codes of information, perception, cognitive processing and responding. From here springs the fundamental assumption of Wicken’s model, concerned with parallel processing: “When a task requires processes at the same dimension (such as giving two spoken answers simultaneously), parallel processing of information generates performance problems. However, if task performance addresses different dimensions, parallel processing is possible without competition for attentional capacity, for example, one can read a word while pushing a button” (p. 56). Competition will take place only when two tasks draw on the same pool of resources; when they draw on separate resource pools, there will be no competition.

It is precisely Wickens’ idea of separate resource pools that constitutes the central component of Robinson’s stance on the issue of attention allocation. Robinson claims that as the three areas of production, fluency, accuracy and complexity, draw on different resource pools, there is no tension between accuracy and complexity on resource-directing dimensions. As tasks become more complex along resource-directing dimensions, learners can attend to more than one area of production simultaneously. More specifically, Robinson claims that accuracy and complexity of speech are not in competition for attentional resources. This claim is directly related to the predictions of the Cognition Hypothesis, which claims that a complex task is overall favorable to the quality of speech production: during task performance, the learner is not forced to prioritize one area of performance over the others, as increased task demands are beneficial to overall development. The area that increased task demands may be detrimental to, is fluency. Overall, however, in Robinson's understanding, complex tasks have a function of forcing the learner to stretch their linguistic repertoire to meet the demands the task places on the learner. As a consequence, they foster greater precision and force the learner to search for more linguistically complex solutions when faced with increased cognitive demands, at the expense of fluency.

To sum up this section, as previously mentioned, in his model Wickens postulates that greater success in performance is possible when two tasks belong to different codes and/or modalities, rather than to the same one. Wickens' predictions inform the predictions of the Cognition Hypothesis insofar as increasing the mental effort "needed to make more demanding *cognitive/ conceptual* distinctions in language should therefore prime learners – and direct their attentional and memory resources – to aspects of the L2 system required to accurately understand and convey them, thereby facilitating selective attention to, and "noticing" of these (Robinson, 1995) and so speeding up L2 grammaticization in conceptual domains" (Robinson & Gilabert 2007, p. 164).

Aside from the psychological basis, the predictions of the Cognition Hypothesis draw on parallel claims in other areas, such as functional/ cognitive linguistics (Givón, 1985), L1 developmental psychology (Cromer, 1974), and SLA research. The Cognition Hypothesis broadly claims that

> "increasing the cognitive demands of tasks contributing to their relative complexity along certain dimensions will (a) push learners to greater accuracy and complexity of L2 production in order to meet the consequently greater functional/communicative demands they place on the learner and (b) promote heightened attention to and memory for input, so increasing learning from the input, and incorporation of forms made salient in the input, as well as (c) longer term retention of input; and that (d) performing simple to complex sequences will also lead to automaticity and efficient scheduling of the components of complex L2 task performance" (Robinson, 2003, pp. 47-48).

From the perspective of parallels between L1 and L2 acquisition, the Cognition Hypothesis is generally related to Cromer's (1974) Cognition Hypothesis of language acquisition, which claims that "conceptual and cognitive development, creates the impetus for language development" (p. 49). While first language acquisition and adult language acquisition both draw on cognitive processes, the reasons why this happens are different. As pointed out,

> "the complex notions are not available to very young children, while they are available but not accessed in early stages of adult language acquisition … It is possible to stage increases in the cognitive demands of language learning tasks which recapitulate the ontogenetic course of conceptual development in childhood" (Slobin, 1993 in Robinson, 2003, p. 50).

To sum up Robinson's stance on tasks, in the Cognition Hypothesis and associated models (Triadic Componential Framework), and the theoretical underpinnings of his theories, at the level of pedagogic intervention Robinson proposes scaling up the cognitive complexity of tasks. Achieving an optimal balance in the distribution of attention resources is, in his approach, what leads to changes in interlanguage, and is the objective of a pedagogic intervention, and not striking a balance between meaning (content) and form (accuracy and complexity), as postulated by Skehan. Also, unlike Skehan, he does not postulate developing and administering tasks of "appropriate levels", but tasks characterized by *low to high cognitive complexity levels*, demanding more linguistic, and other types of resources, which the learners have at their disposal. Such cognitive challenge is expected to maximize opportunities for IL development. According to Robinson, it is the learner's exposure to cognitively very challenging tasks that causes the mobilization of all their resources, and therefore interlanguage development occurs. Rather than establishing "appropriate levels of difficulty", and exposing the speaker to the cognitive challenge he or she is able to respond to, Robinson advocates introducing maximum cognitive complexity in a sequence of tasks, which forces the learners to push their linguistic resources beyond the usual repertoire of familiar structures and forms, and so they incorporate into their output linguistic solutions which they would otherwise not employ. The basic area in which Robinson and Skehan clearly agree is that learners are sensitive to task design manipulations and that these manipulations have the potential to bring about changes in learner's interlanguage, perhaps leading to development over time.

3.8 Robinson's and Skehan's models juxtaposed

The two hypotheses presented in this chapter have received an increased interest from researchers, syllabus designers and teachers. They have been researched with multiple learner populations, both in laboratory and classroom settings and under a variety of research paradigms. Despite a large body of research carried out, the results obtained hitherto have provided a rather mixed picture of the effect of manipulating task demands on speech production and other outcomes. Multiple reasons have contributed to this state of affairs, such as task design, measures employed, or the lack of common agenda in operationalizing constructs, to name a few. Despite the fact that Robinson's is much more comprehensive and detailed proposal, both theories are highly speculative and far from definitive. Extensive research into the different variables in the TCF has not yielded sufficient evidence to support or reject either of them. In practice neither of the hypotheses, nor the research carried out within them, fully specifies the notions of "simple" and "complex", making it challenging to determine how simple or how complex is sufficient to have an effect, and they fail to provide uniform operational definitions or frameworks for investigating the aforementioned concepts.

While Skehan's moderate, or minimalist, approach may at first seem a handicap, the multitude of variables in Robinson's approach probably allows for too many interpretations, if not over-interpretations, of what variables such as "elements" or "reasoning demands" refer to in task design. This is evidenced in the number and scope of different operational definitions adopted across studies, which may contradict Robinson's assumption about a taxonomical framework of task characteristics, which, as he claims, should be (i) be motivated by a theory, (ii) be empirically researchable, identifying task characteristics and dimensions that predict differential effects on L2 performance, and (iii) be operationally feasible (Robinson, 2007, p. 13). Following Kuiken and Vedder (2007) "one may wonder whether the variables distinguished in the Triadic Componential Framework, and particularly the 2007 version, meet Robinson's second and third constraint of being empirically researchable and operationally feasible" (p. 278).

A somewhat underrepresented issue in both approaches is that of the role of individual differences, for example in the area of proficiency. Basically both authors do not go beyond admitting that they exert some sort of influence. Robinson (2007) claimed that "individual differences in ability and affective factors relevant to the cognitive demands of tasks will increasingly differentiate learners' speech production, and interaction and uptake, as tasks increase in complexity" (p.196). Along similar lines, Skehan and Foster (2001) stated the following:

(…) conceptual and empirical research is needed to better distinguish between the effects of IDs[2] on task performance and the effects of tasks themselves. Given that performance priorities are unavoidable, different learners may characteristically commit attention, when limited, to one of the areas covered here, e.g., the conversational risk-taker. Conversely, the task itself may lead to selective attentional commitments, e.g., structured tasks and greater fluency. It is difficult to know at present how to distinguish between these different sources of influence without additional experimental study, e.g., assessing whether dependable processing-based IDs exist. Performance dimensions in themselves are unrevealing" (p. 198).

3.9 The SSARC model of pedagogic task sequencing

What can be observed about all the previous models presented throughout this chapter is that the issue of sequencing generally lacked specification. It was shown that early proposals rather than disentangling different groups of factors, put them under the same conceptual category. The basis for sequencing was therefore dubious at best.

The objective of this section is to present the only – to my knowledge – fully-fledged proposal of sequencing in applied linguistics as developed by Robinson (2010), the SSARC model of pedagogic task sequencing. Prior to developing this model, Robinson addressed the urge to establish sound criteria for sequencing in his previous work, and one of its manifestations comes from his 2001 publication: "The development of theoretically motivated, empirically substantiable, and pedagogically feasible sequencing criteria has long been acknowledged as a major goal of research aimed at operationalizing task-based approaches to syllabus design" (Robinson, 2001, p. 27).

What merits attention in the above quotation are the ideas of "theoretically motivated, empirically substantiable, and pedagogically feasible". Robinson clearly rejects arbitrariness or randomness of criteria, and postulates the development of a

[2] Individual differences

rigorous model or models, which must meet three essential requirements: 1) being well-grounded in theory, 2) meeting the criterion of being researchable, and 3) lending itself to pedagogic interventions. While this list of criteria is not necessarily a complete or definite one, it is compelling in the sense that it captures three dimensions a potential model should address: theory, research, and pedagogy. Robinson's SSARC model of pedagogic task sequencing, the object of this section, is an example of such a framework.

Crucial to understanding Robinson's stance on sequencing is the separation of cognitive factors from other kinds of factors or sources of influence, as observed in the Triadic Componential Framework. This is important insofar as Robinson suggests taking task complexity, and its two dimensions – resource-directing and resource-dispersing variables – as *the sole basis for sequencing criteria*, which constitutes the first task sequencing principle in his framework (Robinson, 2010). Robinson supports his claim by stating that such sequencing helps

> "ensure deep semantic processing (Craik & Lockhart, 1972; Hulstijn, 2001, 2003) rehearsal in memory (Robinson, 2003) and elaboration and successful transfer of the particular 'schema' for interactive or monologic task performance to real-world contexts of use (Schank, 1999; Schank & Abelson, 1977)" (Robinson, 2010, p. 247).

What such sequencing implies in practical terms is that tasks can be sequenced along such variables as ±reasoning demands, ±here and now, ±elements (resource-directing dimensions), or ±planning time, ±single task, ±few steps (resource-dispersing dimensions), and not on the basis of, for example, participation variables (e.g., ±open solution, ±few participants), or participant variables (±same proficiency, ±same gender). Taking this as a starting point, he claims that "task *sequencing* is done by designing and having learners perform tasks simple on all the relevant parameters of task demands first, and then gradually increasing their cognitive complexity on subsequent versions" (Robinson, 2010, p. 246-247).

The second sequencing principle deals with the order in which different task complexity factors should be manipulated. Robinson proposes three stages in implementing a sequence of tasks:

1. tasks simple on all dimensions (that is, both resource-directing and resource-dispersing);
2. tasks cognitively demanding on *resource-dispersing dimensions*;
3. tasks cognitively demanding on both *resource-directing and resource-dispersing dimensions*.

Illustrating these principles with a task design example of a narrative task, such a task is first performed with provision of planning time and in the here-and-now condition (task is simple on all dimensions). On a subsequent version, the same task is narrated in the present, without provision of planning time (task is simple on the here-and-now dimension and complex on the planning time dimension). Finally, the most complex task version is narrated in the past with no provision of planning time (task complex on both resource-directing and resource-dispersing dimensions). This example can be extended to other resource-directing and dispersing variables, the idea being that increasing the cognitive complexity of task along resource-directing dimensions follows the increases along the other group of variables. It is argued that increasing cognitive complexity along the resource-directing dimensions triggers attention paid to form–function mappings, and so potentially promotes interlanguage development, while complexity increased along resource-dispersing dimensions "promotes increasing automatic access to current linguistic resources" (Robinson, 2010, p. 247).

The fundamental part of the SSARC model is the rationale as to how each of the above mentioned steps in task sequencing implementation affects the learner's interlanguage resources, and each of them is assigned a specific function. In step 1 (task simple on both resource-directing and resource-dispersing dimensions), task performance relies on the simple, stable (SS) state of current L2 proficiency. Cognitively more demanding step 2, consisting for example in removing planning time or adding multiple steps, fosters consolidation, and access to and automatization (A) of the learner's current interlanguage system, which is consistent with the procedural demands posed by resource-dispersing dimensions described earlier. Finally, by introducing maximum cognitive complexity (task complex both along resource-directing and resource-dispersing dimensions), step 3 leads to interlanguage restructuring, and to the development of new form-function/concept mappings. It also leads to maximum complexity (C) of the interlanguage as well as its destabilization. In step 3, attention is divided over many non-linguistics aspects of the task (resource-dispersing dimensions), and simultaneously resource-directing dimensions direct the attention to the linguistic aspects of the task.

There are several key features in Robinson's proposal which distinguish it from the models presented before. First, unlike early theorizing on grading and sequencing, the SSARC model is theoretically driven. Second, both the task taxonomy reflected in the Triadic Componential Framework and in the SSARC model separate cognitive factors from other kinds of factors (task condition and learner factors), and suggest scaling up the complexity of tasks only on the basis of their cognitive complexity. Third, as could be observed in previous proposals, factors clearly belonging to different categories were conflated, some being cognitive in nature, and therefore potentially lent themselves to be deliberate manipulations. Others, on the other hand, seemed to be external sources relatively alleviating or adding to the overall cognitive complexity of tasks. Clearly, separating these factors in the SSARC model presents an advantage as it distinguishes what is a controllable, manipulable internal feature from those sources of influence which are fundamentally different in nature. With these considerations in mind, Robinson's taxonomy of task characteristics, and general approach to task complexity and sequencing represent an important conceptual advancement in the treatment of task complexity and task sequencing. Potential shortcomings and drawbacks of this model will be presented in the Discussion chapter.

3.10 Summary of the chapter

The hypotheses presented in this chapter, the Trade-off Hypothesis and the Cognition Hypothesis, as well as the models associated with them (Skehan's three-way distinction and Robinson's Triadic Componential Framework), have inspired a substantial number of empirical investigations in the TBLT domain. While these are reported in the next chapter, a brief note relevant to the information presented in this chapter is that, when investigating the two hypotheses, researchers often times do not go beyond stating that Skehan and Robinson stand in opposition to each other regarding the predictions they make. As stated by Ortega (2004), "Despite some similarities, the mechanisms behind acquisition proposed by Skehan (1998) and Robinson (2001a) are clearly different, and it is unfortunate that these differences are often times ignored, including specific treatments" (p. 25-26, original in Spanish). This chapter, by juxtaposing Skehan's and Robinson's stances, has tried to illustrate that these two very influential authors exhibit differences in how they conceive of much broader issues – task, task-based instruction, and how they see the learner in a task-based approach. It is probably in these areas, rather than at the level of predictions, that these authors exhibit fundamental differences, an issue largely ignored in studies investigating these frameworks.

3.11 Implications for the current study

The information presented in this chapter has informed the current study in a number of ways. First, Robinson's Triadic Componential Framework was taken as a basis for selecting and manipulating variables. The manipulated variables were reasoning demands and elements, and the choice of these two is further explained in the Methodology chapter.

The issue of sequencing in the current dissertation is explored from the point of view of the SSARC model of pedagogic task sequencing. While the framework described here constitutes the theoretical basis for subsequent task sequencing decisions, not all considerations were taken into account: only resource-directing dimensions were manipulated in order to explore their effects on the learners' fluency, accuracy, and complexity behaviors. In the current dissertation task sequencing was operationalized as *delivering a sequence of three tasks manipulated along two resource-directing variables in one sitting and at short time intervals*. The ostensibly facilitative role of sequencing from simple to complex over alternative sequencing orders was put to test by having a group of participants perform tasks in the simple-complex sequence, and another group in randomized sequence. Details of this design are provided in the Methodology chapter.

As was observed in this chapter, the issue of individual differences is a largely underrepresented one in current theorizing about task complexity. This dissertation aims to cloze this gap by investigating task-based performance of speakers of different proficiency levels.

Having presented the conceptual work related to the concept of task complexity and sequencing, the next chapter deals with a selected portion of TBLT investigations into the concept of cognitive task complexity.

CHAPTER 4
EMPIRICAL RESEARCH INTO THE VARIABLES ±ELEMENTS AND ±REASONING DEMANDS

4.1 Introduction

The previous chapter showed, among other things, that a task can be conceived of as a set of variables amenable to a deliberate manipulation. The aim of this chapter is to present a critical overview of research carried out into two selected variables from the Triadic Componential Framework relevant to this dissertation: ±number of elements and ±reasoning demands.

The objective of the following review is three-fold. First, it aims to present the research carried out in the area of task cognitive task complexity, and it focuses particularly on those studies which investigated the influence of the manipulation of the two aforementioned variables on a range of phenomena, including language production, development or interaction. A total of thirty studies are reviewed, and the body of research includes journal articles, book chapters, and published and unpublished doctoral dissertations and master's theses.

Second, it offers a critical perspective on the reviewed studies in terms of their task design, with a special focus on such methodological issues as the operationalization of cognitive task complexity, independent measurement of hypothesized cognitive complexity differences, employed outcome measures, and, finally, results.

Third, the theoretical overview and evaluative analysis lead to the contextualization of the current study in light of the previous research. In this respect, the chapter outlines major areas of overlap and difference between the current study and previous research, as well as discusses the original contribution of this dissertation.

The chapter is organized in the following way. First the two variables, reasoning demands and elements, are presented in a top-down fashion, starting with references to cognitive psychology, and the scope is subsequently narrowed down to how they are conceived of in the TBLT domain, with references to Robinson's (2001, 2005, 2007) and Robinson and Gilabert's work (2007). This leads to a general overview of studies and to their grouping into three major categories: studies in which tasks were manipulated along ±reasoning demands, studies in which tasks were manipulated along ±number of elements, and studies in which the two variables were manipulated within the same task design, that is, simultaneous manipulation of ±reasoning demands and ±number of elements. In order to facilitate reading, each category of studies is illustrated with two tables, one focusing on task and study design features (e.g., operationalization of cognitive complexity, independent measurement of cognitive complexity, sample size, L1, etc.), and the other one reports the dependent variables, the measures employed, and partially also the results obtained.

This literature review is the result of an extensive literature search and effort was made to retrieve all the literature pertinent to, or shedding light on, the methodological aspects of this dissertation. To this end, the current review includes studies:

1 dealing with both production and acquisition;

2 dealing with various L1s and target languages;

3 that took place in both laboratory and classroom settings;

4 that took place in both foreign and second language contexts;

5 that employed performance, development or interaction outcome measures;

6 that employed various research designs;

7 that measured both oral and written data.

Although this diversity may seem a handicap, given that studies measuring different phenomena, such as production, acquisition, or interaction, are hardly comparable with each other, it was considered that including studies measuring a range of language phenomena helps to gain a complete insight into the body of research in the area of cognitive task complexity, as well as illuminates some of the decisions taken in the current study. This chapter is therefore an attempt at a "systematic secondary review of accumulated primary research studies" (Norris & Ortega, 2006, p. 4), with the scope of analysis narrowed down to those studies, which manipulated two specific resource-directing variables.

4.2 Defining "reasoning demands" and "elements"

In order to understand the underlying principles behind the existence and manipulation of reasoning demands and elements in the TBLT domain, let me recall some of the conceptualizations of the construct of cognitive load outside of the TBLT domain, which were mentioned in the second chapter.

While the concept of cognitive load is relatively novice in task-based investigations, it is a well-established one in other fields, and the TBLT explorations of cognitive load draw majorly on the accumulated knowledge of theorizing on one hand, and empirical investigations on the other hand, carried out in the area of cognitive psychology. Let me recall a definition of cognitive load by Sweller at al.: "Cognitive load is generally considered a construct representing the load that performing a particular task imposes on the cognitive system" (p. 266). Cognitive load is therefore the amount of computation necessary for a successful completion of a task, be it a pedagogic task requiring the use of language, or a different kind of task in any domain. Reviewing this concept in cognitive psychology, it was concluded that the higher the intrinsic cognitive load of a task, the more complex a task becomes for the performer.

As was seen in chapter 3, in TBLT tasks, one of the possible sources of cognitive load are the design factors of a task, the presence and intensity of which either exacerbates or alleviates the overall mental load imposed on the learner. In the current dissertation they are reasoning demands and number of elements, two factors along which cognitive complexity can be manipulated. Both these variables fall into the category of resource-directing dimensions in the Triadic Componential Framework. These dimensions are design features of tasks, and “increasing complexity along these dimensions is potentially a means of directing resources to a wider range of functional and linguistic requirements” (Robinson, 2001, p. 35). Manipulating the cognitive complexity of a task, then, ostensibly draws the task performer’s attention to the linguistic requirements needed for successful task completion. The demands a task poses “can be met by using specific features of the language code” (Robinson, 2001, p. 31).

4.2.1 Reasoning demands

“Reasoning demands” are the demands a task puts on the learner’s reasoning ability, and in general terms they are related to the amount of thinking or on-line computation a learner must do in order to complete a task. In cognitive psychology terms, “the more interacting variables to be processed in parallel, the higher the demand” (Halford et al., 1998, p. 806). In TBLT, reasoning demands can therefore be described as those components of a task which force the learner to reason about motives and/ or actions, and to justify choices, among others (Robinson, 2007). In a simple task learners are typically asked to merely transmit or exchange information, whereas a complex task requires them to carry out one or multiple mental operations in order to meet the task's cognitive demands. As was already suggested in previous task complexity proposals, "tasks requiring selective information transmission and reasoning to establish causality, and justification of beliefs are more complex than tasks requiring non-selective information transmission, without these demands" (Robinson, 2001, p. 38).

In the TBLT domain, three types of reasoning have been distinguished: (1) spatial reasoning, associated with a relative ease or difficulty with which speakers must make reference to physical location; (2) causal reasoning, understood as information transmission versus discovering links and establishing relationships between events; and (3) intentional reasoning, understood as information transmission versus understanding people's "intentions, beliefs and desires and relationships between them" (Robinson & Gilabert, 2007, p. 165). Let me illustrate each type of reasoning with sample tasks available in TBLT literature and manipulated along reasoning demands.

Spatial reasoning

The studies, which manipulated the variable ±spatial reasoning, typically employed a scenario involving physical location (frequently labeled as a "map task"), which generally required the speaker to make reference to the position of task components manifesting themselves as different animate and/ or inanimate objects. In a number of studies, the distinction between a simple and a complex task condition consisted in whether or not task input provided the learner with already existing objects, the logic behind such operationalization being that it is cognitively more demanding to describe physical location when scaffolding in the form of already existing objects is unavailable. A different kind of operationalization was one in which the task performer was required to provide directions on a map, and they therefore had to describe a route from point A to point B, making reference to the different landmarks, paths, and locations. In the latter scenario, typically the more landmarks were present, the more cognitively complex a task was considered to be. This was so because the components of a task belonging to the same type (e.g., people, objects, animals, etc.) had to be distinguished from each other in order to successfully complete the task, by making reference to slight differences, which potentially prompts the use of more advanced syntax or more elaborate lexis.

Causal reasoning

One of the typical examples of a task in which causal reasoning is involved is a picture arrangement task. Usually in a simple task of this kind, the learner was required to narrate a story in which the sequence of events was correctly arranged, whereas a complex task required the learner to establish the correct sequence in order for the story to be coherent and logical, and narrate it as a second step. This logic could be simplified by saying that a simple task involved information transmission, whereas a complex task involved reasoning prior to narrating. As Robinson (2005) noted, "tasks which require no causal reasoning to establish event relations, and simple transmissions of facts, compared to tasks which require the speaker to justify beliefs, and support interpretations of why events follow each other by giving reasons, require, in the latter case, expressions, such as logical subordinators" (p. 5).

Intentional reasoning

Finally, intentional reasoning in the TBLT domain is related to understanding motives, intentions, and assigning mental states to others. Forcing the learner to carry out these complex mental operations is expected, for instance, to bring about the occurrence of verbs which were found to occur later in L1 acquisition, and specifically, "psychological, cognitive state verbs (e.g., know, believe, suppose, think), [the use of] which introduces complex syntactic complementation" (Robinson, 2005, p. 5). Intentional reasoning can be manipulated, for example, by requiring the learner to successfully understand a character in a story to have a certain intention (in a picture/cartoon narration task). Failure to do so may compromise the storyline inherent in a task, and consequently the completion of a task is partial at best.

On the basis of the above-mentioned definitions and examples, it can be broadly stated that reasoning demands is the function of the amount of thinking inherent in a task, and the number and intensity of mental operations a task requires from a learner. Following Halford et al. (1998), "the more interacting variables to be processed in parallel, the higher the demand" (p. 806).

4.2.2 Elements

Turning to the dimension "elements", it has been extensively used in cognitive psychology and other fields, and the research into this variable in the TBLT domain has been considerable in amount. The fundamental idea behind cognitive complexity manipulated along this dimension is that the build-up of elements triggers a task more complex. However, this dimensions entails two sources of cognitive complexity: one is the number of elements itself, and the other one is how these elements interact with each other (Sweller & Chandler, 1994, Chandler & Sweller, 1996).

The empirical investigation into the variable “elements” has been characterized by a plurality of operationalizations adopted by the researchers in the domain, given the fact that this construct encompasses a wide range of task components. In previous research “elements” have referred to such components in task design as: number of landmarks in a map task, number of electronic devices, photographs of people, number of options or considerations to take into account when taking a decision, or financial resources, to mention just a few. It could thus be broadly stated that the dimension of “elements” refers to these components of a task which are subject to be counted, and it is the number of occurrences of that component, and the relationship between the different components, that marks the difference between a simple and a complex task condition.

In the case of a simple task, “few easily distinguished versus many similar elements are involved” (Robinson, 2007, p. 18), or “distinguishing between and selectively referring to one or more among many elements or objects” (Robinson, 2007, p. 194). Previous task complexity proposals suggested that “tasks requiring a few clearly different elements to be distinguished from each other (e.g., trees, apples and clouds) are easier than tasks requiring many similar elements to be distinguished (e.g., cars in a traffic jam, buildings and streets on a map)” (Robinson, 2001, p. 38).

However, as demonstrated in the previous chapter, current theorizing about elements does not offer specific guidelines as to the number of characters, considerations, or other components falling under the category “elements”, which should or must be incorporated into a task in order to distinguish between its simple and a complex task condition. Therefore the issue of how many elements are sufficient to distinguish a simple task from a complex one is largely undefined and it is the researcher’s task to operationalize the construct of cognitive complexity at the theoretical level and verify it by means of external validation in order to ensure that the *hypothetical designed* differences in complexity match *perceived* differences in complexity.

One of the available illustrations of cognitive complexity manipulated along the dimension of ±elements is the previously mentioned map task. In terms of the dimension in question, a simple task might include a small number of elements, which are clearly dissimilar from each other, and therefore they do not pose a challenge in terms of differentiating between them, whereas a complex task includes multiple, often times identical elements, with only slight variations. In the case of a map task, "tasks requiring complex spatial reasoning, event construal, and reference to motion, can be expected to lead learners to attempt to use developmentally later acquired lexicalization patterns for describing motion events" (Robinson & Gilabert, 2007, p. 166). However, as will be seen later in this chapter, including elements in a map task is only one of the possible operationalizations of this variable.

Comparing the analysis of elements and reasoning demands, the latter seems to be more thoroughly defined across studies, which perhaps makes this variable more operationally feasible and more easily adaptable to task design than the dimension of elements. In the case of the former, although the notion of "elements" is intuitive at first sight - the more occurrences of a certain component, the more complex a task is – importantly, available theoretical models do not go beyond proposing this variable as a potentially manipulable one, and empirical research has by no means produced a homogenous picture of linguistic behavior as a result of adding or taking away elements.

To sum up this section, in the case of both variables, the build up of cognitive complexity is not, or should not be, the function of being exposed to "more of the same variable", but the function of how these variables interact with each other. In other words, the difference between a simple and a complex task is quantitative *and qualitative*, in the sense that it is the relationship between the different task components rather than the components themselves, that constitutes the degree of difference. Consequently, the difficulty related with solving communicative tasks stems from the fact that the different elements must be processed simultaneously in working memory (Pollock et al., 2002).

Given that defining constructs and operationalizing variables is currently one of the most challenging aspects of the TBLT research agenda, this literature review places special emphasis on how researchers went about operationally defining reasoning demands and elements, and how these theoretical concepts were adapted to task design across studies. Results of each study will also be presented and discussed in terms of the quantity and variety of measures used.

4.3 Research into the variables "number of elements" and "reasoning demands" in the Triadic Componential Framework

Before proceeding to analyzing the studies, let me recall the fundamental claim of the Cognition Hypothesis. Broadly speaking, it posits that as pedagogic tasks are made cognitively more complex, the lower the fluency in terms of speech rate, and higher structural and lexical complexity, and accuracy. These claims have been broadly tested in TBLT literature, and the studies, which are presented in this chapter, have explored them by means of the manipulation of two aforementioned variables.

The studies presented here can be grouped into three major categories:

(1) studies in which task complexity was manipulated along the variable ±reasoning demands;

(2) studies in which task complexity was manipulated along the variable ±number of elements;

(3) studies in which task complexity was simultaneously manipulated along the variables ±reasoning demands and ±number of elements.

Although this review aims for comprehensiveness when reporting the relevant aspects of previous research, the scarcity of information provided in several primary studies made it impossible to report all the information in desired detail. Where encountered operationalizations in primary studies were minimal, so is their reporting in this chapter. Where such was the case, it was pointed out in footnotes.

When classifying studies into the different categories, the primary authors' original manipulations were maintained. That is to say, the decision to include a study as belonging to either ±reasoning demands or ±elements category was consistent with how the authors of studies conceived of their manipulation, and not how the manipulations were understood by the author of the dissertation.

In the case of those studies, which employed two or more different tasks, one of which was manipulated along ±reasoning demands and the other one along ±elements, they were classified as belonging to two different sections in this literature review, and were reported separately in the main text and in the tables.

4.4 Studies in which task complexity was manipulated along ±reasoning demands

This section presents research carried out into the variable ±reasoning demands from the Triadic Componential Framework. The information pertinent to the studies reviewed in this section is summarized in tables 3 and 4.

The first study to explore reasoning demands was that by Niwa (2000), which enquired about the effects of individual differences in working memory, intelligence and aptitude on performance, in addition to investigating reasoning demands. The tasks used in this study were picture arrangement tasks from the WAIS-Weschler Adult Intelligence Scale, a test of cognitive ability for adults[3]. In this study, 22 Japanese learners of English performed a series of four narrative tasks in which the number of pictures ranged from 3 to 6. Mixed results were found for the Cognition Hypothesis: on the complex task, greater lexical complexity and partly structural complexity were observed, to the detriment of accuracy and fluency.

The Weschler Adult Intelligence Scale was also used in Robinson's 2007 study, which enquired about the effect of intentional reasoning on production, interaction, and uptake. In this study, learners were required to arrange pictures in tasks of three complexity levels manipulated via intentional reasoning demands. In the simple task, a character had to be understood to have the intention of building a house.

In the complex version, the main character's intentions were a result of other peoples' perceptions. The most complex task required the learner to "conceptualize, frame and communicate the sequence" (Robinson, 2007, p. 198). Concerning the results, an effect was found on interactional moves: the complex task triggered more turns and a greater occurrence of clarification requests and confirmation checks; however, enhanced cognitive complexity did not have an influence on performance except for one specific measure of syntactic complexity (conjoined clauses), and the emergence of psychological and cognitive state terms. Lexical complexity and accuracy were not affected by the degree of cognitive complexity.

[3] This study does not provide further details about cognitive complexity operationalization.

Table 3. Summary of studies which manipulated the variable ±reasoning demands

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Niwa (2000)	Based on WAISRPA; ST = set 1 +CT =set 9 tasks between simple and +C = sets 3 and 7	4	none	Story telling (O) (from wordless comic strips)	21	Japanese/ English/FL
Robinson (2007)	ST: a character must be understood to have the intention of building a house; CT: the main character's intentions are a result of other people's perceptions +CT: the learner must "conceptualize, frame and communicate the sequence" (Robinson 2007, p. 198)	3	AVQ (9-point Likert scale) Difficulty estimation ST: *M*=2.5, *SD*=1.7 CT: *M*=3.9, *SD*=2.4 +CT: *M*=6.6, *SD*=2.4	Narrative (from a set of pictures) (O)	42	Japanese/ English/ FL
Nuevo (2006)	ST: narrating a story from ordered pictures CT: narrating a story from scrambled pictures	2	none	Narrative (from a set of pictures) (O)	113	Multiple (20)/ English/SL
Gilabert (2007c)	*Fire chief task* ST: similar characters, plenty of resources, few unconnected factors. CT: specific types of people, few resources, intricately related and dynamic factors	2	AVQ (9-point Likert scale) Complex task perceived as more difficult ($p<.001$)	Decision-making task (O)	42	Spanish/ English/ FL
Gilabert and Barón (2013)	Task 1 same as Gilabert 2007c Task 2 (Party task) ST: 2 vs. 3 guests invited to the party Few mental skills (comparing, deducing) CT: 7 vs. 4 guests invited to the party multiple step decisions	2	AVQ (9-point Likert scale) Time on task Time estimation task Complex tasks perceived as more difficult ($p<.05$)	Decision-making task (O) Party task (O)	36	Spanish/ English/FL

Note. TC=task complexity; N=sample size; L1=first language; L2=second language; FL=foreign language; SL=second language; ST=simple task; CT=simple task; CT=complex task; +CT=the most complex task, AVQ=affective variables questionnaire; O=oral task; W=written task; *M*=mean; *SD*=standard deviation; NSS=native speakers; C1=condition 1; C2=condition 2; C3=condition 3; C4=condition 4; WAISRPA= Wechsler Adult Intelligence Scale-Revised PA.

Table 3 continued. Summary of studies which manipulated the variable ±reasoning demands

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Lee (2002)	*Car crash task* ST: 2 cars, 2 car types, 0 pedestrians, 3 way-intersection, 2 moving directions CT: 3 cars, 3 car types, 1 pedestrian, 3-way intersection, 3 moving directions +CT: 4 cars, 4 car types, 2 pedestrians, 4-way intersection, 4 moving directions *Body functions task* ST: 6 words, 3 stages, 2 arrows; CT: 11 words, 4 stages, 3 arrows, +CT: 16 words, 4 stages, 7 arrows.	3	none	Two picture description tasks (O)	75	English/ Korean/ SL
Baralt (2010)	ST: no need to reason about the intentions of characters' actions CT: need to reason about the intentions	2	AVQ (6-point Likert scale) Time judgment task Qualitative comments	Narrative (O, W-CMC)	70	English, Korean Mandarin Chinese/ Spanish/SL
Kim (2009)	ST: narrating from ordered pictures; CT: narrating from scrambled pictures.	2	AVQ (9-point Likert scale)	Picture narration task (O)	34	Multiple/ English/ SL
Choong (2011)	ST: narrating a story from ordered pictures CT: narrating a story from scrambled pictures	2	Subjective ratings by NSS (4-point scale[4]) C1: M=2.5; C2: M=3.7; C3: M=1.7; C4: M=2.1	Narrative (from a set of pictures) (O)	22	Japanese/ English/ SL
Kormos and Trebits (2012)	ST: narrating a story from 6 ordered pictures CT: narrating a story from unrelated pictures	2	none	(1) Cartoon description (2) Picture narration (O, W)	44	Hungarian/ English/ FL

[4] In this study "1"=the most difficult task, and "4"=the least difficult task.

Table 3 continued. *Summary of studies which manipulated the variable ±reasoning demands*

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Shiau and Adams (2011)	Same as Gilabert (2007c)	2	AVQ (9-point Likert scale) ST: *M*=6.07 CT: *M*=6.80	Same as Gilabert (2007c) (O)	15	Mandarin, Korean, Japanese/ English/ SL
Notash and Yousefi (2011)	Same as Gilabert (2007c)	2	none	Same as Gilabert (2007c) (O)	60	Persian, Turkish, Kurdish/ English/ FL

Note. TC=task complexity; N=sample size; L1=first language; L2=second language; FL=foreign language; SL=second language; ST=simple task; CT=simple task; CT=complex task; +CT=the most complex task, AVQ=affective variables questionnaire; O=oral task; W=written task; *M*=mean; *SD*=standard deviation; NSS=native speakers; C1=condition 1; C2=condition 2; C3=condition 3; C4=condition 4; WAISRPA= Wechsler Adult Intelligence Scale-Revised PA.

Table 4. Dependent variables and their measurement in studies which manipulated the variable ±reasoning demands

STUDY	Dependent variables	Measurement of dependent variables
Niwa (2000)	*Fluency*	(1) words/ second; (2) second/ pause; (3) words/ pause.
	Syntactic complexity	(1) words/ T*; (2) SN/T*.
	Lexical complexity	(1) TTR.
	Accuracy	(1) error-free T-unit
Robinson (2007)	*Fluency*	(1) speech rate; (2) words/C-unit.
	Lexical complexity	(1) TTR.
	Syntactic complexity	(1) Cl/C-unit; (2) words per turn*.
	Accuracy	(1) percentage of error free C-units.
	State terms	(1) psychological*; (2) physiological; (3) emotion; (4) desire; (5) cognitive*; (6) all psychological state terms per turn*; (7) cognitive state terms per turn* .
	Indices of complex speech	(1) infinitival phrases; (2) conjoined clauses*; (3) wh-clauses.
	Interactional moves	(1) nº of turns taken; (2) clarification requests; (3) confirmation checks for listener/sequencers on each task version.
	Pre-task input	(1) exact uptake (exact, unaltered phrase); (2) partial uptake; (3) exact uptake per turn; (4) partial uptake per turn.
Nuevo (2006)	*Interactional opportunities*	(1) recasts; (2) clarification requests; (3) confirmation checks; (4) comprehension checks.
	Orientation to form	(1) other repair; (2) self-repair*; (3) hypothesis testing; (4) metalinguistic talk.
	Type of orientation to form	(1) target like; (2) non-target like; (3) no resolution.
	Linguistic focus	(1) lexical; (2) syntactic; (3) morphological; (4) phonological; (5) orthographic.
Gilabert (2007c)	*Accuracy (self-repair)*	(1) errors/AS; (2) ratio errors/words; (3) error-repairs/AS, (4) ratio error-repairs/words; (5) all repairs/AS; (6) all repairs/ 100 words*; (7) percentage of self-repairs; (8) ratio repaired /unrepaired errors; (9) corrected ratio repaired/ unrepaired errors*.
Gilabert and Barón (2013)	*Pragmatic moves*	(1) nº of pragmatic moves x 60 seconds; (2) nº of pragmatic moves x 100 words; (3) nº of pragmatic moves per turns
	Types of pragmatic move	(1) possibility /probability; (2) conditionals; (3) should + have to; (4) impersonal; (5) decision; (6) want statement
Lee (2002)	*Accuracy*	(1) error-free C-units ratio; (2) error-free clauses ratio; (3) correct particle use in obligatory and non-obligatory contexts* (car task)
	Syntactic complexity	(1) clauses/C-unit; (2) inflectional suffixes/C-unit; (3) coordination index (independent cl coordinators/combined cl)
Baralt (2010)	*Development of Spanish subjunctive*	(1) past subjunctive forms; (2) indicative past tense forms; (3) present subjunctive
Kim (2009)	*Occurrence of LREs*	(1) grammatical LREs; (2) lexical LREs; (3) total LREs
	Resolution of LREs	(1) correctly resolved LREs, (2) unresolved LREs, (2) incorrectly resolved LREs, (4) total LREs.
Choong (2011)	*Content complexity*	(1) Idea units;
	Syntactic complexity	(1) words/T
	Evidence of reasoning demands	(1) frequency count of indicators of intentional reasoning demands; (2) frequency count of indicators of spatial reasoning demands; (3) frequency count of indicators of causal reasoning demands

Table 4 continued. Dependent variables and their measurement in studies which manipulated the variable ±reasoning demands

STUDY	Dependent variables	Measurement of dependent variables
Kormos and Trebits (2012)	*Fluency*	(1) SRA;
	Accuracy	(1) error-free clauses; (2) error-free relative clauses; (3) error-free past-tense verbs; (4) error-free verbs*(O)
	Structural complexity	(1) clause length*(W); (2) ratio of subordinate clauses; (3) ratio of relative clauses*(W)
	Lexical complexity	(1) D
Shiau and Adams (2011)	*Accuracy*	(1) errors/AS; (2) errors/words; (3) self-repairs/AS; (4) self-repairs/words
	Syntactic complexity	(1) Cl/AS; (2) words/Cl
	Syntactic variety	(1) number of different grammatical verb forms used*
	Lexical complexity	(1) Guiraud's Index*
Notash and Yousefi (2011)	*Uptake*	(1) focus on form episodes with learner responding to corrective feedback
	Successful uptake	(1) correct use of the corrected form immediately following correction

Note 1: SN=S-node; TTR=type-token ratio; Cl=clause; AS=AS-unit; LREs=language-related episodes. T=t-unit; SRA=unpruned speech rate (Rate A); O=oral task mode; W=written task mode;

Note 2: An asterisk indicates a significant difference detected in favor of the complex task condition. Such signaling was employed only in those studies which measured CAF.

A version of a picture arrangement task was also used in a study on the development of English past forms by Nuevo (2006). In an oral narrative task, 113 learners were required to tell a story in the past. In this study the difference between the simple and the complex task consisted in that in the simple condition the frames were arranged in the correct order, whereas in the complex task they were not. Regarding the operationalization of reasoning demands, the lack of arrangement required the learners to discover the correct sequence of events, in addition to narrating the story in the past. This study found no support for the development of the target feature (English past simple) or the amount of between-participant interaction, except for one measure (confirmation checks).

In a study by Gilabert (2007c), a decision-making fire-chief task under two cognitive complexity conditions was performed by 42 Spanish learners of English. The tasks represented a scenario in which there was a building on fire and the learner, in the role of a fire-chief, had to decide which characters should be given priority when saving them, and justify their choices. In this task, reasoning demands were operationalized as specific roles assigned to characters, available resources, and the presence or absence of "difficult conditions". In the simple version, the characters were assigned no specific roles, there was a moderate degree of danger, a wide range of resources was available, and there were few unrelated factors. In the complex version, characters were assigned roles, there were differences in the degree of danger, few resources were available, and there were many related factors. This study focused specifically on the accuracy of speech production, and it found that on two out of nine measures used, accuracy levels were statistically significantly higher in the case of the complex task condition (all repairs per 100 words and corrected ratio repaired/ unrepaired errors).

The fire-chief task was also used in Gilabert and Barón (2013), which enquired about pragmatic moves in task-based performance. In this study, increased demands of the complex version resulted in increased frequency of using conditional sentences. This study also used what was called a “party task”, in which participants engaged in a role-play where one participant had to ask the other one for permission to hold a party in the house they shared. However, both had certain conditions, such as the number of people they wanted to invite, or the food to be served, and they also had to consider the conditions of a third flat mate who was absent from the conversation. The difference between the simple and the complex version was operationalized as conflicting interests, in four categories, between the conditions of two participants involved in the role-play and those of the absent flat mate. Results indicated that the complex party task led to a higher occurrence of pragmatic moves in the categories “probability/ possibility”.

A different perspective on exploring reasoning demands was offered by Lee (2002), who investigated the role of task complexity on oral accuracy and complexity. Two narrative tasks were used with 75 English speakers of Korean: “car crash” and “body functions”, with three complexity levels per task. In the most complex version of the car crash task, reasoning demands were operationalized as the number and variety of cars, road conditions, and the variety of directions in which the cars were moving. The participants acted as eyewitnesses and they were required to describe the accident in as much detail as possible. In the body functions task, reasoning demands were manipulated along the number of words in the text, the number of stages involved, and the number of arrow movements in the pictures. This task showed selected aspects of the functioning of the human body system and in both task complexity conditions the participant’s task was to describe the pictures in detail.

Regarding the results, a trend was observed towards a trade-off effect between accuracy and complexity on the task of medium complexity level. The results of the body task were largely unsystematic, with different measures revealing different patterns in both accuracy and complexity. The three cognitive complexity levels of tasks did not yield statistically significant differences in either of the two tasks.

In a study by Baralt (2010), which investigated the effect of task complexity on the development of Spanish subjunctive, 70 participants were involved in retelling a story in dialogic production. In the first task, a family discovered that something valuable they had in their house was missing and they wrongly accused the housekeeper of having stolen it. In the second task, two men, who were keen on football, were invited by a scout to play with the city's team. Out of the two players one of them was much better, and he had to choose whether to play professionally or help his not so good friend. In this study the degree of cognitive complexity levels was based on the fact that the simple version did not require the learner to reason about the intentions of people's actions, and the complex task did. In addition to exploring task complexity, this study also measured the impact of mode: face-to-face (FTF) versus computer-mediated communication (CMC) mode. The results showed that increased cognitive complexity was more conducive to language development, but differences were found between the two modes. In the FTF mode the +complex triggered the most development, whereas it did not affect the participants in the CMC mode.

A study by Kim (2009) measured the impact of cognitive complexity differences on the occurrence and resolution of language-related episodes (LREs). Thirty-four learners of English engaged in the performance of a picture narration task; in the simple condition, the pictures were ordered, whereas in the complex condition they were not. The complex task condition led to significantly more instances of language-related episodes.

Choong (2011) enquired about the manipulation of two cognitively different tasks on production in terms of content complexity, syntactic complexity, and evidence of reasoning demands[5]. In this study, 22 Japanese learners of English were required to orally narrate a story from correctly arranged pictures (simple task), and to narrate a story from unordered pictures (complex task). The complex task condition triggered significantly more idea units[6], and a higher frequency count of indicators of spatial and causal reasoning demands, but no effect was found on syntactic complexity.

In a study by Kormos and Trebits (2012), which investigated the impact of cognitive task complexity on performance in the spoken and written mode, 44 Hungarian learners of English performed two tasks: a cartoon description and a picture narration task. The difference in cognitive complexity levels was operationalized as ordered versus unordered pictures. The effect of cognitive complexity turned out to be minor, with the complex task triggering more errors in the oral tasks (one out of four accuracy measures), and structural complexity in the written tasks, as measured by the ratio of relative clauses and clause length (two out of three measures).

Shiau and Adams (2011) set out to investigate the impact of cognitive complexity on performance, and employed the fire-chief task previously used in Gilabert (2007c). While the complex task condition did not have an influence on accuracy or syntactic complexity, it did affect syntactic variety (as measured by the

[5] In this study, apart from cognitive task complexity, the independent variable of contextual support was investigated.

[6] Idea units were operationalized as "the meaningfulness and elaboration of propositional information within the parameters of the task" (Choong, 2011, p. 14).

number of different grammatical verb forms used) and lexical complexity (as measured by Guiraud's Index), with only one measure per dimension used.

The same task in its simple and complex condition was used in a study by Notash and Yousefi (2011). This study enquired about uptake, operationalized as focus on form episodes with learner responding to corrective feedback, and successful uptake, defined as correct use of the corrected form immediately following correction. The two versions of the task were performed by 60 learners of English, who represented a range of L1 backgrounds. The complex task resulted in a higher occurrence of successful uptake.

4.4.1 Reasoning demands: operationalization of cognitive complexity

In the category of studies reported above, the tasks the learners had to perform were, broadly speaking, narratives or decision-making tasks. The tasks employed in the studies reviewed here required the learners to reason, justify, and give opinions, which is generally consistent with the way reasoning demands were earlier defined at the theoretical level. In terms of task outcome, learners were mostly asked to provide the correct arrangement of task input presented to them in the form of pictures, but other tasks (e.g., fire-chief) also included the component of making a decision before justifying an action. It is justifying choices, giving reasons, et cetera, that distinguished a complex task from a simple one, the latter being typically associated with narration or description. In the case of this variable, the number of complexity levels involved ranged from two to four, with three studies employing more than two cognitive complexity levels (see table 3, column "Nº of TC levels").

What can be seen in this group of studies is that the relative size of difference between a simple and a complex task version has been typically associated with information transmission or description of some kind in the simple task versus engaging in higher-level mental processes, such as justifying, reasoning, among others, in the complex task. The hypothesized degree of difference between cognitive complexity levels can therefore be conceived of as the relative need of online computation required from the learner. Simple tasks have typically required the speaker to transmit readily available information which needed to be communicated, whereas in complex tasks such simple information transmission was substituted, or complemented, by more advanced mental operations.

In those studies, which employed a narrative, the relative difference between the simple and the complex task version consisted in whether or not the task required plot conceptualization prior to engaging in task performance. Following this line of thought, simple tasks were the ones which asked the learner to merely *describe a set of correctly arranged pictures*, whereas complex tasks required the mental operation of *discovering the correct sequence prior to task performance*. An exception to this pattern is the study by Lee (2002), and more specifically, the body functions task. Across the three tasks of different cognitive complexity levels, the participants were merely required to describe the pictures, whereas in other studies the mental operation of describing was present only in the simple task condition. It raises the question of whether this task was manipulated along reasoning demands, or if the manipulated variable was, in fact, number of elements. It also potentially raises the issue of the relative difficulty of disentangling number of elements from reasoning demands.

4.4.2 Reasoning demands: independent measurement of cognitive complexity differences

In the group of studies discussed here, as table 3 demonstrates, task difficulty was measured independently from task performance in seven out of twelve studies. By far the most common way of verifying the degree of task complexity difference has been the technique of subjective difficulty ratings by task performers themselves (6 studies), but one study (Choong, 2011) employed native speaker ratings instead. The fact that not all studies reported the means obtained makes it challenging to compare the results on a common scale, but two studies which did provide the means (Robinson, 2007, and Shiau & Adams, 2011) demonstrate that a task hypothesized to be very complex in one study may be a simple one in another study. In Robinson (2007), M=6.6 for the most complex task, whereas in Shiau and Adams (2011), M=6.07 and M=6.80 for the simple task and for the complex one, respectively.

4.4.3 Reasoning demands: confounding and mislabeled variables

It may be the case that in some studies variables other than reasoning demands came into play. For example in the study by Nuevo (2006) it could be argued that the action of reconstructing events from memory, a mental operation indispensible for task performance in this task design, was a resource-dispersing dimension adding to overall cognitive complexity.

In the study by Lee (2002), the fact that the learners were looking at the picture while describing it could also be interpreted as +contextual support. In the same study it could be argued that cognitive complexity was manipulated along ±number of elements in addition to ±reasoning demands, as in the complex versions it was the *number of cars* in the car crash task and the *number of arrows* in the body functions task that were the object of cognitive manipulation.

4.4.4 Reasoning demands: emerging patterns in findings and conclusion

Despite the fact that issues such as the ones mentioned above are subject to debate, the studies reviewed here overall comply with general assumptions of reasoning demands as laid out by Robinson. With a few exceptions, the different task designs reported here can be described as quite consistent across the different studies in the sense that in order to successfully complete a complex task, a series of mental operations needed to be carried out. At the level of task design, and cognitive complexity operationalization, a certain degree of unanimity can therefore be observed. However, despite such relative homogeneity in operational definitions of cognitive complexity, studies have yielded very unsystematic results. None of the studies provided full support for the Cognition Hypothesis (nor, for that matter, for the Trade-off Hypothesis). The general picture of findings is a mixed one, with the benefits for accuracy being somewhat clearer than in the case of the other dimensions: in three studies which investigated CAF, accuracy increased as a result of engaging in the complex task. The other CAF dimensions showed very little variation under the simple and complex conditions.

4.5 Studies in which task complexity was manipulated along ±number of elements

This section reviews those studies in which the variable ±elements from the Triadic Componential Framework was manipulated to examine its effect on different outcome measures in a second language. As was mentioned in the previous chapter, elements refer to those components of a task, which are subject to be counted, and the difference between a simple and a complex task condition is typically conceived of as a build up of elements, and also the complexity of their interrelatedness. Similarly to the previous section, the main focus is placed on the operationalization of cognitive complexity, the measures employed, and the results obtained. The body of research presented dealt mostly with production as opposed to acquisition, and in almost all studies complexity, accuracy, and fluency measures were used to track changes in performance, with one study inquiring about the effect of increased complexity on negotiation of meaning. The information about the studies relevant to this section is summarized in tables 5 and 6.

Table 5. Summary of studies which manipulated the variable ±number of elements

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Robinson (2001)	*Number of elements* ST: small area; CT: large map area. *Prior knowledge* ST: participants familiar with the area; CT: participants unfamiliar with the area	2	AVQ (9-point Likert scale) ST: M=3.5 (SD=2.1); CT: M=5.4 (SD=2.3)	Instruction-giving map task (O)	44	Japanese/ English/ FL
Gilabert (2007c)	*Number of elements* (map task) ST: few, easily distinguishable elements CT: many similar elements	2	AVQ (9-point Likert scale) Complex task perceived as more difficult (p<.05)	Instruction giving map task (O)	42	Spanish/ English/ FL
Gilabert, Barón, and Llanes (2009)	Same as Gilabert (2007c)	2	AVQ. Complex task perceived as more difficult (p< .05)	Same as Gilabert (2007c)	60	Spanish/ English/ FL
Kuiken, Mos, and Vedder (2005)	Number of criteria to take into consideration when choosing a resort or B&B. ST: 3 requirements CT: 6 requirements	2	none	Personal letter to a friend (W)	62	Dutch/ Italian/ FL
Kuiken and Vedder (2007)	Same as Kuiken, Mos, Vedder (2005)	2	none	Same as Kuiken, Mos, Vedder (2005)	159	Dutch/ Italian, French/FL
Kuiken and Vedder (2008)	Same as Kuiken, Mos, Vedder (2005)	2	none	Same as Kuiken, Mos, Vedder (2005)	167	Dutch/ Italian, French/FL
Michel, Kuiken, and Vedder (2007)	ST: 2 elements CT: 6 elements	2	none	Descriptive, argumentative task (O)	44	Moroccan, Turkish/ Dutch/SL

Table 5 continued. *Summary of studies which manipulated the variable ±number of elements*

STUDY	**Cognitive TC operationalization**	**Nº of TC levels**	**Independent TC differences measurement**	**Pedagogic/ research task type**	**N**	**L1/L2/ Context (FL vs. SL)**
Kim (2009)	*Beach scene task* ST: few elements[7] CT: many elements	2	AVQ (9-point Likert scale) *Results not provided*.	Picture difference task (O)	34	Multiple/ English/ SL
Michel (2011)	ST: 4 (people to choose from (total of 4 possible combinations) CT: 6 people to choose from (total of 9 possible combinations)	2	AVQ (5-point Likert scale) ST: M=3.6 (SD=.06), CT: M=3.5 (SD=.07)	Argumentative tasks (study task, dating task) (O)	64	Moroccan, Turkish/ Dutch/ SL
Michel (2013)	Same as Michel (2011)	2	none	Same as Michel (2011)	64	not reported/ Dutch/SL
Sasayama (2011)	ST: 2 easily distinguishable characters (female and male); CT: 9 not easily distinguishable characters (some same-sex characters)	2	none	Narrative based on a picture story (O, W)	10	Japanese, Korean/ English/ SL
Levkina and Gilabert (2012)	ST: 2 apartment descriptions (holiday destinations) CT: 4 descriptions Other variable: planning time	2	AVQ C2 vs. C4 (p=.001) C2 vs. C3 (p=.036) C3 vs. C4 (p=.014)	Decision-making task (O)	42	Spanish/ English/ FL
Sasayama and Izumi (2012)	ST: 2 easily distinguishable characters CT: 6 not easily distinguishable characters	2	none	Narrative task (O)	23	Japanese/ English/ FL

Note. TC=task complexity; N=sample size; L1=first language; L2=second language; FL=foreign language; SL=second language; ST=simple task; CT=simple task; CT=complex task; +CT=the most complex task, AVQ=affective variables questionnaire; O=oral task; W=written task; M=mean; SD=standard deviation; C2=condition 2; C3=condition 3; C4=condition 4[8].

[7] The author did not articulate what specific task components this variable referred to (e.g., characters, objects, etc.).
[8] The four conditions in Levkina & Gilabert (2012) were: 1) +Planning time, +few elements; 2) –Planning time, +few elements; 3) +planning time, -few elements, 4) –planning, -elements.

Table 6. Dependent variables and their measurement in studies which manipulated the variable ±number of elements

STUDY	Dependent variables	Measurement of dependent variables and results
Robinson (2001)	*Fluency*	(1) words/C-unit.
	Accuracy	(1) error-free C-units.
	Lexical complexity	(1) TTR*.
	Structural complexity	(1) Cl/C-unit.
	Negotiation of meaning	(1) clarification requests; (2) confirmation checks.
Gilabert (2007c)	*Accuracy (self-repair)*	(1) errors/AS*; (2) ratio errors/words*; (3) error-repairs/AS; (4) ratio error-repairs/words*; (5) all repairs/AS*; (6) all repairs/ 100 words*; (7) percentage of self-repairs*; (8) ratio repaired /unrepaired errors*; (9) corrected ratio repaired/unrepaired errors*.
Gilabert, Barón, and Llanes (2009)	*Interaction moves*	(1) clarification requests; (2) confirmation checks; (3) comprehension checks; (4) recasts; (5) LREs.
	Accuracy	(1) repairs*.
Kuiken, Mos, and Vedder (2005)	*Accuracy*	(1) 1DE/T; (2) 2DE/T; (3) 3DE/T *; (4) TE/T*.
	Syntactic complexity	(1) Cl/T; (2) dependent cl/cl.
	Lexical complexity	(1) TTR; (2) TTR corrected for text length.
Kuiken and Vedder (2007)	*Accuracy*	(1) grammar error/T; (2) lexical error/T*(It, Fr); (3) orthography error/T; (4) appropriateness error/T, (5) other error/T;
	Lexical variation	(1) lexical frequency profile* (Italian);
Kuiken and Vedder (2008)	*Accuracy*	(1) errors/T* (It, Fr), (2) 1DE/T* (It, Fr); (3) 2DE/T* (It, Fr); (4) 3DE/T.
	Syntactic complexity	(1) Cl/T, (2) dependent cl/cl
	Lexical complexity	(1) TTR* (Fr); (2) TTR corrected for text length
Michel, Kuiken, and Vedder (2007)	*Fluency*	(1) SRA; (2) SRB; (3) filled pauses/100 words.
	Accuracy	(1) Errors/AS*; (2) lexical errors/AS; (3) omissions (or articles, verbs and subjects)/AS; (4) self-repairs/errors; (5) percentage of self-repairs/total number of words.
	Structural complexity	(1) Clauses/AS; (2) subordinate clauses/clauses.
	Lexical complexity	(1) Guiraud's Index; (2) lexical words/total no of words*.
Kim (2009)	*Occurrence of LREs*	(1) grammatical LREs; (2) lexical LREs; (3) total LREs.
	Resolution of LREs	(1) correctly resolved LREs, (2) unresolved LREs, (2) incorrectly resolved LREs, (4) total LREs.
Michel (2011)	*Fluency*	(1) SRA; (2) SRB; (3) repairs/AS; (4) filled pauses/AS.
	Accuracy	(1) lexical errors/AS; (2) morphosyntactic errors/AS; (3) determiner errors/AS.
	Structural complexity	(1) subordination index; (2) words/clause.
	Lexical complexity	(1) Guiraud's Index*.

Table 6 continued. Dependent variables and their measurement in studies which manipulated the variable ±number of ±number of elements

STUDY	Dependent variables	Measurement of dependent variables and results
Michel (2013)	*Frequency of conjunctions*	(1) target conjunctions/100 words.
	Occurrence of conjunctions	(2) number of participants using a conjunction at least once.
Sasayama (2011)	*Accuracy*	(1) error-free Cl/T;
	Structural complexity	(1) Cl/T* (W)
Levkina and Gilabert (2012)	*Fluency*	(1) SRB.
	Accuracy	(1) errors/AS.
	Structural complexity	(1) Cl/AS.
	Lexical complexity	(1) Guiraud's Index*.
Sasayama and Izumi (2012)	*Fluency*	(1) % of repeated words and phrases; (2) syllables/ per total number of seconds.
	Accuracy	(1) % of error-free clauses; (2) % of accurate use of noun modifiers.
	Structural complexity	(1) Cl/T; (2) use of noun modifiers.
	Lexical complexity	(1) mean segmental TTR.

Note 1. TTR=type-token ratio; Cl=clause; AS=AS-unit; LREs=language-related episodes; T=t-unit; 1DE=first degree errors; 2DE=second degree errors; 3DE=third degree errors; TE=total number of errors;t=t-unit; It=Italian, Fr=French; SRA=unpruned speech rate (Rate A); SRB=pruned speech rate (Rate B); W=written task; O=oral task.

Note 2. An asterisk indicates a significant difference detected in favor of the complex task condition. Such signaling was employed only in those studies which measured CAF.

The exploration of number of elements was initiated by Robinson (2001) in what was called a "map task" and it has thus far been used in a number of studies. In his seminal study on number of elements, Robinson (2001) operationalized "elements" as the size of the map: there were few elements in the simple task and multiple elements in the complex one. Forty-four participants were randomly assigned the role of the speaker or listener. In each dyad the speaker had to describe the route on a map from point A to B and the listener had to draw the route and get to point B. Cognitive complexity was operationalized as the number of elements and the degree of familiarity with the area. The results indicated that it was the complex version of the task that generated more lexical complexity (as measured by type-token ratio), but greater fluency was found on the simple task version. Structural complexity was unaffected by simple versus complex task condition.

An instruction-giving map task was also used in Gilabert (2007c) and Gilabert, Barón, and Llanes (2009), although these two studies explored different phenomena (self-repairs vs. amount of interaction, respectively). In the former study, which measured the impact of task complexity on accuracy, cognitive complexity was operationalized as few easily discernible landmarks and moving along a single lateral axis in the simple task, and the complex version involved many not so easily distinguishable points of reference, in addition to moving along vertical, lateral, and sagittal axes. This operationalization was based on insights from cognitive psychology, in particular on the concepts of "landmark identification, path selection, direction selection and abstract environmental overviews" (Gilabert, 2007, p. 224). In the description of the number of landmarks no specific figures were provided, and the notions of "few" and "many" were employed instead. The complex task condition led to significantly more accurate linguistic output on eight out of nine employed measures.

The latter of the mentioned studies (Gilabert, Barón, & Llanes, 2009) used the same task as the one described above, but it enquired about the effect of cognitive task complexity in three task types on interaction moves and repair behavior. Regarding interaction, statistical significance in favor of the complex task version was detected on four measures.

Two studies investigated the dimension of elements in written production, where number of elements was operationalized as conditions to take into account when taking a decision. In a study by Kuiken, Mos, and Vedder (2005), 62 native speakers of Dutch studying Italian were asked to write a letter to a friend, in which they had to decide on a holiday destination. A number of requirements (3 in the simple version and 6 in the complex one) had to be taken into consideration when making the decision (considerations included aspects such as location, access to facilities, meal plan, etc.). In this study, enhanced accuracy in the complex task condition was observed in the case of all employed measures, with two of them showing statistically significant differences. By contrast, greater structural and lexical complexity was detected in the case of the *simple* task condition.

This study was replicated by Kuiken and Vedder (2007) using different measures. It explored the impact of increased complexity on accuracy and lexical variation with 76 Dutch speakers of French as a foreign language. The complex task condition triggered fewer errors on three out of four accuracy measures and greater lexical complexity as measured by type-token ratio, but the simple task condition led to slightly more structural complexity compared to the complex task.

The same set of tasks was used in a study by Kuiken and Vedder (2008), which enquired about the effect of cognitive complexity on accuracy and complexity. The participants were 167 Dutch L1 students of Italian and French. In both target languages, accuracy, as operationalized by different degrees of errors (a total of four), was found to be improved as the result of engaging in the more complex task. Of the other measures, only higher lexical complexity levels were detected in the case of French on type-token ratio. Structural complexity remained intact as a result of increased cognitive complexity.

A study by Michel, Kuiken, and Vedder (2007) operationalized “elements” as the number of electronic devices in a decision-making oral task. The participants were asked to leave a message on the answering machine of a friend, giving advice about buying either an MP3 player or a mobile phone. Regarding the operationalization of elements, in the simple task version two electronic devices were involved and six in the complex one, and in both complexity levels the devices differed from each other in seven relevant features such as price, color, capacity, et cetera, although this component of task design was not labeled “elements”. Regarding the results, the complex task condition led to greater accuracy on one out of five measures (errors/AS-units, and to enhanced lexical complexity as detected by lexical words/total number of words. Fluency decreased on the complex task condition, and structural complexity remained intact across the different conditions.

A study by Kim (2009) investigated the impact of cognitive complexity on the occurrence of language-related episodes in a picture difference task. The study involved 34 speakers of low and high proficiency levels, who performed the tasks in pairs. The picture difference task was based on a “beach scene”, and the difference between the simple and the complex version was operationalized as the number of elements present in each condition[9]. The results indicated that the complex task condition led to a significantly higher occurrence of LREs, particularly so in the case of low proficiency participants, and under the same task condition slightly more LREs were resolved correctly.

The effect of number of elements on oral performance was further investigated in two studies by Michel (2011 and 2013), in which two decision-making tasks (a dating task and a study task) were used.

In terms of the operationalization of elements, in the simple condition of the dating task the participants were asked to identify 2 out of 4 contestants (and 4 out of 6 in the complex version) who would make the best male-female couples based on several characteristics. In the study task the learners were supposed to decide which 2 out of 4 females (simple version) versus 4 out of 6 (complex version) would make the best study couple on the basis of a number of characteristics such as, for example, age, nationality, or reading behavior. In both simple and complex versions the number of characteristics to take into account was six. However, in both tasks only the number of contestants was labeled as "elements", and not the characteristics to take into account (which, however, were labeled as "elements" in other studies).

The 2011 study by Michel measured all four dimensions of performance, whereas the 2013 focused specifically on use of conjunctions. In the former study, greater lexical complexity was detected by means of Guiraud's Index, the other dimensions remaining intact. Particularly striking are the virtually identical figures obtained for both structural complexity measures employed. This finding was further reinforced in the 2013 study: only one out of five relevant conjunctions was affected

[9] It is not articulated in the study which aspect of the task the variable elements referred to, or the exact difference in the number of elements between the simple and the complex task condition.

by the complex version of the task, but *in the opposite direction to that predicted by the Cognition Hypothesis*: a higher score was found in the case of the simple task.

In a study by Sasayama (2011), elements were operationalized as the number of characters (2 in the simple task vs. 9 in the complex task), and the relative difficulty involved in distinguishing between them (female and male characters in the simple task, vs. some same-sex characters in the complex task). Written and oral narratives based on picture stories were administered to ten Japanese and Korean learners of English as a second language, in an attempt to measure the impact of cognitive complexity on accuracy and structural complexity. While accuracy, as measured by error-free clauses/t-unit was not affected by increases in cognitive load, a statistically significant difference in favor of the written complex task condition was detected in structural complexity (clauses/t-unit).

Levkina and Gilabert (2012) set out to investigate the impact of two cognitively different tasks, the difference between which was operationalized as the number of holiday destinations to choose from (2 in the simple task condition and 4 in the complex task condition)[10]. Forty-two Spanish learners of English engaged in the performance of these tasks, and their output was found to be significantly lexically more complex (as measured by Guiraud's Index) as a result of enhanced cognitive complexity. By contrast, fluency, accuracy, and structural complexity remained intact.

In a study by Sasayama and Izumi (2012), which measured the impact of cognitive complexity and pre-task planning on oral performance, two picture narration tasks were administered to 23 Japanese learners of English. Both cartoon stories were sequenced in the correct order, and the difference between the simple and the complex task condition was operationalized as the number of characters involved

[10] This study also investigated planning time.

in the story: two in the simple task and six in the complex task. Aside from the sheer number of characters, the difference consisted also in the relative difficulty associated with distinguishing between them (easy to distinguish in the simple task and difficult to distinguish in the complex one). While no effect was found on structural complexity or lexical complexity, the participants delivered their speech in a more accurate and fluent way *under the simple task condition*.

4.5.1 Number of elements: patterns in findings

Extensive research has been carried out into the dimension number of elements, each making its original contribution to the knowledge of language production, and in particular, performance under simple and complex task conditions. The various tasks mostly required the participants to make a decision, taking into consideration the number of elements. In all studies in this category, learners were confronted with two levels of cognitive complexity of the tasks. All studies contributed only partial and frequently minor evidence in favor of the Cognition Hypothesis, with most studies yielding a significant difference for only one of the dimensions, and sometimes only for one of the measures employed for a particular dimension. A systematic pattern, which emerged across the studies was that, overall, *accuracy and lexical complexity tended to increase as a result of enhanced cognitive task complexity*. At the same time, it is noteworthy that of all studies presented here, *only one found an effect of cognitive complexity on the structural complexity* of output (Sasayama, 2011), and it was in the case of the written production mode.

This state of affairs can be due to a number of reasons, one of them being the fact that, as was previously mentioned, the variable in question (number of elements) allows for a wide range of interpretations. As the available conceptual frameworks do not specify which aspects of tasks are subject to be labeled as “elements”, the adaptations of this dimension to the different task designs is preliminary and quite experimental. Depending on the task, “elements” have been options to choose from, considerations, conditions, resources, et cetera. The only consistent aspect across the studies which manipulated the variable “elements” is that a simple task was associated with few elements and a complex task was associated with many elements; however, it is fundamentally unclear how many elements are sufficient to distinguish two or more tasks in terms of their cognitive complexity. A closer inspection of how constructs were operationalized reveals several possible shortcomings.

4.5.2 Number of elements: degree of difference between a simple and a complex task

One aspect of design which stands out in this group of studies is that of the criteria adopted when establishing “simple” versus “complex” task conditions. As could be noticed in the above review, simple tasks were normally those, which included two elements, and complex tasks tended to include four or six. Why researchers opted for this choice, in a number of studies, however, seems to be a quite arbitrary decision and one based on intuition rather than empirical evidence or objective criteria. One of the exceptions is the study by Michel (2011), where it is argued that insights from cognitive psychology showed that “our working memory and reasoning limitations share a central capacity that is limited to relations between four variables”[11] (Halford et al., 2007, p. 240). Accordingly, the simple task of the present study giving four elements/ combinations should be within the human

[11] It should not go unnoticed, however, that this explanation from cognitive psychology is used as an a posteriori explanation to the obtained results, rather than a theoretical motivation for task design.

capacity limits whereas the complex task with six elements/ nine combinations should be beyond it" (Michel, 2011, p. 87).

Despite this evidence from cognitive psychology, in Michel's study, 4 elements in the simple version versus 6 in the complex one triggered only a minor qualitative change: lexical complexity increased in the complex task condition. Accordingly, one wonders what effect there might be if the number of elements had been stretched far beyond 4 or 6, in this study and others. For example, would the study by Michel have yielded different results had the number of features of electronic devices been made greater, and not only the number of electronic devices involved. In this and several other studies, authors have acknowledged that one possible explanation for the results obtained was perhaps insufficient differences between the simple and the complex task conditions. This is evidenced in the participants' subjective ratings in Michel (2011): on a 5-point Likert scale, the two tasks of hypothetically different cognitive complexity levels were found to be perceived as virtually identical in terms of difficulty (*M*=3.6 in the simple task and *M*=3.5 in the complex task), showing that the two tasks probably belonged to the same level of cognitive complexity, rather than falling into the simple and complex classification of the cognitive complexity continuum.

Of the studies analyzed here, only two allow for a comparison of the means obtained: Robinson (2001) and Michel (2011). When comparing the subjective ratings in these two studies on a common 9-point scale[12], it can be noticed that the simple tasks in both studies received a value of *M*=3.5 and *M*=6.48, and the complex task was judged as *M*=5.4 and *M*=6.3, respectively, exhibiting a substantial discrepancy,

[12] Michel (2011) used a 5-point scale. For the purpose of comparison of her results with those in Robinson's (2001) study, the results were transformed to reflect numerical values on a 9-point scale.

particularly in the case of the simple task. This once again shows a lack of uniformity from one study to another in what constitutes a simple and a complex task.

Interesting findings are those obtained in the study by Michel, Kuiken, & Vedder (2007), in which all dimensions of performance except for structural complexity, were boosted as a result of engaging in the complex task. The fact that structural complexity remained intact might indicate that in fact there was a partial trade-off effect between accuracy and complexity, but it could also be the case that structural complexity is sensitive to other variables than resource-directing ones, or to other resource-directing variables than the ones discussed here. I will return to this issue in the Discussion chapter.

4.6 Studies in which task complexity was manipulated simultaneously along ±reasoning demands and ±number of elements

The third and last category of empirical research concerns those studies in which the two variables, ±reasoning demands and ±elements, were manipulated simultaneously within the same task design. The information pertinent to the studies reported in this section is summarizes in tables 7 and 8.

In two studies by Ishikawa (2008, 2011) three tasks were used: no reasoning task (control task), simple task, and complex task. Whereas the "no reasoning" task required the participants to merely describe relations between people, and therefore reasoning demands were null, in both simple and complex task conditions the participants were required to attribute intentions and mental states to others in a situation in which human relationships changed in the workplace (e.g., between an employee and a secretary). In the simple task, 2 section members were involved in the change, and in the complex task 4 members were involved. Regarding the operationalization of complexity, the simple and complex tasks, as opposed to the control or "no reasoning" task, required the participants to successfully understand the psychological and other mental states which brought about a change in relationships between people. Increased intentional demands triggered more repair fluency behavior, but did not affect speech rate, and an increase in favor of the complex task condition was observed in all the other dimensions (however, only one measure per dimension was employed).[13]

The analyses carried out in this study were complemented by those in the study by Ishikawa (2011), which enquired about task difficulty, as measured by responses to an affective variables questionnaire, and the correlation between this subjective perception and performance in the L2[14]. As cognitive complexity increased, a greater number of negative and positive correlations could be observed between the different questionnaire items (Concentration, Time pressure, Anxiety, Stress, Difficulty, Interest, Ability, and Motivation), importantly indicating that the more complex tasks get, the more important learner predictions of task difficulty become.

[13] In this study, the participants were provided with three minutes' planning time prior to task performance.
[14] Performance data in terms of the dimensions of fluency, lexical and structural complexity, and accuracy, were taken from Ishikawa (2008).

Table 7. Summary of studies which simultaneously manipulated the variables ±number of elements and ±reasoning demands

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Ishikawa (2008)	ST[15]: describing human relationships between section members CT: thinking up the trouble triggered by a mistake and report about hypothetical human relationship changes +CT: same as CT but more section members involved (2 in CT vs. 4 in +CT).	3	none	"Oral report" (O)	26	Japanese/ English/FL
Ishikawa (2011)	Same as Ishikawa (2008)	3	AVQ. 10-point Likert scale *Results for Difficulty* ST: *M*=6.21; *SD*=2.17; CT: *M*=6.54; *SD*=2.04; +CT: *M*=7.67; *SD*=1.31.	Same as Ishikawa 2008	26	Japanese/ English/FL
Kim and Ventura (2011)	ST: information exchange CT: decision-taking +CT: RDs + elements (factors needed to take into consideration when taking a decision)[16]	3	none	Decision-making task (O)	128	Korean/English/ FL
Révész (2011)	ST: 500,000 dollars distributed among 3 projects; CT: 10,000,000 dollars distributed among 6 programs	2	AVQ. 7-point Likert scale *Difficulty: M*=3.76; *SD*=2.27[17]	Argumentative task (O)	43	multiple/ English /SL

[15] The original nomenclature employed in this study was "No reasoning task (or control task)", "Simple reasoning task", and "Complex Reasoning task". For purposes of clarity and comparability with other studies, in this table "no reasoning task" was coded as simple task (ST), the simple reasoning task was coded as "Complex task" (CT), and the complex reasoning task was coded as "Very complex reasoning task" (+CT), in line with other codings in this chapter.

[16] The information about the number of factors (=elements) was not provided.

[17] The values in the table represent participants' ratings on a 7-point scale. Values of 0 and 7, respectively, would indicate the highest possible rating in favour of the simple versus complex task versions (Révész 2011, p. 175).

Table 7 continued. Summary of studies which simultaneously manipulated the variables ±number of elements and ±reasoning demands

STUDY	Cognitive TC operationalization	Nº of TC levels	Independent TC differences measurement	Pedagogic/ research task type	N	L1/L2/ Context (FL vs. SL)
Kim (2012)	ST: information-gap task CT: information gap + decision-making (2 criteria) +CT: information gap + decision-making (4 criteria)	3	none	Decision-making task (O)	191	Korean/English/ FL
Malicka and Levkina (2012)	ST: 6 furniture items, 2 points of reference provided CT: 15 furniture items, no points of reference provided	2	(1) AVQ. 6-point Likert scale[18]. HP. ST: *M*=4.22; CT: *M*= 4.06 LP. ST: *M*=3.94; CT: *M*=3.65 (2) Time judgment task *Real time vs. estimated time:* 60% of HP speakers estimated the time correctly; 40% did in the LP group.	Instruction-giving task ("map task") (O)	37	Spanish, Catalan/ English/FL

Note. TC=task complexity; N=sample size; L1=first language; L2=second language; FL=foreign language; SL=second language; ST=simple task; CT=simple task; CT=complex task; +CT=the most complex task, AVQ=affective variables questionnaire; O=oral task; W=written task; *M*=mean; *SD*=standard deviation.

[18] In this study, "1"= most difficult task and "6"= least difficult task.

Table 8. Dependent variables and their measurement in studies which simultaneously manipulated the variables ±number of elements and ±reasoning demands

STUDY	Dependent variables	Measurement of dependent variables
Ishikawa (2008)	*Fluency*	(1) SRA; (2) SRB; (3) Dysfluency ratio*;
	Accuracy	(1) error-free T-units*
	Structural complexity	(1) SN/T*
	Lexical complexity	(1) Guiraud's Index*
Ishikawa (2011)		*This study measured correlations between task difficulty and the performance data obtained in Ishikawa (2008)*
Kim and Ventura (2011)	*Development of simple past tense morphology*	(1) nº of activity and stative types with past tense morphology/total nº of activity and stative types in OC.
Révész (2011)	*Accuracy*	(1) errors/AS*; (2) error-free AS/AS*; (3) self-repairs/errors
	Structural complexity	(1) Cl/AS
	Lexical complexity	(1) D*
	Conjoined clauses	(1) unbound; (2) bound; (3) final adverbial; (4) initial adverbial
	Occurrence of conjunctions	(1) and; (2) but*; (3) so*; (4) because*; (5) when/if
	Learning opportunities	(1) LREs/AS; (2) confirmation checks/AS; (3) clarification requests/AS; (4) recasts/AS; (5) metalinguistic talk/AS.
Nuevo, Adams, and Ross-Feldman (2011)	*Modified output*	(1) self-repair; (2) pushed output; (3) total modified output.
	Language learning	(1) gain scores for locatives.
Kim (2012)	*Interaction opportunities*	(1) all LREs involving questions; (2) LREs involving advanced questions.
	Question development	(1) movement to a higher stage in the developmental sequence.
Malicka and Levkina (2012)	*Fluency*	(1) SRA; (2) SRB.
	Accuracy	(1) Errors/AS; (2) Errors/Cl; (3) Spatial expressions errors/AS* (HP); Spatial expressions errors/Cl* (HP).
	Structural complexity	(1) words/AS*; (2) words/Cl*; (3) SN/ASU.
	Lexical complexity	(1) Guiraud's Index* (HP); (2) TTR; (3) Nº of types* (HP).

Note 1. SRA=speech rate A; SRB=speech rate B; SN=S-nodes; T=t-unit; Cl=clause; AS=AS-unit; LREs=language-related episodes; HP=high proficiency; LP=low proficiency; TTR=type-token ratio.

Note 2. An asterisk indicates a significant difference detected in favor of the complex task condition. Such signaling was employed only in those studies which measured CAF.

In a study by Kim and Ventura (2011), four different tasks related to university life were used: describing events at a university festival, hosting an American friend, sharing an experience from university orientation, and preparing for a mayoral election campaign. There were three complexity levels to each task: “simple”, “complex” and “+complex”. In the simple task the participants were only required to exchange information, in the complex task they were asked to take a decision, and in the +complex task they had to take a decision bearing in mind certain considerations (elements). However, no information was provided regarding the exact difference in the number of considerations between the complex and the +complex task. Compared to the simple task, the two complex tasks resulted in more development of past tense morphology. However, statistically significant levels of difference were not reached between complex and +complex tasks. The authors report that this could be due to insufficient differences between the complex and +complex task versions.

In a classroom-based study by Révész (2011), a simple and a complex version of an argumentative task was used. In this task the participants played the role of members of a personal trust foundation board and they had to assess two proposals for funding. The difference between the simple and the complex task was operationalized as the available economic resources ($500,000 vs. $10,000,000 in the simple and the complex task, respectively), and the number of projects the resources could be allocated to (3 vs. 6). The complex task version required the learners to justify the reasons for their choice. The study found that accuracy was greater in the case of the complex task (on two out of three measures), and so was lexical complexity (as measured by D). A mixed picture was found for structural complexity, in which the general measure (Clauses/AS-units) detected a decrease from the simple to the complex task condition, but the results for three out of five specific measures employed (operationalized as different conjunctions) demonstrated greater structural complexity as cognitive complexity increased.

In the study by Nuevo, Adams, and Ross-Feldman (2011), a narrative targeted at past tense and a decision-making task targeted at locatives were used, with three complexity levels each. In the narrative task, the participants were required to collaboratively write up a story. In the simple task, they were given pictures arranged in the correct order, and in the complex one plot information was not available. In the decision-making task the participants had to come up with the best sitting arrangement for guests. In the simple version there was one obvious sitting arrangement, which clearly matched the information about the guests. By contrast, in the complex version there were several imperfect solutions in addition to having to deal with more guests. Concerning the results, increased complexity did not foster modified output. The only area in which a significant effect was found was that of self-repair in the case of locative forms.

Kim (2012) enquired about the effects of three cognitively different tasks on the emergence of interaction opportunities and the development of question formation. 191 Korean learners of English engaged in task-based performance, and the hypothesized difference in cognitive complexity consisted in simple information transmission (simple task), information transmission and decision-making involving two criteria (complex task), and finally, information transmission and decision-making involving four criteria (the most complex task). Regarding interaction opportunities, the results indicated that the most complex task triggered a significantly higher occurrence of LREs than the complex one, and the same held true when the other two pairs of tasks were compared. The very complex task, compared to the simple one, led to a significantly higher occurrence of LREs involving advanced questions. Regarding question formation, the participants in the most complex task were shown to move to a higher stage in the developmental sequence.

Finally, Malicka and Levkina (2012) inquired about the differences in how designed cognitive complexity differences in tasks influenced the production of speakers of different proficiency levels in L2 English. In both proficiency groups the learners performed a simple and a complex task manipulated along spatial reasoning demands and elements, in which the learners were required to explain in as much detail as possible where in an apartment they would like their newly purchased furniture to be placed. In the case of the high proficiency learners, accuracy, lexical and structural complexity were boosted as the result of increasing cognitive complexity. In the low proficiency learners, fluency scores were higher on the complex task, with accuracy being negatively affected.

4.6.1 Simultaneous manipulation of ±reasoning demands and ±number of elements: emerging patterns

A few patterns emerge in the group of studies presented above. In most task designs, the participants were required to perform the action of narrating or describing in the simple version, as opposed to decision-making or justifying their choices in the complex version, taking into consideration the number of considerations involved. In this category of studies (i.e., simultaneous manipulation of two variables), the number of complexity levels was two (two studies) or thee (four studies).

The manipulation of elements and reasoning demands within the same task was predicted by Robinson as "likely to require a wider range of language than simpler tasks, for example greater use of lexical connectors, subordination, complex noun phrases, and a wider variety of attributive adjectives" (Robinson, 2001, p. 38).

The studies presented here overall seem to be in line with this prediction: the simultaneous manipulation of two variables, in this case number of elements and reasoning demands, seemed to be conducive to changes in performance and the results obtained in this category seem to be more systematic than those obtained in the studies in which the two variables were manipulated separately. More specifically, in those studies which measured CAF, accuracy and lexical complexity were boosted as a result of engaging in the most complex task. On the other hand, while in one study (Ishikawa 2008) structural complexity obtained a higher score on the complex task condition (with the provision of planning time), two studies found the opposite trend: it was *the simple task* which triggered more structurally complex speech.

The results obtained have to be treated with caution, however. For instance, in one study (Révész 2011), the fact that accuracy was positively affected by increased demands, and syntactic complexity was not, may indicate that in fact there was a trade-off effect between these two dimensions. On the other hand, it could also be the case that the design of a third and most complex task simply did not provide an opportunity for more complex syntactic structures to emerge when performing the task, perhaps because the differences in cognitive complexity were not salient enough.

It must be borne in mind, however, that an important part of the predictions of the Cognition Hypothesis is that resource-dispersing variables (e.g., planning time) should be kept low. In most of the studies reported here, except for one (Ishikawa, 2008) planning time was not provided, further exacerbating the cognitive load inherent in the tasks, in addition to the other manipulated variables.

4.6.2 Degree of difference between the simple and the complex task conditions

In none of the studies reported here the differences between all three complexity levels reached statistical significance, with some studies revealing differences between the simple and the most complex version, and others between the simple and the complex one. This possibly raises the issue of insufficiently salient differences between the different designed cognitive complexity levels between at least two tasks. The differences in cognitive complexity might therefore not have been stretched enough among the different task conditions to bring about predicted qualitative changes. One possible explanation for this finding, in the case of those studies in which the medium complexity level was complexified by adding more elements, could be that *elements on their own did not have the potential of changing the qualitative nature of speech performance*. The three categories of studies have so far been discussed in separation. Considering them jointly prompts a series of reflections and observations regarding both strengths and weaknesses, which apply to all three categories.

4.6.3 All categories of studies and all types of manipulation: shortcomings and challenges

A large body of research thus far has investigated how the manipulation of variables from the Triadic Componential Framework affects language production and development. Despite the fact that the studies reviewed here have been carried out within the same framework, they often lack unanimity and they present a series of challenges and drawbacks at the level of task design. A general tendency is that studies have largely produced unsystematic results and to date no clear patterns have emerged regarding how designed differences in task complexity affect performance.

Regarding task design, a crucial decision a researcher faces is that of establishing the difference between a simple and a complex task. The majority of studies did not provide a rationale for the hypothesized cognitive complexity differences. Only a handful of studies attempted to measures cognitive complexity differences independently from task performance, and it was typically carried out by means of participants' subjective ratings on an affective variable questionnaire, which inquired about the cognitive load of designed tasks, or more recently used time judgment or stimulated recall. Except for a few studies in which complexity was manipulated taking into consideration criteria from cognitive psychology, the decision about a task being "simple" or "complex" seems to be, at least in some cases, an arbitrary choice rather than a theoretically driven and empirically based decision. As was admitted in some of the research reviewed here, it might have been the case that the results obtained were due to insufficiently salient differences between the complexity levels of tasks. This, in the case of several studies, might have had an effect on the results obtained (a minor effect of the complex task on production, no effect at all, or mixed results).

Related to this is the fact that the operationalizations of central constructs in studies were often (over)simplified and of arguable validity. As was mentioned in chapter 2, there is a need to operationalize constructs at different levels and these are: the theoretical level of cognitive constructs (related to the ability to learn new languages), observational level of behavioral constructs (behavioral manifestations of aptitude), and an operational level of empirical constructs (test, measures, observation schemes) (Housen & Bulté, 2012). In a vast majority of studies, operationalizing constructs can be described as peripheral, which may have consequences varying from decisions about task design in a single study, to generalizability of findings across studies.

Operationalization in this body of work is normally carried out at a theoretical level, with no information about external measures employed to ensure differences in complexity, and in some studies it was simply lacking. There exist guidelines as to the criteria for what can be considered a valid or acceptable definition of constructs. According to Norris & Ortega (2003, 2012), they "should be defined in specific terms, such that observable behaviors may be obviously linked with them, and they should provide a clear identification of the theoretical assumptions that they represent" (Norris & Ortega, 2003, p. 720). However, rather than being a starting point in any study, in the available TBLT literature construct operationalization is still frequently at the periphery of empirical investigation.

As was previously suggested, the combination of weak task design and poor operationalizations in some studies may have had an effect on the results obtained which have been quite inconsistent across studies. It seems to be the case, though, that mixed results are more clearly observable in the case of some categories of studies than others. It seems overall that the simultaneous manipulation of the two variables had a greater effect on performance than manipulating these two variables separately. The dimension of elements, although quite intuitive, allows for such a wide array of operationalizations and interpretations that its effect on production may be as varied as the interpretations it allows. Whereas in the case of reasoning demands there is a certain degree of consistency across studies in what constitutes a simple and a complex task, and therefore a minimum systematicity in results obtained can be expected, in the case of elements, considering the lack of such guidelines, the results across studies represent a range of linguistic behaviors, with the different dimensions largely behaving in unpredictable ways.

An exception to this is the study on written production by Michel, Kuiken and Vedder (2007). Two factors can explain the findings obtained in this study. It might be that taking into consideration as many as six factors when making a decision was simply sufficient to trigger changes in the output. It cannot be neglected though that this study explored the role of complexity in written production, where inevitably the variable of planning time comes into play[19] and the pressure of providing an instant response, as is the case in oral tasks, is minor, if not inexistent. An overall conclusion is perhaps that the manipulation of elements alone is not sufficient to generate changes in quality, but only quantity of production.

It must be remembered, however, that the dimension "number of elements" is not only conceived of as the *number* of elements per se, but it also involves the *relationships between elements*, as was earlier suggested in cognitive psychology literature. Mental load is comprised not of how many elements need to be attended to, but how many must be attended to in parallel. In the body of work reported here, only one task design incorporated these premises, the fire-chief task (Gilabert, 2007c).
As noted by Michel (2011): "the manipulation of the single factor ±number of elements results in a quantitative change only, that is, 'more of the same'" (p. 88). It is therefore possible that those designs which manipulate the variable "number of elements" should additionally incorporate the dimension of *relationship among elements*, which could potentially exhibit a more profound impact on production.

In the case of reasoning demands, although a certain degree of consistency in instructions and operationalizations can be observed, studies have also produced

[19] Unless the task must be completed within an allotted time frame.

mixed results, with somewhat clearer trends observed towards greater accuracy in the complex task condition. It could be argued that studies carried out in this last category are more in line with theoretical premises present in the Cognition Hypothesis: tasks should represent as a sequence, and not a dichotomy. Where three or more cognitive complexity levels were involved, a continuum of complexity is observed rather than two extremes, as is the case in those task designs which employed a simple-complex dichotomy.

The issue strictly related to the results obtained is that of the measures employed. As could be observed, in those studies where CAF measures were used to capture the different linguistic phenomena, studies have used a wide variety of measures. Jackson (2013) reported as many as 86 different CAF measures employed to track changes in language production, which clearly impedes comparing results across studies. Other issues related to measures are: a) the use of general versus specific measures, b) the number of measures used for each of the dimensions, and c) the (in)adequacy of the measures chosen.

Some measures may have been too general to capture differences in production between the tasks of different complexity levels. Thus the use of more sensitive specific measures may be central to revealing qualitative changes in performance. In the studies reported in this review, the number of measures used for each dimension ranged from *one* per dimension to up to *ten* per dimension. Again, given a lack of common agenda or set of principles regarding what can be described as an acceptable or optimal measurement of the dimensions of speech production, researchers mostly opt for employing few (and mostly general) measures per dimension. Drawing conclusions about results on the basis of such minimal evidence seems to be an oversimplification and it perhaps does not do justice to the multifaceted nature of speech production. Using few measures has consequences for interpreting the results obtained in a single study, and for generalizability of findings across studies, and, ultimately, for confirming or rejecting the hypotheses under investigation. Therefore, rather than employing one measure per dimension, the use of several measures should be advocated, reflecting the different aspects of the same dimension; for example fluency can be broken down into speed fluency, repair fluency, and breakdown fluency. Employing a battery of measures rather than a single measure potentially helps to track complex relationships not only between dimensions of performance (such as accuracy and structural complexity), but also within them (such as, e.g., subordination ratio and mean length of clause) which, in turn, renders a study's results more reliable and generalizable.

It was the case in two studies that the Cognition Hypothesis was considered "confirmed" on the basis of only one measure per each of the dimensions (fluency, accuracy, and complexity). In Ishikawa (2008), the result for fluency was confirmed on only 1 out of 4 measures, and it was confirmed in the case of all the other dimensions. The question of whether the results would have been different had more measures been employed for each of the dimensions of production is an intriguing one. It raises the issue of the convenience of employed measures versus completeness or viability of results obtained, and more importantly, of a possible 'accidentalness' in the results obtained. Following Norris and Ortega (2009), "measurement practices in relation to CAF must become considerably more organic, in the sense that they need to capture the fully integrated ecology of CAF development in specific learning contexts over time, so as to help us understand how and why language develops or not within them" (p. 556).

Finally, two other issues can explain the relative inconsistency in the results obtained, and these are: derivation of tasks and suitability of learner populations to the designed tasks. The majority of studies reported here lacked an account on the process of derivation of tasks or a history of their development. Tasks were mostly derived in what seems to be an arbitrary fashion and were in most cases equally randomly administered to the learners. The process of Needs Analysis identified by Long (2006) as one of the prerequisites to any language program, has not been reported with the exception of three studies in this domain. One explanation which can account for this fact that is that in the majority of studies the learners were involved in a general English language course, which to some researchers (e.g., Skehan) impedes carrying out a needs analysis. However, this does not justify the lack of information regarding the origin of the tasks or how they fit the context of the study.

This challenge is related to the issue of the suitability of the designed materials to the learner population(s) performing them. In the majority of studies there seemed to be a mismatch between the designed tasks and the learners involved in doing them. With several exceptions, authors have failed to provide a rationale as to why the tasks the learners were exposed to were pertinent to the syllabus they were following at the time of engaging in task-based performance, how the tasks fit their interests, or their relevance to the learning process. Learners are bound to respond better (potentially with more motivation or interest) to those instruments which they see as connected either with their immediate learning context (classroom reality) or with their future needs.

The scarcity of information provided about the analyzed population of learners themselves is also notable. Apart from demographic data typically including the participants' age and L1, many studies in the domain do not (thoroughly) report such basic data as the participants' proficiency level, simplifying this variable to very vague notions of "beginner", "intermediate", or "advanced", which, besides being incomplete, leads to little comparability across research contexts and results.

Apart from the issues mentioned above, there a number of other problems which will only be mentioned here. Regarding task design, one might wonder whether a task performed under a dialogic condition is simpler or more complex than the same task performed in a monologic condition (e.g., Michel et al., 2007). Probably depending on the interlocutor's characteristics, and the relationship between two task performers, the dialogic mode can be, in addition to cognitive complexity, an extra source of influence either increasing the overall complexity further, or decreasing it.

Finally, most studies only enquired about the second stage in Robinson's SSARC model, that is, mainly resource-directing variables were manipulated, whereas in terms of task sequencing Robinson suggests three steps, as laid out in chapter 2: manipulating resource-directing dimensions (stage 1), manipulating resource dispersing dimensions (stage 2), and manipulating both dimensions simultaneously (step 3). In this sense, research into the effect of task features is still in progress, and the issue of sequencing, understood as employing the three steps in sequencing, has been virtually untouched. Most studies, as could be seen in the above review, employed a simple-complex dichotomy rather than a sequence of tasks.

On the basis of the above review of studies and their analysis, it can be observed that there is a lack of common agenda in a number of issues such as construct definition (the three levels mentioned before), measures (their optimal number, variety and adequacy), and various issues related to task design. Although every study is an original contribution to our understanding of the processes underlying producing or developing a second language, the lack of a uniform scheme in investigating these phenomena is a factor which leads to an incomparability of results and to concerns about the validity of findings in terms of confirming or rejecting the theories which guide them. There is an obvious need to replicate more studies in order to see how the same task affects production in various learner populations and with different research designs. Obtaining information about individual tasks, dichotomies, and sequences, and how these affect performance in different settings, can serve as a starting point to creating tasks administrable in the classroom context and extending task-based experiments to task-based curricula.

4.7 Summary of the chapter

This chapter has reviewed the research on task complexity relevant to the current study, and more specifically, it has focused on the empirical investigations carried out into two variables from the Triadic Componential Framework: ±reasoning demands and ±number of elements. An overview of studies was followed by a critical evaluation of the research carried out thus far in terms of methodological challenges. Major aspects dealt with have been task design and methodology, and among these, critical issues have been construct operationalization, designed differences in cognitive complexity of tasks, the use of external measures of cognitive complexity, the suitability of designed tasks to the learner population, the derivation of tasks and several issues related to measures.

4.8 Contextualizing the current study in light of previous work

The current dissertation aims to provide further insights into the dynamics between task complexity and L2 production. This section provides an outline of the investigated phenomena, pointing at the same time to some of the methodological issues encountered in previous research and the ways in which the design of the current study attempted to overcome them.

The fundamental component which this study has in common with previous work is that it investigates the impact of resource-directing variables on performance, understood as fluency, lexical complexity, structural complexity, and accuracy. Regarding task design and the manipulation of variables, it resembles those studies in which two variables, reasoning demands and elements, were manipulated within the same task design. It therefore resembles task design employed in studies such as Ishikawa (2008), Kim (2009), Kim and Ventura (2011), Révész (2011), and Nuevo, Adams, and Ross-Feldman (2011), all of which were reviewed in this chapter.

In the current study, complexity is manipulated along ±reasoning demands in the case of the simple and complex task, and the +complex task is manipulated along both ±reasoning demands and ±elements. An original contribution of this study is that it goes beyond investigating a dichotomy and it explores a phenomenon barely touched upon in previous research, that is, the issue of task sequencing. More specifically, this study researches individual task performance versus two sequencing orders, simple-complex and randomized. When tracking the qualitative changes in performance under the different task complexity conditions, in this study a battery of measures have been employed. While most of them are general production measures, in order to render comparability with previous research, a specific measure was used in the case of accuracy: target-like use of prepositions.

One of the major differences between this dissertation and the majority of studies reported here is that the current study was carried out in an English-for-specific-purposes setting. The process of Needs Analysis was therefore employed in order to derive pedagogic tasks (such was the case with the fire-chief task reported in Gilabert, 2005, and Gilabert & Barón, 2011). Although not all insights gathered from NA could be taken into account when making subsequent decisions about task design, care was taken for the designed tasks to resemble situations encountered in real life. As a consequence, the designed tasks are well-suited to the participants performing them in that they are related to the participants' future occupations.

Finally, regarding the hypothesized differences between the task complexity, which, as shown throughout this chapter, is an emerging, rather than common procedure, a total of three validation studies were carried out to ensure that designed differences in cognitive complexity match perceived differences. Two of them were carried out prior to the main study, and an affective variables questionnaire was also administered as part of the main study. The contextualization of the current study in the light of previous research takes me to the presentation of the investigated claims.

4.9 The claims under investigation

As was already mentioned in chapter 3, the main prediction of the Cognition Hypothesis is that increased cognitive complexity of tasks (simple-complex order) leads to an optimal performance simultaneously in the areas of accuracy and complexity, to a possible detriment of fluency. Robinson emphasizes that it is a specific *sequence* of tasks that has a potential to bring about qualitative changes. However, it is largely undefined how many individual tasks the sequence should include and how many cognitive levels a sequence should be comprised of, how many performances of how many sequences are necessary to lead to a qualitative change, and possibly at which intervals these are supposed to be performed. The objective of this study is to shed light on the dynamics between task complexity, sequencing, and language performance. More specifically, this dissertation aims to explore the aforementioned prediction about increasing complexity from two points of view: individual, immediate task performance on the one hand (different speakers performing tasks of different complexity levels), and different sequencing orders (order of increasing complexity versus randomized order) on the other hand (the same speaker performing a sequence of three tasks). The common goal of these two perspectives is to explore the impact of increased task complexity on four dimensions of L2 production: fluency, accuracy, and structural and lexical complexity.

Individual task performance provides information about how the different cognitive levels of tasks affect performance independently from each other. That is, this condition (i.e., performing three tasks of different levels of cognitive complexity) tests the possible effect on production of each of these tasks separately. The two sequencing orders, on the other hand, provide insights into whether, and how, output gets modified as a result of engaging in performing three tasks of different cognitive levels in a subsequent manner. In other words, the objective is to explore whether qualitative changes in performance, if at all detected, are unique to the sequencing order proposed in the SSARC model, or whether they occur as a result of engaging in performing any sequence of tasks.

4.10 Research questions and hypotheses

In light of the predictions of the Cognition Hypothesis and models associated with it (Triadic Componential Framework and the SSARC model of pedagogic task sequencing), and considering relevant research carried out to date, the current study aims to answer the following research questions.

4.10.1 Research questions related to the independent variable of cognitive task complexity

Research question 1a. How do tasks of different cognitive complexity levels, as manipulated along the variables ±reasoning demands and ±number of elements, affect task performance in the condition of sequences, as measured by fluency, accuracy, and complexity?

Hypothesis: Tasks of three different cognitive complexity levels were designed for the current study: a simple task, a complex, and a +complex task. Following the prediction of the Cognition Hypothesis, the complex task is predicted to lead to greater accuracy and complexity compared to the simple task. +Complex task, in turn, is expected to boost performance in the area of accuracy and complexity compared to the simple and the complex task. The opposite pattern might emerge in the case of fluency: following the predictions of the CH, the simple task will trigger the highest fluency, and it will decrease as cognitive complexity increases.

Research question 1b: How do tasks of different cognitive complexity levels, as manipulated along the variables ±reasoning demands and ±number of elements, affect individual, immediate task performance, as measured by fluency, accuracy, and complexity?

Hypothesis: In order to answer this research question, different participants performed a single task of a specific cognitive complexity level (i.e., either the simple, complex, or +complex task). Given that, to my knowledge, previous research did not include such a scenario, a null hypothesis is adopted, that is, the oral production of speakers performing different tasks will not be differentially affected by tasks of different cognitive complexity levels.

4.10.2 Research questions related to the independent variable of task sequencing

Research question 2a*:* How do different sequencing orders (simple-complex vs. randomized) affect performance in terms of CAF?

Hypothesis: Given that the sequencing order proposed by the Cognition Hypothesis has not been explored by contrasting it with an alternative order, be it complex-simple or randomized, the null hypothesis is adopted in the case of this research question. There will be no difference in performance between the simple-complex and the randomized condition.

Research question 2b: Does performing a task in a sequence lead to qualitatively different output compared to performing a task in isolation?

Hypothesis: Following the predictions of the Cognition Hypothesis, it is expected that performing tasks from simple to complex will positively influence the dimensions of accuracy and complexity, with fluency being potentially negatively affected. The simple task will elicit least accuracy and structural and lexical complexity compared to the two tasks of higher cognitive complexity levels. The complex task is expected to bring about qualitatively different speech than the simple task in the case of accuracy and complexity, and engaging in the +complex task is expected to further boost performance in these areas. The opposite trend is predicted to emerge in the case of fluency. Following the prediction of the Cognition Hypothesis, performance will decrease proportionally to the increases in cognitive complexity.

4.10.3 Research questions related to the independent variable of L2 proficiency

Research question 3a. How does task complexity affect the production of speakers of different proficiency levels, as measured by CAF?

Hypothesis: Given the role of proficiency in L2 performance is an unaddressed area both at the level of theoretical and empirical research, the null hypothesis is adopted in the case of this research question. There will be no difference in the way task complexity affects speakers of different proficiency levels in the L2.

***Research question 3b*.** Do low-proficiency learners benefit more from the simple-complex sequence than from the randomized one?

Hypothesis: Given that neither the issue of sequencing nor the issue of L2 proficiency as an individual difference have been addressed in literature, a null hypothesis is adopted in the case of research question 3b. There will be no difference in the way low proficiency speakers perform on the simple-complex versus randomized sequence.

Prior to presenting the experimental design in the current study, the upcoming chapter describes the process of Needs Analysis carried out for this dissertation, which was the source for subsequent development of research tasks.

CHAPTER 5
NEEDS ANALYSIS

5.1 Introduction

One of the constructs relevant to the current dissertation covered in chapters 3 and 4 was that of task selection. In chapter 3 it was shown that choosing sound criteria for selecting tasks for pedagogic or research purposes is important in order to ensure informed task design. Chapter 4, on the other hand, revealed that few empirical studies set out to select tasks in accordance with principled criteria. One of the proposals mentioned regarding the selection of tasks is Long's (2005, 2006) idea of Needs Analysis. Taking into consideration the theoretical premises by Long, and in order to fill the "task selection gap" observed in previous research, in the current dissertation a Needs Analysis was employed as a first step towards deriving research tasks, describing them in terms of variables, and exploring ways in which they could be manipulated.

5.2 Theoretical basis for Needs Analysis

As Long (2005) pointed out, "Every language course should be considered a course for specific purposes, varying only (and considerably, to be sure) in the precision with which learner needs can be specified - from little or none in the case of programs for most young children to minute detail in the case of occupationally-, academically-, or vocationally-oriented programs for most adults" (p. 1).

From here springs Long's (1990, 2005) idea of Needs Analysis (NA), that is, that in any language program the selection and sequencing of input the learners are exposed to should be motivated by empirical evidence rather than draw on vague intuition or feelings or impressions of what should, or must, be taught and (hopefully) learned. Following this idea, designing any language program requires a combination of the knowledge of applied linguists and expert or insider knowledge, whereas traditionally many programs were constructed on the basis of the insights from applied linguists or language teachers only. As a consequence, various language features and largely imprecise notions of "difficulty" or "usefulness" have served as a starting point in developing most syllabi.

In a Needs Analysis model, traditionally used "linguistic units (such as words, structures, notions or functions) are replaced by task as the unit of analysis" (Gilabert, 2005, p. 182) when it comes to deciding on what to teach, why, and in what order. Especially in those learner communities, which have a common goal in learning a foreign language, discovering what needs to be learnt by consulting experts, collecting surveys, doing observations, et cetera, in a given learning context, is an indispensible step for taking informed decisions about the curriculum.

Needs Analysis can be a valuable source of information about what constitutes daily tasks in a given domain, the frequency of performing each of them, the different levels of difficulty of the performed tasks, and the standards according to which the tasks need to be completed. Applying all these to curriculum design, the insights obtained from NA can facilitate taking informed decisions at four major levels of syllabus design: how pedagogic tasks can be drawn from target tasks (*task selection*); what is considered a simple and a complex task (*task complexity*); the optimal order of presenting them (*task sequencing*); and what constitutes a "good performance" of a task (*task assessment*). The knowledge collected about these four areas creates a sound foundation for translating the observed phenomena into a range of pedagogical options.

Given that the study reported here was done in an English for specific purposes context, discovering learner needs was considered a prerequisite for taking informed decisions about the critical issues of selecting the tasks and subsequently sequencing them. It has to be clarified that the objective of this NA was not to create a complete curriculum with individual units, but rather to use knowledge about learners' needs to derive several target tasks and apply them to the research context for the purpose of this dissertation only. Therefore a selective approach was taken towards the obtained data in that only a part of the information was used, and it served as a foundation for creating research tasks.

It must be pointed out here that Needs Analysis, although important at the level of task design, is not the object of study *per se* in this dissertation neither as a theoretical concept nor a practical method. Information provided on both theoretical and empirical levels is only meant to illuminate decisions regarding specific aspects of task design.

5.3 Needs Analysis carried out for this dissertation

The overarching goal of performing Needs Analysis for this dissertation was to discover, in general terms, how the different daily pieces of work in a hotel receptionist's job translate themselves, or fit into, the concept of "task". This NA was therefore carried out with the following questions in mind: 1. Can the daily tasks performed be described in terms of research tasks? 2. Can the components of the different tasks be described in terms of dimensions or variables? 3. If so, in what ways can these variables be manipulated to establish different levels of cognitive complexity? In order to explore these issues, two NA techniques were applied, interviews and observations in the workplace, which were suggested in literature

(Long, 2005, 2006) as illuminative, and they will be described in the following paragraphs.

5.4 Interviews

Regarding the first employed technique - semi-structured interviews - a total of seven were carried out, both with professionals in the field of tourism and tourism students in internships. The technique of interviews was used as a retrospective tool to reflect on the daily tasks and to gain insights into what the informants perceived to be a simple versus a complex task. As a result of employing this technique, a comprehensive list of potential target tasks was created potentially amenable to a research context.

Interviews revealed several categories of tasks receptionists have to deal with, for example: problem-solving tasks (e.g., overbooking), decision-making tasks (e.g., recommending a place to eat out), information-exchange tasks (providing customers with directions to get somewhere), and information-gap tasks (e.g., check-in task). A separate category of tasks is that in which the receptionists deal with requests on the customer's arrival and during their stay. The frequency of performing each of these tasks varies considerably from one day to another, but some of them evidently occur repeatedly, such as checking in the customers and solving various problems (e.g., malfunctioning of a device), while others are more out of the ordinary and therefore more sporadic (such as a client's request to translate the entire restaurant menu into a foreign language). Such broad scope of tasks allows for the occurrence of various interactional patterns, different cognitive demands placed on the person performing a task, and on-line computation options. Therefore they fall into the classification of task features proposed by Pica et al. (1993). In terms of information requester and supplier, the tasks are both one-way and two-way, in terms of goal orientation they are convergent and divergent, regarding outcome options they can be both open and closed, an in terms of interaction they can be tasks in which information exchange is required or optional. Quite a large variety is also observed in discourse genre as the tasks include narratives, instructions, descriptions, and reports, among others. Finally, in terms of cognitive categorization, they may be information-, reasoning-, and opinion-gap tasks.

Some of the tasks and categories of tasks mentioned above clearly lend themselves to be described in terms of their relative simplicity or complexity. For example, the tasks of giving recommendations and providing directions were reported by almost all informants as those in which there was a clear distinction between "easy" and "difficult" ones. In the case of giving recommendations, a simple task was reported to be one in which the receptionist was familiar with the area, knew the specific types of restaurants, and when there were multiple options to choose from (from the leaflets available at the hotel). A complex task is one with scarce options and the receptionist not knowing the area very well.

In the case of a direction-giving task, a complex version was reported as one in which the receptionist had to provide directions on the phone to someone who was driving a car, got lost on their way to the hotel, and did not know how to get to the hotel. This task presents an entire range of complexity features: understanding where exactly the customer is, being familiar or not with the area, lack of face-to-face communication, and potential anxiety accompanying it, the use of multiple resources to solve the problem, and probably also time pressure. An easy task consists in providing directions within a familiar, probably small area, with one or a couple of transportation options, and providing directions in face-to-face interaction. A task which combines the two tasks mentioned above is far from uncommon: recommending a place and giving directions to get to that place, which could either be conceived of as one single task with two components, or two separate tasks taking place subsequently.

The situation of overbooking was generally perceived by the informants as a difficult task to deal with. Some of the reported features contributing to perceived difficulty were the alternative options to choose from, discovering the best match for the client taking into consideration their circumstances and characteristics, and, last but not least, customer status (e.g., clients in the loyalty program). However, as was reported by some informants, the level of difficulty of this specific task frequently depended heavily on the customer's attitude; that is, on how demanding, anxious, and understanding they were. Yet another category of tasks was that of dealing with various incidents inside or outside of the hotel (e.g., missing items, or reporting a theft), which were considered by the participants as more or less difficult depending on the amount of narration required, internal complexity of the situation, and narrating in first versus third person.

This preliminary analysis reveals that some of the internal features of the aforementioned tasks probably lend themselves to be described in terms of selected variables from the Triadic Componential Framework. Some of the resource-directing dimensions are the options to choose from (±*number of elements*), apologizing and justifying (±*reasoning demands*), reporting an incident to a superior (±*perspective-taking* and possibly ±*here-and-now*), and providing directions (±*spatial reasoning*). On the other hand, resource-dispersing dimensions include the number of repetitions of the same situation (±*task familiarity*), the time available to respond to a problem, request, et cetera (±*planning time or ±time pressure*), number of stages involved in task performance (±*few steps*), and how related they are (±*independency of steps*). Some of the episodes reported by the informants probably go beyond the scope of the variables the Triadic Componential Framework features, for example face-to-face communication versus communicating at a distance or over the phone, which is a potential source of complexity in multiple tasks performed daily in many occupations.

Whereas many of the described tasks possess the right characteristics for being described in terms of the simple-complex dichotomy, and these are crucial to this study, performing a series of simple tasks for extended periods of time is not uncommon in a receptionist's job, and probably across professions. What adds to the relative simplicity of some tasks is the fact that every day there are bound to be multiple opportunities to perform them, such that each individual performance is a rehearsal before performing it again. Task repetition then plays a crucial role in developing the necessary skills to achieve satisfactory standards of performance on all tasks, both simple and complex.

5.5 Observations

The insights obtained in interviews were further complemented by a series of three four-hour long observations done in two different hotels in Barcelona. Observations provided an opportunity to collect discourse samples and they turned out to be particularly interesting for the variable ±reasoning demands incorporated into the task design. In observations, attention was paid both to how the receptionists solved the problems arising from an incomplete L2 system, and how they dealt with complex tasks. Instances of both linguistic and cognitive complexity could be observed and analyzed.

It was observed that the performed tasks represented a range of mental operations the receptionists engaged in, such as describing, explaining, apologizing, and convincing. These were observed to be present in a series of situations which required dealing with clients, ranging from giving instructions and recommending, through solving a problem with the malfunctioning of a device, to assigning a different room or hotel. Generally, cognitive complexity was observed to be the function of the number of the necessary mental operations and their diversity: the greater their number, and the more varied they were, the higher the cognitive challenge, and the greater mental burden the task imposed on the receptionist.

In terms of assessment, good performance on the aforementioned tasks was measured in terms of providing a satisfactory solution to a problem, very often with very little time available, few resources, and varying degrees of anxiety accompanying the choices taken and the solutions provided (which matches the idea of limited planning time and high levels of stress involved in performing complex tasks).

Overall, both interviews and observations provided useful insights into a receptionist's job. Although some parts of the information obtained were intuitive and may seem commonsense, they provided the necessary empirical evidence about cognitive complexity, and other sources of complexity, in this domain. As will be seen in the next chapter, this NA was a starting point for selecting target tasks, it helped take an informed decision about which tasks to design (task selection), how easy or difficult they should be (task complexity), and in which order they should be presented to the participants (task sequencing).

The NA carried out for this dissertation revealed that complexity is a multi-faceted construct in a hotel receptionist's job, in that internal complexity of tasks is frequently aggravated or alleviated by the characteristics of the interlocutor. An internally simple task can turn into a complex one depending heavily on this circumstance and others, such as the pressure to do the task on time, the number of people in the party, or interlocutor gender. Many of these considerations were taken into account when making decisions about task design for this dissertation. This and other issues are reported in the upcoming chapter.

CHAPTER 6
METHODOLOGY

6.1 Introduction

This chapter presents the experiment carried out for this dissertation. It starts with a general explanation of the context of the study. The research tasks created for the purpose of the experiment are then reported; two versions are presented in the order they were developed in. The first experimental version of research tasks was eventually redesigned and was a step towards creating the final version of tasks. Regarding the final version, an operationalization of the construct of cognitive complexity is offered in the context of the proposed task design, followed by the results of the pilot study intended to validate the hypothesized cognitive complexity differences. This is followed by the design of the experiment. Finally, the measures and statistical tools are described.

6.2 Participants

The experiment reported in the current dissertation was carried out in an English for specific purposes context, and specifically, in a curriculum focused on tourism. The subjects (*N*=117) were L1 Spanish and Catalan learners of English. They were all undergraduate students at a private college of tourism in Barcelona, and at the time of data collection there were enrolled in their first, second, or third year of university degree. The age of the participants ranged between 19 and 31 (mean age=21 years). English was an obligatory subject on the curriculum for all the participants involved, and the participants represented levels B1 to C1 in the Common European Framework of Reference for Languages.

6.2.1 Measuring participants' proficiency

The participants' proficiency was measured by means of Oxford Placement Test, which was extensively used in previous research (e.g., Tavakoli & Foster, 2008, Murphy & Roca de Larios, 2010, and Ahmadian, 2011, 2012). It is a multiple choice test with 60 items and it is targeted mainly at lexis and syntax. The overall pool of students to whom this test was administered was 205, and, on the basis of the test results, out of this pool 117 took part in the actual experiment. The test was administered during class time. The participants were told that this test was an independent measurement of their proficiency level.

In the current study, the participants were divided into "low" and "high" proficiency levels, as identified by means of Oxford Placement Test. "Low" proficiency participants were those who scored less than 29 on the placement test, and those who scored over 35 were classified as "high proficiency". The means obtained for both groups were: M=24.34 (SD=3.08) for the "low proficiency speakers", and M=39.5 (SD=4.30) for the "high proficiency speakers."[20] It must therefore be noted that the two groups did not represent two extremes of complexity, but a continuum of proficiency levels. A t-test showed that the difference between the two groups was statistically significant (p=.000).

[20] Later in this chapter a more detailed account of the means obtained is presented, because it distinguishes between scores obtained by the different conditions (simple-complex sequencing, randomized sequencing, individual task performance).

6.3 Development of the pedagogic tasks: two tasks versions and two pilot studies

The findings obtained from Needs Analysis prompted the design of pedagogic tasks. Two versions of the tasks - the first exploratory version and the final version - are discussed here. Only the final version of the tasks was used in this dissertation; however, given the fact that the first version influenced a series of decisions relating to the final version, its inclusion in this chapter was considered necessary to shed light on the final design. This section reports, in the case of the final draft, the results of the pilot study whose objective was to obtain information about the perceived difficulty of the tasks, which, in turn, leads to the operationalization of cognitive complexity.

6.3.1 First version of tasks: description, piloting, and findings

On the basis of the information obtained through Needs Analysis, two tasks were created in the first version, a “check-in task” and a “recommendation-giving task” (Appendix 1, pp. 329-340). There were three complexity levels to each of the tasks. In Pica et al.’s (1993) terms, these tasks were two-way, open, and convergent tasks. Two subjects were involved in the performance of each of the tasks, one of which played the role of the receptionist and the other that of the client. Both participants were given prompt cards, which explained their roles.

The check-in task was based on the idea of conflict of interests and it was targeted at the elicitation of pragmatic features and overall politeness on the part of the receptionist. The receptionist’s role was to do his or her best at accommodating the requests of the client, but these could be only partially accommodated due to the hotel’s policy, current circumstances, or the client’s supposed misunderstanding of the hotel's rules. Therefore none of the solutions offered by the receptionist was fully satisfactory to the client. The client's role, as specified in the task instructions, required them not to agree to any of the solutions offered by the receptionist, and to be consistent in their demands. In terms of the Triadic Componential Framework, the operationalization of increased complexity consisted in adding the number of requests and their level of sophistication to subsequent task versions.

The recommendation-giving task required the receptionist to suggest dining options to clients representing a range of profiles. It was designed to elicit the different linguistic forms pertinent to the language of making suggestions and giving recommendations. Through mental processes such as describing and convincing, the receptionist's role was to encourage the client to agree to the solution offered by the receptionist. There were six dining options to choose from and the designed differences in complexity levels consisted in the changes in the clients' characteristics: these were expected to place higher cognitive demands on the receptionist in the complex task as compared to the simple one.

These tasks were tested in a pilot study, the primary objective of which was to obtain information about whether the designed differences in complexity matched the learners' perceived differences in complexity. The second objective was to gain information about the kind of discourse which emerged as a result of participants engaging in task performance, and to what extent the instructions were followed. The piloting of this first version took place in the classroom context with students of tourism. Participants' speech was recorded and transcribed, providing a database of 12 dialogues.

The technique used to obtain insights about the cognitive complexity of tasks was participants' subjective ratings of the difficulty of tasks on a 9-point Likert scale

(Appendix 2, p. 341). The results of this pilot study, as measured by an affective variables questionnaire, did not reveal differences between the different task levels, as identified by descriptive statistics. In the case of the check-in task, the means obtained for the simple, complex, and the most complex tasks were, respectively, M=4, M=3.85, and M=3.71. In the case of the recommendation giving task, the simple task was assessed as M=5.14, the complex task as M=4.14, and the most complex task, as M=5. The tasks were generally considered to be of comparable, if not very similar, levels of cognitive complexity on the grounds that the values assigned to the tasks of hypothetically different levels were, overall, very similar. While this was particularly true of the check-in task, the results of the recommendation-giving task revealed that the simple and the most complex task were perceived as representing similar levels of cognitive challenge, and both were assessed as *more complex than the task representing medium cognitive complexity level*, which did not match the idea that simple tasks should be perceived as easy ones and complex tasks as difficult ones. The differences were therefore assessed as not salient enough in order for the tasks to be classified as "simple" versus "complex". Despite the fact that there was a mismatch between designed differences in cognitive complexity and the participants' perception, this pilot study guided certain features of the final task design.

First, one of the insights obtained from the previously described NA was confirmed, that is, that in a two-way task cognitive task complexity was mediated by the individual characteristics of the interlocutor (the client, who, in this case, was played by another student), in addition to the internal characteristics of the designed tasks. When comparing the results of the pilot study with the transcriptions of students' production, it became clear that the greater demands the client put on the receptionist (e.g., pressure, persistence, and an overall adamant attitude), the more complex the receptionist perceived the task to be. While this was not the case in all performances, this external source of complexity was considered beyond the control of the researcher, unless trained interlocutors were to play the role of the client. In other words, having an interlocutor potentially conflated different sources of influence: task-internal and individual factors.

Second, the subjects in the pilot study also reported that it was not clear to them when they could consider the task completed, as the information on the prompt cards explicitly instructed them to disagree with each other (both tasks were open and convergent). These instructions were considered confusing because it seemed that the task could "go on forever", as one of the participants reported.

These observations influenced one of the crucial decisions in the final task design, that is, that the tasks be *monologic, one-way tasks*. This decision might to some extent compromise the authenticity of the tasks, but as this study does not intend to measure interaction, this decision was considered necessary to focus on internal cognitive complexity of the task and eliminate possible external sources of influence, which might either intensify or lessen overall complexity. It was also decided that, rather than targeting specific features of language code (pragmatic features or the structures pertinent to making a recommendation), performance would be measured on four dimensions of language production, fluency, accuracy, and structural and lexical complexity. This decision was also motivated by the fact that the final version of the research tasks did not target any specific lexis or syntax, and analyzing speech using CAF measures ensured comparability of this study findings with previous research.

This pilot study also revealed that perhaps an affective variable questionnaire alone may not be a sufficient predictor of differences in cognitive complexity. The participants found it challenging to verbalize their motivations for judging a task as easy or difficult, and one of the few reasons that were provided was that they lacked certain linguistic items to express themselves with an acceptable degree of accuracy. The received feedback did not address the potential role of few versus multiple elements, or the amount of thinking required, as factors contributing to complexity. Most probably this state of affairs was because the complexity levels were not sufficiently salient, so when difficulty was measured on the second version of tasks, complementary techniques were employed to ensure a more complete insight into the construct of cognitive complexity.

6.3.2 Final version of tasks: scenario, instructions, and operationalization of cognitive complexity

Given the results of the first pilot study and the feedback obtained from the participants, the first version of tasks was redesigned and new tasks were developed. A sequence of three tasks was designed, entitled "*Welcome to the 'solving problems at reception training session' series*".

6.4 Scenario

In this version of tasks, the participants were asked to imagine that they had just started working as hotel receptionists and the hotel they worked for was famous for its personal touch when dealing with its clients. According to the hotel's policy, all new employees had to engage in training sessions in order to practice their skills and become more self-confident when it comes to dealing with the clients.

The designed tasks represented three situations in which the receptionist had to solve a problem at the hotel reception. The receptionist's role was to relocate the clients to different rooms or hotels due to a problem which arose in the hotel. There were two main components to each of the three tasks, and these were: 1) the profiles of the clients, and 2) the room or hotel options to choose from. In the simple task the participant had to describe the different options offered by the hotel, and in the other two tasks (the complex and the most complex task), the participant needed to take a decision about which room or hotel option best matches each of the clients on the basis of a series of characteristics made available about both the clients and the room and hotel options. Except for the simple task, the tasks were therefore based on a mismatch between the clients' requests and the options offered by the hotel. There was not a single perfect solution the receptionist could offer, but rather several imperfect ones, which rendered the task open.

6.5 Operationalization of cognitive complexity

A top-down approach was adopted when dealing with operationalization, that is, complexity is discussed at three levels, starting with the most general one, the theoretical one which takes us back to the Cognition Hypothesis, followed by the observational level, and finally, the tasks are presented at the lowest, operational level of empirical constructs, in which they are discussed in terms of the manipulated variables. When dealing with the last of the mentioned levels, first the two manipulated variables – reasoning demands and elements – will be discussed separately, and then each of the tasks will be analyzed.

6.5.1 Theoretical level

As was mentioned in the previous chapter, this dissertation adopts Robinson's Cognition Hypothesis as its theoretical basis. Briefly explained, the CH suggests that increased cognitive complexity leads to qualitative changes in language production. The dimensions of accuracy and complexity are expected to benefit from increases in complexity, whereas performance in the area of fluency may deteriorate. Such a pattern of linguistic behavior is possible because there is no *single* volume of attention which runs out of resources, as a result of which two areas of performance can be attended to simultaneously. Increases in complexity may trigger these performance patterns both in immediate, individual task performance in the L2, and possibly also in L2 development. Drawing on this idea, a sequence of three tasks was created for this dissertation, with salient differences in cognitive complexity. In light of the CH, each of these tasks has a potential of contributing to performance in unique ways, and performing a sequence of tasks may lead to qualitative differences in the areas of fluency, accuracy and complexity.

6.5.2 Observational level

In Robinson's Triadic Componential Framework, associated with the Cognition Hypothesis, cognitive complexity is broken down into more or less operationally feasible and researchable variables, in terms of which components of tasks can be described, and which need to be defined within a specific research design. In this framework, resource-directing and resource-dispersing variables are controllable by the researcher and they can therefore be deliberately manipulated in task design. "Number of elements" and "reasoning demands", two dimensions under investigation in this study, are two variables that can influence the relative complexity of a task. As was mentioned before, “reasoning demands” refers to the amount of online computation required to perform a task, whereas “number of elements” refers to the number of occurrences of a specific task component. In the task design reported in this study, regarding elements, increased cognitive complexity consists in adding more elements to the existing ones, and in the case of reasoning demands it involves augmenting the number of mental operations involved so as to pose a challenge so substantial on the learner’s attentional resources that it leads to the restructuring of the L2 system. Linking this design feature to the issue of task sequencing and Robinson’s SSARC model (Robinson, 2010), when tasks are performed in a sequence, a simple version of a task on both reasoning demands and number of elements plays a facilitative role of stabilizing the L2 system, while a complex task is supposed to automatize the access to L2 forms, and the very complex tasks is expected to restructure the L2 system in search of optimal ways to deal with high cognitive demands. In the task design adopted in this dissertation then, the expected qualitative changes in performance are the result of accumulated strength of the two variables: reasoning demands and elements. They are expected to place increasingly higher demands on the speaker, and it is the gradual build-up of demands that is hypothesized to bring about qualitative changes in performance of the different tasks.

6.5.3 Operational level

6.5.3.1 Reasoning demands

In terms of reasoning demands, the tasks designed for this dissertation were manipulated along the build-up of mental operations needed for successful task completion. “Mental operations” were conceived of as engaging in thought processes such as describing, recommending, apologizing, and justifying. Increased cognitive complexity was the function of the number of these operations and their simultaneous occurrence within a single task.

6.5.3.2 Number of elements

The variable “number of elements” embraced two components of tasks: client profiles and, depending on the task, hotel or room options. It must be clarified, though, that cognitive task complexity did not refer to the number of clients or room options, but rather to *the number of characteristics thereof*. That is to say, only those components of the tasks the number of occurrences of which changed across the different task versions were labeled “elements”. Therefore, the number of the clients being three, or the number of room/ hotel options being five across the task versions, was not referred to as “elements”, because they were kept constant across task versions. The following paragraphs offer the operationalization of complexity (both reasoning demands and elements) in each of the three tasks. The operationalization is summarized in table 9.

Table 9. Operationalization of cognitive complexity in the current study

	Reasoning demands	**Elements**	
	Mental operations required	*Categories of characteristics*	
		CLIENT	ROOM/ HOTEL
Simple task	describing	Brief description	- price - location - view - meal plan - discount - facilities (6 items)
Complex task	describing apologizing recommending	- original booking - budget - view - meal plan	- location - price - view - meal plan
+Complex task	describing apologizing recommending justifying	- original booking - price - length of stay - requests - what they like about the hotel	- location - public transport - price - internet - sea view - terrace - parking lot - swimming pool - meal plan - hotel availability

6.5.4 Operationalization of cognitive complexity in the simple task

In terms of reasoning, in the simple task version the participant was only required to perform the mental operation of *describing* the rooms currently available at the hotel. In this task version the profiles of the clients varied in terms of who the clients were, but only a general description of three typical clients was provided: a young couple, scholars, and groups of friends. The dimension of elements was operationalized as the information about the rooms the receptionist had to communicate to the clients, and it included location, price, view, and meal plan. However, the presence of the number of elements in this task did not imply choosing or prioritizing. This task could be successfully transacted simply by providing the customers with all the relevant information. Therefore this task can be described as a *one-way information-exchange task with simple information transmission*.

A number of factors rendered this task simple: few mental operations involved, simple information transmission rather than having to contrast and compare pieces of information, and the fact that "number of elements" was *present* in the task, but it was not *manipulated*. In terms of cognitive psychology, this task is composed of isolated, rather than interactive elements, understood as the instructions to follow and the actual task components, and it is due to the isolated character of these elements that they can be "easily held and processed in working memory" (Pollock et al., 2002, p. 66). All of the attentional resources necessary for task completion were focused on *transmitting* basic pieces of information and there was no need to split attentional resources between information transmission and another mental operations.

The simple task is displayed in Appendix 3 (p. 342).

6.5.5 Operationalization of cognitive complexity in the complex task

In the complex task version the mental operation of describing the available options was complemented by two other mental operations: *apologizing* to the client for the situation and *recommending* them the best option. Demands were therefore higher in this task than in the previous one as more actions were required for the task to be successfully completed, and the participant had to find the best match for each client. The attention had to be split between the instructions, client characteristics, and room options.

Cognitive complexity was also simultaneously increased along the dimension of elements. This variable was represented by two task components: the clients' and the rooms' characteristics. In the former component, four categories of characteristics were included: original booking, budget, view, and meal plan. The characteristics of the clients are better understood in terms of categories of elements, given that sometimes more than a single piece of information was provided (e.g., the element "original booking" included information such as the number of rooms, the type of room, and the date of arrival). As for the characteristics of the rooms, four pieces of information (=number of elements) were provided about every room, and these were: location, price, view, and meal plan. This was largely similar to the simple task. A simultaneous increase in reasoning demands and elements required a great deal of attention being allocated to multiple task components, which was expected to place higher demands on working memory, and to require greater control overall of one's linguistic resources.

Linking the sort of manipulation adopted in this task to insights from cognitive psychology, it can be stated that this task was built of multiple relations which had to be processed in parallel in order to successfully transact the task: there was high element interactivity, because different pieces of information had to be held in working memory in parallel (comparing and contrasting clients' characteristics and room options), alongside several mental operations.

The complex task is displayed in Appendix 4 (p. 346).

6.5.6 Operationalization of cognitive complexity in the +complex task

In terms of reasoning demands, all the components of the complex task (i.e., apologizing, describing, and recommending) were present in the +complex task version. An additional component in the +complex version was that of *justifying* one's choice when recommending an option. Compared to the complex task, in the +complex task more categories of elements were involved, both in the characteristics of the client and in the available options. There were therefore many related, interacting factors, which could not be considered in isolation, in a step-by-step fashion. It was the accumulation of these factors which rendered this task's intrinsic load higher than that in the other two tasks. Whereas in the previous task the participants had to choose different rooms for the clients, in this task the learner had to choose a different hotel due to overbooking. This task was based on such a hypothetical scenario because, as was found in NA, the situation of overbooking was generally reported to be a difficult one to deal with.

In terms of the clients' characteristics, more categories of elements were present (five, as opposed to four in the complex task). The categories were: originally booked rooms, price, length of stay, special requests, the clients' likes and dislikes about the original hotel, and 'additional information' (the latter category applied to only one of the clients). The number of hotels' characteristics was ten, and these were: location, availability of public transport, price, availability of internet, sea view, terrace, parking lot, swimming pool, meal plan, and availability of the hotel. As was the case in clients' profiles, in room and hotel options it was also the case that in some categories more information was provided than in others (such as the category of prices in task 3, in which four different room types and their prices were provided). As was the case in the complex task, in this one the participant had to discover the best match for each of the clients, taking into consideration all the available bits of information. Also, similarly to the complex task, cognitive complexity in the +complex task was manipulated both along ±number of elements and ±reasoning demands, with one more mental operation (justifying) to carry out and more elements to take into account.

It was the simultaneous rather than individual occurrence of these two variables that distinguished the simple task from the other two tasks. The accumulation of characteristics and mental operations was considered to gear the participants' attention simultaneously to various components of the tasks (instructions, specific client profiles, and options to choose from), which was expected to make the most complex task version more complex than the simple one. In contrast to the simple and the complex task, the task at hand required balanced attention to, and parallel processing of the instructions, elements inherent in tasks, and multiple mental operations.

The +complex task is displayed in appendix 5 (p. 350).

6.5.7 Summary of cognitive complexity manipulation

The manipulation of cognitive complexity adopted in this chapter was the outcome of insights from empirical evidence previously identified through Needs Analysis and the theoretical basis provided by the Triadic Componential Framework, and more specifically, two resource-directing variables: ±reasoning demands and ±number of elements. As was reported in the literature review chapter, a combination of these two variables within a single task design was adopted in several previous studies[21]. Although measuring the joint effect of merging these, or any two resource-directing variables, rather than their separate effect is still in its infancy, the research carried out so far has hinted at the possibility that variables may be better able to bring about qualitative changes in production when they operate jointly rather than in isolation.

6.6 Rationale for the language of input

In all three tasks both the instructions and the input were provided in L1 Spanish, the rationale for which was two-fold. First, by providing the input in the speakers' L1, the tasks were closer to realistic tasks than if the input had been provided in the target language. In NA it was shown that when faced with a situation when the use of L2 English was required impromptu, it was quite uncommon to have the necessary information needed to complete a task immediately available in a language different from the mother tongue.

Second, one of the concerns about performance was that if the input had been presented in the target language, the subjects would have been provided with ready-made solutions when transacting the task. Performing the task would have perhaps largely involved repeating, or reading out, the information available in the prompt cards, and linguistic output would have therefore not been the subject's unique output, but rather, in the best case, a rephrasing of the information from the input.

[21] See the following studies covered in the literature review section: Ishikawa (2008); Kim (2009); Kim & Ventura (2011); Révész (2011); and Nuevo, Adams, and Ross-Feldman (2011).

To conclude, providing input in the L1 served a double goal of making the tasks resemble similar tasks encountered in an authentic hotel reception setting, and ensuring that the learners used their own linguistic resources.

6.7 Piloting of the pedagogic tasks: independent task complexity measurement

The manipulation of cognitive complexity adopted in these tasks was subject to an external measurement to validate the perceived hypothetical differences in complexity. As was mentioned in the literature review chapter, measuring task complexity independently form task performance is crucial to a sound task design, and it was identified as a basic yet often neglected procedure in empirical TBLT research.

Three techniques were used to measure differences between the three tasks: an affective variable questionnaire, time judgment task, and stimulated recall. In the affective variable questionnaire, the participants were required to mark their judgments on a 9-point Likert scale (1 being “low cognitive load” and 9 being “high cognitive load”). In the time judgment task the participants were asked to assess the time it took them to perform the task and their estimations were subsequently compared with the real time. Finally, stimulated recall was carried out with four participants, and they were required to reflect on complexity in a retrospective way. This qualitative measure of cognitive complexity complemented the other two quantitative techniques and it provided first-hand information about whether, and why, certain tasks were considered more difficult than others.

6.7.1 Procedure

There were 23 participants in this pilot study. The tasks were piloted in two sessions (in some cases individual sessions were scheduled). Each participant played the role of a hotel receptionist and they had a potential 'client' sitting opposite them (another student), whose task was to keep track of the real time it took the receptionists to complete the task. Each of the receptionists was given a 22-page dossier which included, in the following order: the task, time judgment task, and an affective variable questionnaire (the latter two are displayed in appendix 6[22], p. 354). This procedure was repeated three times, once for each of the three tasks. To counterbalance the sequence of tasks, half of the participants performed a simple-complex order and the other half performed a randomized order. Stimulated recall was done with several participants once the entire sequence had been completed.

6.7.2 Results

6.7.2.1 Affective variable questionnaire

The participants' perceived difficulty, as measured by an affective variable questionnaire, was subjected to statistical analysis using descriptive statistics, and a paired samples t-test. The results of descriptive statistics are displayed in table 10.

[22] The affective variables questionnaire enquired about four items (appendix 6, but this section reports the results obtained only for the items "Difficulty" and "Mental effort".

Table 10. Descriptive statistics (difficulty and mental effort)
Means and standard deviations

Dependent variable	**Simple task**		**Complex Task**		**+Complex task**	
	M	***SD***	***M***	***SD***	***M***	***SD***
Difficulty	3.35	1.36	4.3	1.27	5.8	1.82
Mental Effort	3.48	1.5	4.43	1.19	6.1	2.02

As can be observed in the descriptive statistics, there was a gradual pattern of increase in perceived difficulty in the case of the two variables: the simple task was perceived as the least difficult one, the complex task as more difficult than the simple one, and the +complex task as the most difficult one. It can also be observed that the means obtained for mental effort were generally slightly higher than those for difficulty. It also seems that the perception of difficulty between the complex and +complex task was greater than that between the simple and the complex task.

A paired-samples t-test was conducted to further investigate the perceived levels of difficulty of the tasks. The results indicated that for each variable (difficulty and mental effort), the three comparisons (simple vs. complex task, complex vs. +complex task, and simple vs. +complex task) turned out to be statistically significant. Given these results, it was concluded that the operationalization of cognitive task complexity was matched by the task performers' perception.

6.7.2.2 Time Judgment Task

Time Judgment Task measured the differences between perceived time it took the participants to perform a task and the real time. The results are displayed in table 11.

Table 11. Time judgment task. Descriptive statistics (means and standard deviations)

Simple Task		**Complex task**		**+Complex task**	
M	*SD*	*M*	*SD*	*M*	*SD*
73.6	74.61	56.21	60.17	54.94	82.47

Table 11 compares the results obtained for the perception of time it took the subjects to perform the task. The means denote the difference, in seconds, between the estimated and the actual time it took the participants to perform the tasks. The prediction was that as complexity increases, the greater should be the gap between the real time and the estimated time. Although this measure has not been extensively used in TBLT literature, examples of studies which employed it include Baralt (2010), Malicka & Levkina (2012), or Révész, Michel, & Gilabert (2012). As can be observed in descriptive statistics, the opposite trend was revealed: the greatest mismatch between the real and guessed time was detected in the case of the simple task, and the values obtained for the two other tasks were lower, the difference between them being marginal. The differences were not statistically significant.

6.7.2.3 Stimulated recall

As was mentioned before, the two quantitative techniques were complemented with a qualitative, retrospective measure of stimulated recall. The participants were asked two basic questions: 1. Did you think there were differences in difficulty between the tasks you performed? 2. Why did you think they differed in difficulty?

In three out of four cases the information obtained in stimulated recall overlapped with the hypothesized differences in cognitive complexity. In their answers the participants indicated that the more options they had to deal with (i.e., client and room or hotel characteristics), the more difficult the task was because they had to pay simultaneous attention to many aspects of the tasks. They also indicated that task difficulty increased when more dense information was included in the instructions, that is, if they had to perform several actions at the same time (i.e., describing, apologizing, convincing, etc.). These insights matched the designed differences in cognitive complexity in both of the manipulated variables, number of elements and reasoning demands.

6.7.3 Pilot study: summary of results and conclusion

Out of the three techniques used in the study, differences in the perception of difficulty were confirmed by two of them (affective variables questionnaire and stimulated recall). The results of time judgment task were assessed as inconsistent with the two other techniques used, possibly pointing to the fact that this technique might not have been sensitive enough to capture the differences in perceived complexity. In the process of collecting data about time judgment task, several episodes were observed in between-participant interaction, which were thought to have possibly influenced the results (instances of possible cheating when estimating the time, participants forgetting to assess the tasks on completion, which resulted in some data missing, etc.). Also, at the beginning of the data collection each participant was asked to estimate each task, which was considered to have influence on the patterns of estimation itself: if the participant knew they had to judge every task, it might have drawn their attention to how much time they took. Therefore when approximately half of the data were collected, the rest of the participants were asked to assess tasks randomly, without knowing whether or not they would be asked to estimate the invested time. All these episodes might have influenced the results.

Taking into consideration the tendencies observed in descriptive statistics, as well as the insights gained through stimulated recall, it was considered that the designed differences in cognitive complexity were large enough for the tasks to represent different levels of cognitive challenge. Given that two out of the three techniques pointed in the desired direction (i.e., simple tasks were considered "easy" and complex tasks were considered "difficult"), this information was considered sufficiently robust to state that designed differences in complexity reflect perceived differences and they were therefore ready to be used in the main experiment. The upcoming sections provide information about the experiment itself, and issues such as design, procedure, statistical tests, and measures are dealt with.

6.8 The experiment

6.8.1 Study design

This study explored the effect of task complexity and task sequencing on production. More specifically, two variables from the Triadic Componential Framework, ±number of elements and ±reasoning demands, were manipulated across three tasks to investigate their effects on individual task performance versus two sequencing orders of tasks, simple-complex and randomized.

There were therefore two groups in the study:

Group 1: individual task performance (baseline data)

Group 2: sequences (simple-complex vs. randomized)

In the first group (individual task performance), three subgroups of subjects were involved, and each performed *one task of a specific cognitive complexity level*; that is, each subgroup performed the simple, the complex, or the +complex task. 18 participants performed the simple task, 19 participants performed the complex task, and 20 participants performed the most complex task. The objective in this group was to measure the effect each of the individual tasks had on performance.

In the second group (simple-complex and randomized sequencing), the three tasks were performed *by the same participant in a subsequent fashion*. As was mentioned in chapter 3, the available theoretical frameworks (in this case the SSARC model of pedagogic task sequencing) does not define how many tasks a sequence should be comprised of, which calls for an operationalization of this construct. “Sequencing” was operationalized as performing a series of three tasks in one sitting, at short, 1-minute intervals, with unlimited online planning time. In the simple–complex condition, the subjects (*N*=30) performed all three tasks in the order of increasing cognitive complexity, starting with the simple task and finishing with the +complex one. Each participant in this condition performed exactly the same order.

Randomized sequencing was operationalized as administering the tasks in one of the five possible orders, which results from a total of three tasks:

a) simple, complex, +complex;

b) complex, simple, +complex;

c) complex, +complex, simple;

d) +complex, simple, complex;

e) +complex, complex, simple

In this condition the subjects (*N*=30) performed a randomized sequence of tasks, with six subjects assigned to one of the randomized orders described above. Specific information about each of the conditions is summarized in Table 12 and explained below.

Table 12. Experimental design of the study

Experimental design								
Individual task performance			***Sequence***					
			Simple – complex	*Randomized*				
N=18	*N=19*	*N=20*	*N=30*	*N=6*	*N=6*	*N=6*	*N=6*	*N=6*
S	C	+C	S C +C	C S +C	C +C S	S +C C	+C S C	+C C S

S= "Simple task"; C= "Complex task", +C = "Most complex task"

Within the sequencing conditions (simple-complex and randomized sequencing), half of the participants in each condition were "low proficiency" (*N*=15), and the other half were "high proficiency" (*N*=15). As was explained at the beginning of this chapter, the participants took a 60-item placement test. The means obtained by each group were as follows:

1) low proficiency participants (both simple-complex and randomized sequencing conditions): *M*=24.34 (*SD*=3.08);
2) high proficiency participants (both simple-complex and randomized sequencing conditions): *M*=39.5 (*SD*=4.30);
3) group performing task 1 (individual task performance): *M*=28.88, *SD*=7.67;
4) group performing task 2 (individual task performance): *M*=32.36, *SD*=6.63;
5) group performing task 3 (individual task performance): *M*=33.78, *SD*=9.54.

A series of t-tests revealed that the differences were statistically significant between low proficiency participants and high proficiency participants (groups 1 and 2 above; *p*=.000), but none of the other comparisons was significant (simple-complex vs. randomized sequencing group, and individual task performance vs. the sequencing groups).

6.8.2 Data collection procedure

Data collection took place in the first semester of the academic year 2012/ 2013, from November 2012 until January 2013, in a total of approximately 10 sessions. All participants were informed at least a week prior to data collection that they would participate in an experiment. The data collection took place during regular class time in large classrooms with each student performing the task individually one-on-one with the researcher[23]. The other students were assigned other English tasks in the meantime. As the participants did not know that only students of certain proficiency levels were the object of the study, the data were collected with all students present in class at the time of data collection, but only the ones matching the desired proficiency level were included in subsequent analysis. Those students who could not attend class on the day of the data collection were contacted by the researcher, and a separate arrangement was made with them. In those cases the recording took place outside regular class time, at a time convenient to the students.

In each classroom, prior to starting data collection, the context and general instructions to all tasks were read out aloud by the researcher in Spanish (appendix 7, p. 355). The participants learned that they would participate in a training session for hotel receptionists, and that they would perform one or several tasks, in which they would have to solve a problem at a hotel reception. They were told that prior to performing the task they would have one minute to get familiar with the instructions.

[23] Despite the data being collected during class time, it is a laboratory study given that the administered tasks were not part of the curriculum.

After the general instructions were read out, the researcher indicated whose turn it was to perform the task.

In the classrooms where the data collection took place, the participant was sitting facing the researcher, at the teacher's desk. First the participant received the input for a task (i.e., instructions, client profile, and hotel/room characteristics). Then the participant was informed whether they would perform one or three tasks. The participants had one minute to familiarize themselves with the task input, which was sufficient to get a general idea of what the task was about, and scan the instructions rather than read for details. When one minute passed, the researcher asked the participant if they had any questions. The researcher then turned on the recorder, and the participant was asked to start performing the task immediately. There was no interlocutor, and the participant was instructed that the researcher would not interfere when the task was being performed. There was no time limit to performing the task, and the participant normally signalled that they finished the task by saying "that's it".

Immediately after the task was completed, the participant was given the affective variable questionnaire. In order to avoid misinterpretations, the scale points were explained (1=easy task, 9=difficult task), and the participant marked their answers by putting a cross in the pertinent box. Filling in the post-task questionnaire took no longer than one minute per task. The affective variables questionnaire was the same as the one used in the first pilot study (Appendix 2, p. 342).

The procedure described above (i.e., performing the task and filling in the questionnaire) was repeated three times with those participants who engaged in performing a sequence of three tasks. Performing a single task took on average four and a half minutes in the case of the simple task, and the complex and +complex task each took on average seven minutes to perform. Filling in the questionnaire typically took no longer than one minute. Therefore the whole procedure took on average 4-5 minutes in the case of those participants who performed only one task (so, individual task performance group), and approximately nineteen minutes in the case of those participants who performed a sequence of three tasks. The overall pool of data consisted of 237 sound files (over twenty-one hours of oral speech). The data were recorded with several digital recorders Marantz PMD620.

6.8.3 Independent and dependent variables

The independent variables in the current study were task complexity, task sequencing, and learner proficiency, and dependent variables were fluency, accuracy, and lexical and structural complexity. The CAF triad was chosen for three reasons. First, it is a widely accepted view that aspects of L2 production are multi-faceted in nature, and that they can be represented by means of fluency, accuracy, and complexity (Ellis & Barkhuizen, 2005; Skehan, 1998; Housen & Kuiken, 2009). Second, the choice of CAF over other dependent measures also ensured the comparability of findings in the current study with previous research, such as the studies analyzed in the literature review chapter. Third, this study did not target any language feature in particular (in terms of lexis, syntax, or pragmatics), so analyzing speech in a three-dimensional way (i.e., by means of CAF) was considered an optimal approach.

6.8.4 Measures

In those studies, which measured CAF, capturing the different linguistic phenomena was carried out using a variety of measurements. As was mentioned in the literature review chapter, several shortcomings related with the measures encountered in some of the TBLT studies have been: 1) the number of measures used (capturing only one aspect of a dimension, such as speech rate in the case of fluency); 2) the scope of measures used (general vs. specific), and 3) adequacy of the measured used to the analyzed learner population.

In order to overcome some of these shortcomings, and in an attempt to reflect the multi-componential nature of speech production, the current study employed both a wide variety of general measures, and one specific measure. General performance measures were employed as, following Skehan and Foster (1999), “these have been argued to be the most sensitive to differences between groups of learners in experimental studies” (p. 229). An equally important reason for using general measures was to ensure the comparability of findings between the current study and previous (and future) research, and their well-established nature and validity due to their extensive use in CAF research. In response to recent calls for specific production measures (Norris & Ortega, 2009; Robinson & N. Ellis, 2008; Robinson, Cadierno, & Shirai, 2009), this study also employed a specific measure, as these were claimed in literature to be better able to capture differences in performance.

In this study, the different components of the CAF triad were defined following Skehan and Foster (1999), p. 96-97:

Fluency: the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems.

Accuracy: the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error.

Complexity/range: the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system.

6.8.4.1 Fluency measures

Skehan (1998) distinguished three aspects of fluency: speed fluency, which is concerned with the rate of speech delivery; repair fluency, which is marked by a range of hesitation phenomena, and breakdown fluency, which measures pausing behavior. These three building blocks of fluency reflect Segalowitz's (2010) idea of fluency being a multidimensional construct. This study focused on the first two of these phenomena, speech rate and repair fluency, and it examined them by means of three measures: unpruned and pruned speech rate (as a measure of speed and density of delivery) and dysfluency ratio (as a measure of repair behavior).

6.8.4.1.1 Unpruned speech rate (rate A)

This measure was calculated by dividing the total number of syllables by the total time, and multiplying it by 100. In this measure, the original text was taken into account, keeping the file intact in terms of repetitions, false starts, and self-repairs.

6.8.4.1.2 Pruned speech rate (rate B)

This measure was calculated by dividing the total number of syllables in a narrative by the total time, and multiplying it by 100. In this measure, the texts were pruned: prior to the calculation, they were previously cleaned by deleting from them repetitions, restarts, and self-repairs. Each of these categories is defined below, when discussing the Dysfluency ratio measure.

The texts were divided into syllables manually, and both for Rate A and Rate B, a CLAN formula was employed to perform the calculation of the total number of syllables. Both unpruned and pruned speech rate have been used extensively in those studies which measured CAF, for example Robinson (2007), Kormos and Trebits (2012), and Michel et al. (2007).

6.8.4.1.3 Dysfluency ratio

In order to measure dysfluency, the narratives were coded for the following categories of dysfluency markers: repetitions, false starts, and lexical and morphosyntactic repairs. Dysfluency was calculated by dividing the total number of these phenomena by the total time, and multiplying it by 100. What follows is the conceptualization of these phenomena following Foster et al. (2000).

<u>Repetition</u>

Following Foster et al. (2000), “A repetition is where the speaker repeats previously produced speech (…) This is a device which may be used to hold the floor, or to allow time for planning on line” (p. 368). In the context of the current study, “previously produced speech” encompassed such instances of production as repeating, twice or multiple times, the same syllable, word, string of words, or sentence.

Example of a repetition (REP)

I have some [/] REP some bad news for you
(*repetition of a single word*)

<you have> [/] REP you have views to the swimming pool
(*repetition of a string of words*)

this hotel here stay <in the cent@i> [/] REP in the center of the city
(repetition of a string of words with one unfinished word)

When repetition was a means to achieve the effect of emphasis, such repetition was not deleted, as for example in:

it is a very very beautiful hotel.

False start

False starts were defined as "an utterance which is begun and then abandoned" (Foster et al., 2000, p. 368). False starts were instances of abandoning a started utterance, and starting a new utterance, or reformulating a message. False starts, although most typically identified at the beginning of an utterance, were also encountered in the middle of an utterance.

Example of a false start (FST)

<it have see to the> FST [///] you can see the sea <with the> [///] FST and you have a balcony to see (…)
(false start detected at the beginning of an utterance)

you can search in this point of the city a lot of shows a lot of restaurants a lot of [///] FST I [/] REP I believe that this hotel is hmm@p [/] REP is perfect for [/] REP for you
(false start detected in the middle of an utterance)

Repair

Following Foster et al. (2000), "repairs" were identified to occur "when the speaker identifies an error either during or immediately following production and stops and reformulates the speech"[24] (p. 368). They were conceptualized as instances of reviewing one's speech, as demonstrated by a participant's substitution of a wrongly used lexis or grammar structure for a correct one. Lexical and morphosyntactic repairs were distinguished. The former had to do with reviewing word choice understood as a single word or two- or multiple-word structures, such as phrasal verbs or collocations. The latter had to do with reviewing syntactic structures, including a range of aspects such as, for example, verbal tense or verb morphology. Initially the category "other repair" was also contemplated for these dysfluency phenomena which potentially escaped the scope of definitions provided above. However, after analyzing a small sample of data it was considered unnecessary as the observed instances of dysfluency fit into the categories established above.

Examples of lexical repair (LREP)

> we have also the emperador hotel the localization [//] LREP location <of the> [/] REP of the hotel it's [/] REP it's similar
>
> we can reservate [///] LREP book the [/] REP the rooms that are near

Examples of morphosyntactic repair (MSREP)

> <I sure> [//] MSREP I'm sure that you can be good in this hotel
>
> you booked a [/] a REP suite <in the last> [/] MSREP on the last floor
>
> this hotel <have a good> MSREP [//] has a good location
>
> basically <the most hmm@p cheap> [//] MSREP the cheapest one
> that we have is for fifteen euros

Dysfluency ratio was not used as extensively in CAF research as speech rate measures, and some examples of studies which employed it were are Hsu (2012) and Ishikawa (2008).

[24] In Foster et al.'s (2000) nomenclature, "repairs" were "self-corrections". They are used as synonyms here.

Although in the current dissertation speech was not analyzed for breakdown fluency, it was argued in literature that analyzing speech rate (so, rate A and rate B) “includes both the amount of speech and the length of pauses, since it takes into account the number of syllables and the total number of seconds in the narrative (Griffiths, 1991 in Gilabert, 2005, p. 210).

6.8.4.2 Lexical complexity measures

Both in empirical TBLT research and in current theorizing about CAF measures (Housen, Kuiken, & Vedder, 2012), there seems to be a consensus that at the observational level of behavioral constructs, the lexical complexity of speech can be broken down into three aspects: density, diversity, and sophistication. In the current dissertation, lexical complexity was operationalized as lexical diversity, which, following Jarvis (2013), is “usually operationalized into measures designed to capture the proportion of words in a language sample that are not repetitions of words already encountered” (p. 88). One of the problems with lexical complexity measures frequently mentioned in literature is that, to some extent, most of them display dependence on text length: the longer the text, the lower the value (Laufer & Nation, 1995; Read 2000; Vermeer, 2000; Jarvis, 2002). In response to this challenge, two measures have been used which are claimed to have overcome the problem of text length: Guiraud’s Index and *D*.

6.8.4.2.1 Guiraud’s Index

Guiraud’s Index (Guiraud, 1954) has been widely used in previous TBLT investigations measuring lexical complexity (e.g., Michel et al., 2007; Michel, 2011;

Levkina & Gilabert, 2012; Malicka & Levkina, 2012), and employing this measure was intended to ensure comparability of the current study with previous research. It was calculated by dividing the total number of types by the square root of the total number of tokens. Although it ostensibly corrects for text length, given the presence of the square root in the formula, as indicated by Jarvis and Daller (2013), "regarding TTR [type-token ratio] and all measures that are calculated solely from type and token frequencies (e.g., Guiraud's Index, Herdan's Index), the problem is that they reduce an entire text to just two categories of words: (a) those that are novel and (b) those that recur" (p. 20). Therefore, this measure perhaps provides only partial insights into the nature of qualitative lexical change from one task to another.

6.8.4.2.2 D

The other lexical complexity measure employed, the *D* index, was based on the work of Malvern and Richards (1997, 2002), Malvern et al. (2004), and Duran et al. (2004). The fundamental idea behind this measure, which distinguishes it from other lexical diversity measures, is that a text represented by this index is a set of type-token ratio values presented as a curve, rather than a single TTR value. In other words, it calculates type-token ratio for different samples of words from a text instead of from the entire text. Following Jarvis (2013), "Because the D index represents the shape and position of the entire TTR curve, this index is assumed not to be affected by text length" (p. 61). In this study, the D_tools software (Meara & Miralpeix, 2004) was used in order to carry out the calculation. Both lexical complexity indices – Guiraud's Index and D - were calculated from pruned texts. The D index was previously used for example in Tavakoli (2009) and Révész (2011).

One shortcoming pertinent to both these measures is that they are text-internal, that is, they calculate lexical complexity solely on the basis of the involved text files, without making reference to existing objective language corpora (as is the case with lexical frequency profiles or Lambda, which measure lexical sophistication), and consequently, they are not sensitive to word difficulty.

6.8.4.3 Structural complexity measures

Given the multifaceted nature of this dimension, three different constructs have been measured: overall complexity, structural complexity at subordination level, and structural complexity at phrasal complexity. Such a thorough approach to measuring syntactic complexity phenomena, although far from common practice in a lot of task-based research, was argued by Norris and Ortega (2009) as reflective of a variety of ways in which speech may get complexified. At the same time, investigating only one of the aforementioned constructs is not only incomprehensive, but may lead to possible misinterpretations. In what follows, the measures used for each of the three sub-constructs are defined.

6.8.4.3.1 Words per AS-unit

Mean length of AS-unit is a measure of overall syntactic complexity, and in this sense it is similar to other length-based measures, such as mean length of t-unit or c-unit. The selected syntactic unit of segmentation, "AS-unit", was chosen due to its suitability for spoken discourse. The definition of AS-unit which guided the subsequent coding process was adapted from Foster et al. (2000): "An AS-unit is a single speaker's utterance consisting of *an independent clause*, *or sub-clausal unit*, together with any *subordinate clause(s)* associated with either (…) *An independent clause* will be minimally a clause including a finite verb" (p. 365, italics in the original). Length-based measures are common in TBLT research, and in particular Words/AS-unit was used for example in Tavakoli and Foster (2008) and Tavakoli and Skehan (2005). An example of coding of speech into AS-units is provided at the end of the next section.

In this dissertation, AS-units were identified from pruned files, that is, in the calculation only the final version of utterance was taken into account. This decision had to do with the fact that low proficiency participants were part of the study. In these participants' speech, a substantial amount of repetition and false start phenomena were identified, frequently including repetitions of entire clauses or sentences, and the inclusion of repeated structures was considered as potentially swaying the results.

6.8.4.3.2 Clauses per AS-unit

Syntactic complexity, understood as changes in subordination ratio, was measured by means of Clauses per AS-unit. It is similar to other subordination indices, such as S-nodes/T-unit or clauses/t-unit, and all these have been extensively used in previous research (e.g., Wigglesworth & Elder, 2010, Albert & Kormos, 2004, and Michel et al., 2007). In the current study, the definition of "clause" was adopted from Foster et al (2000), and it was defined as consisting "minimally of a finite or non-finite Verb element plus at least one other clause element (Subject, Object, Complement, or Adverbial)" (p. 365, capital letters in the original). There were two reasons for choosing subordination ratio as a measure possibly illuminating changes in performance. First, it was chosen considering the participants' proficiency level: it was considered that the participants were all advanced enough to produce subordinate clauses, and for instance coordination ratio would not have been a good predictor of qualitative changes. The second motivation was related with the task design feature of reasoning demands. Taking into account the presence of such reasoning demands as justifying, apologizing, and recommending, the emergence of subordinate clauses was considered relevant to the mental operations that the tasks required from the participants. Particularly the action of justifying was considered to naturally elicit subordination markers such as “because”. This is consistent with the operationalization and adoption of subordination in other pieces of research: for instance according to Matthiessen and Thompson (1988), in English subordinate clauses typically indicate reason or cause, among others.

<u>Example of division of oral speech into Clauses and AS-units</u>

> other option is the junior suite which is twenty euros less for night and is in the second floor (1 AS-unit, 3 clauses)
>
> and other option is the luxury suite in the last floor too but it's twenty euros more expensive (1 AS-unit, 2 clauses)
>
> if you are agree with one of these bedrooms please call me and we can talk about it (1 AS-unit, 3 clauses)

6.8.4.3.3 Words per clause

In order to capture potential differences in performance at phrasal level, the measure Words per clause was used. Following Norris and Ortega (2009), this measure is unique in its contribution to the picture of structural complexity, and distinct from other measures based on length insofar as it “taps a more narrowly defined source of complexification" (p. 561). Other studies in which this measure was used include Ishikawa (2007) and Storch and Wigglesworth (2007).[25]

6.8.4.4 Accuracy measures

Following Pallotti (2009), accuracy refers to the "degree of conformity to certain norms" (p. 592). In calculating accuracy, errors were not classified according to categories of gravity, that is, they were counted on an equally stringent basis. Three categories of errors were calculated: lexical, morphosyntactic, and other errors. The first category concerned those errors which had to do with the word choice, the second category concerned those which had to do with syntax, and the third category encompassed all errors which did not match any of the other two categories, for example pragmatic errors. Pronunciation or intonation errors were not coded, given the difficulty in determining what counts as "error". All narratives were coded for errors by a trained native speaker.

<u>Example of coding for different types of errors</u>
("morphosyntactic"=MSR; "lexical"=LEXR; "other"=OTHER)

> it have MSR nice views
>
> the breakfast is not including MSR
>
> I recommend you MSR the guest room because MSR is not so expensive
>
> you have firstly OTHER one hotel in the center of the city
>
> we can't handle LEXR the room to you
>
> we did LEXR a little mistake and we don't have any more rooms
>
> we cannot keep you here OTHER so my recommendation would be to choose the catalonia hotel instead

---

[25] However, as pointed out by Norris and Ortega (2009), in these studies the measure *words per clause* was used to measure fluency, and not structural complexity.

> there's a meeting room available at every moment OTHER five days a week
>
> it's located in the center of the city and have LEXR the same characteristics as the one you book MSR

Following the coding, errors were calculated by means of two global accuracy measures (Errors per AS-unit and Errors per 100 words) and one specific measure (target-like use of prepositions).

6.8.4.4.1 Errors per AS-unit

Errors/AS-unit were calculated by dividing the total number of errors by total number of AS units in a narrative. This measure was previously used in such previous research as Gilabert (2007c), Shiau and Adams (2011), and Révész (2011).

6.8.4.4.2 Errors per 100 words

Errors/100 words were calculated by dividing the total number or errors by the total number of words, and multiplying it by 100. Other studies which employed this measure include Mehnert (1998), Fortkamp (1999), Sangarun (2005), and Guará-Tavares (2009).

6.8.4.4.3 Target-like use of prepositions

In order to calculate target-like use of prepositions, all correctly and incorrectly supplied prepositions were identified, and their obligatory contexts. Target-like use of prepositions was expressed as a percentage, and was calculated by dividing the number of correctly supplied prepositions by the sum of the number of obligatory contexts and the number of incorrectly supplied prepositions, and multiplying it by

100. This measure was selected in accordance with the following criteria: task-essentialness, participants' L1 and the target language of the study, and participants' current interlanguage state.

Example of coding for prepositions

Incorrect preposition ("WPREP") suppliance

it's in WPREP the third floor and you have a view for WPREP landscapes

the machine that you use for WPREP dry your hair

if there was any problem call to WPREP the hotel

sorry for WPREP the problem

the prices for this room is one hundred forty euros for WPREP night

Preposition omission ("OPREP") (lack of suppliance in obligatory contexts)

is not possible OPREP offer this room

I'm sorry but we need OPREP change you the room

if you want OPREP solve anything else you can call us

Correct preposition ("CPREP") suppliance

you reserved three suites on CPREP the last floor of CPREP the building

if you want we can move you to CPREP a bigger bedroom

I have one problem with CPREP the check-in

it's quite close to CPREP the center of the city

there are a few activities to CPREP do

To my knowledge, measures including prepositions have not been extensively used. Several studies set out to measure spatial prepositions (e.g., Malicka & Levkina, 2012), and a handful of studies in TBLT set out to investigate a conceptually similar measure, target-like use of articles (Gilabert, 2007c; Ahmadian, 2012; Meraji, 2011).

6.9 Data transcription and coding

All data were transcribed in the CLAN mode of CHILDES (MacWhinney 2000). In cases of not clearly audible material, the researcher used her best guess, except for cases of objectively unintelligible and therefore untranscribable data, which was excluded from the transcripts. Such instances were extremely uncommon, however.

6.10 Data screening

6.10.1 Techniques

All data were first subjected to a thorough analysis of the normality of their distributions and other characteristics associated with assumptions for statistical tests. Several statistical and graphical methods were applied in this process, including the values obtained in normality tests, skewness and kurtosis values, and a visual inspection of the data as represented by histograms, Q-Q plots, and boxplots. The data were also checked for outliers. Regarding statistical tests of normality, the Shapiro-Wilk ($\alpha<0.05$) test was selected, as this test is considered stricter than Kolmogorov-Smirnov in samples under 50 participants.

6.10.2 Data sets subject to assessing normality

When assessing normality, a strict approach was employed in the sense that, rather than subjecting the entire data set to a single analysis, the data screening procedure was performed for the different data sets separately. This resulted in a total of five data sets subjected to analysis, and these were: simple-complex sequencing condition (N=30), randomized sequencing condition (N=30), low proficiency speakers (N=30), high proficiency speakers (N=30), and baseline data (N=57). Additionally, given the fact that the analysis of the data obtained for the current dissertation was carried out using multiple CAF measures, the values were analyzed by dimension (i.e., fluency, accuracy, structural complexity, and lexical complexity).

6.10.3 Results of data screening: sequencing conditions and proficiency levels

Table 13 reports the skewness and kurtosis values for all groups except baseline data, which will be treated separately in the next section. In order to assess normality on the basis of skewness and kurtosis values, the approach taken was to divide the values obtained in the table by the standard error of skewness or kurtosis, respectively. If the result was larger than 3.3, the data were assessed to be not normally distributed. As a result of this calculation, the data were assessed as slightly to heavily skewed. Regarding the different speech production dimensions which were subjected to analysis, the data obtained for fluency were overall normally distributed except for several data points, with task 2 showing a pattern of somewhat more bias towards non-normal distribution compared to the two other tasks.

Also, the data for rate fluency exhibited a trend towards more normal distribution than those for repair fluency. Regarding accuracy, it can be observed that the data obtained for Errors/AS-unit and Errors/100 words showed a trend towards a very abnormal distribution, with target-like use of prepositions showing less of this trend. The data obtained for structural complexity showed a strong deviation from normal distribution, with task 3 in the case of two variables (Words/AS-units and Clauses/AS-units) showing a less extreme tendency. Finally, the data obtained for lexical complexity showed a fairly homogenous pattern of normal distribution.

Taking into consideration the skewness and kurtosis values, and the remainder of the above-mentioned techniques, it was concluded that the data obtained for fluency and lexical complexity were normally distributed, while the data obtained for accuracy and structural complexity were assessed as not normally distributed. In order to select the optimal approach to data analysis, several preliminary analyses were conducted using both parametric and nonparametric tests. After several such inspections, two decisions were taken regarding analyzing the data obtained for sequencing conditions and proficiency levels: (1) to analyze normally distributed data (fluency and lexical complexity) using parametric statistical tests, and (2) to analyze not normally distributed data (accuracy and structural complexity) using nonparametric statistical tests.

Table 13. Skewness and kurtosis values: sequencing conditions and proficiency levels

		Sequencing condition				Proficiency			
		Simple-complex		Randomized		Low	High		
	Task	*S*	*K*	*S*	*K*	*S*	*K*	*S*	*K*
Rate A	1	-.376	-.734	.046	.180	.453	-.230	.514	.560
	2	2.957	12.331	-0.59	-.399	.620	.239	2.842	11.953
	3	.218	-.848	.217	1.192	.862	2.013	.063	-.448
Rate B	1	-.266	-.642	.121	-.292	.539	-.304	.478	.195
	2	2.878	11.507	.095	-.629	.757	.344	2.484	9.375
	3	.440	-.222	.602	1.365	1.450	3.578	.237	.495
Dysfl.	1	1.254	2.820	-.027	-.867	.008	-.727	1.178	2.642
	2	2.150	6.279	.810	.060	.810	-.173	2.207	6.686
	3	.607	.377	.586	-.197	.327	-.798	.911	1.173
Err/AS	1	1.462	2.623	.611	-.394	.841	.646	2.607	9.361
	2	3.679	17.013	1.393	2.245	2.284	6.553	.648	.010
	3	2.371	7.969	1.469	2.475	1.821	4.211	-.505	1.519
Err/100w	1	1.789	4.567	.854	.171	1.473	2.762	2.063	7.338
	2	1.578	3.363	1.171	1.091	.971	.603	.687	1.780
	3	1.813	5.092	1.819	3.518	1.482	2.146	-.429	1.410
TLU_prep	1	-1.034	.400	-.747	-.481	-.555	-1.053	-1.027	.686
	2	-1.285	2.440	-.737	.386	-.634	.427	-.158	-.348
	3	-.818	.525	.183	-.864	-.703	.638	.112	-.240
Words/AS	1	.091	.172	2.523	10.648	-.235	-.291	2.480	9.775
	2	1.932	4.881	-.075	1.144	1.254	2.866	1.572	5.526
	3	.761	.295	.094	-.060	.333	.561	.376	-.560
Words/Cl	1	-1.499	4.651	-.354	.532	-.297	1.982	-2.397	8.820
	2	2.176	7.771	.949	.676	2.256	7.784	.842	1.259
	3	1.711	2.495	1.579	2.649	1.076	.783	2.063	3.669

Table 13 continued.

Clause/ASU	1	4.854	24.837	1.924	5.079	2.198	5.776	4.765	24.156
	2	1.568	3.430	-.082	-.223	.849	.852	1.212	2.993
	3	-.672	-.007	-.214	.010	.084	-.070	-1.006	.628
Guiraud's Index	1	.609	.446	.402	-.138	.453	-.023	.783	.014
	2	.373	-.003	-.339	.972	-0.87	-.292	.456	.680
	3	.040	-.825	-.334	.111	.166	-.336	-.166	-.019
D	1	1.045	1.333	.493	.293	.562	-.098	.908	.904
	2	.548	-.460	.171	-.405	.660	-.497	.537	-.301
	3	.642	.595	-.645	.301	.187	-1.095	.842	1.046

6.10.4 Results: baseline data

Table 14 reports the results obtained in the data screening process for individual task performance. The data for accuracy and structural complexity exhibited an overall strong deviation from normality (except for a few measures), with the majority of values obtained for skewness and kurtosis being positive. Regarding lexical complexity, the data showed the same pattern on both of the employed variables (Guiraud's Index and D) for the most complex task. As far as fluency data are concerned, those showed a quite normal distribution compared to the other dimensions.

Fluency being the only dimension showing considerably less deviation from normality than the other dimensions across tasks, a decision was taken to analyze this entire data set (i.e., baseline data) using nonparametric statistical tests.

In the process of screening the data in all data sets (i.e., sequencing conditions, proficiency levels, and baseline data), outliers were detected and these were kept in the data for the final analysis. The rationale behind this decision was three-fold. First, the detected outliers were legitimate ones, that is, they were not a result of an error in the data, but belonged to the overall sample. Second, a large portion of the data was analyzed using nonparametric tests, which do not assume a normal data distribution (Burke, 2001; Orr et al., 1991). Therefore, the presence of an outlier should not cause a bias in the results obtained. Third, in these data points where outliers were detected, preliminary analyses were carried out, both including and excluding outliers. Following these inspections, they were assessed as not influencing the results in any significant way.

Table 14. Skewness and kurtosis values: baseline data

		Baseline data	
	TASK	***Skewness***	***Kurtosis***
Rate A	1	.560	-.311
	2	-.638	1.113
	3	.644	-.466
Rate B	1	.645	-.240
	2	-.468	.725
	3	.881	.007
Dysfluency	1	.403	-1.453
	2	.663	.240
	3	.774	-.211
Errors/ASU	1	2.901	10.015
	2	1.138	2.030
	3	1.254	1.458
Errors/100w	1	1.421	1.846
	2	1.246	1.659
	3	.989	.256
TLU_prep	1	-.501	-.497
	2	-.732	.623
	3	-1.415	2.065
Words/ASU	1	1.315	1.862
	2	.458	1.193
	3	.298	-.619
Words/Clause	1	3.075	11.112
	2	3.724	15.089
	3	.249	.346
Clause/ASU	1	.788	.086
	2	-1.427	.524
	3	.304	-.583
Guiraud's Index	1	-.132	-0.81
	2	-.021	.153
	3	1.162	4.153
D	1	-.495	.443
	2	-.183	.034
	3	1.031	1,495

6.11 Statistical tests

Given that some of the data were normally distributed and others were not, the use of both parametric and nonparametric statistics was warranted. The following statistical analyses were used in the study: 1) descriptive statistics to obtain information about means and standard deviations, 2) one-way repeated measures ANOVA in the case of normally distributed data in sequencing conditions (fluency and lexical complexity), 3) Wilcoxon signed-ranks test in the case of not normally distributed data in sequencing conditions (structural complexity and accuracy), and 4) Mann-Whitney U-test in the case of individual task performance (all speech production dimensions). Two types of effect sizes were reported. For those data analyzed by means of ANOVA, partial eta squared was reported. For all the other statistical tests, the reported effect size was Cohen's *d* (Cohen, 1988). The latter effect size was calculated on the basis of means and standard deviations.

6.12 Adjustment of the p value

As is the case in the field of Second Language Acquisition, the overall alpha level in this study was set at $p<.05$ (Larson-Hall, 2010). However, the use of both parametric and nonparametric tests, as well as multiple dependent variables (a total of eleven), called for the adjustment of the alpha level. Following Tabachnick and Fidell (2013) "the more numerous the DVs[26] (…), the greater the potential distortion of alpha levels" (p. 254). The more dependent variables involved, the greater the likelihood that one or several of them will yield a statistically significant result. Given that data analysis was performed using both parametric and nonparametric tests, a

[26] dependent variables

Bonferroni adjustment in the case of ANOVAs would be only a partial solution to the problem, because the SPSS package does not offer an analogous built-in solution in the case of nonparametric tests. Therefore, both in the case of parametric and nonparametric tests, a decision was taken to divide the overall alpha level (.05) by the total number of multiple comparisons. Multiple comparisons were operationalized as comparisons between the different tasks: the simple task and the complex task, the complex task and the most complex task, and the simple task and the most complex task. There were therefore three multiple comparisons involved in the study, so overall alpha (.05) divided by 3=.016, which is the level at which the experiment-wise *p* value was set. Throughout the Results chapter, the results matching the standard *p* value (.05) were signaled with one asterisk, and the results matching the experiment-wise *p* value (.016) were signaled with two asterisks.

6.13 Summary of the chapter

This chapter explained the methodological decisions taken in the experiment reported in this dissertation. First, the context of the study was outlined, with a special focus on the participants and the measurement of their proficiency level. Following that, the research tasks used in this study were presented, which involved issues such as the derivation of tasks and their development, the operationalization of cognitive complexity, and the pilot studies carried out to validate the hypothesized differences in cognitive complexity. Study design was then described, and details were provided regarding the three experimental conditions under investigation. Finally, the employed statistical test used, coding and measures were presented. The next chapter presents the results obtained in the current experiment.

CHAPTER 7
RESULTS

7.1 Introduction

The aim of this chapter is to present the results obtained in the current dissertation. The chapter is organized in the following way. First, each research question is answered one by one, presenting first the descriptive statistics, followed by the results obtained in the inferential tests. In each of the research questions, the results for each dimension of production are answered separately, and throughout the chapter the different dimensions are presented in the following order: fluency, lexical complexity, accuracy, and structural complexity. For each of the research questions, a general table with descriptive statistics (means and standard deviations) is first reported. Subsequently, each of the dimensions is discussed in terms of the findings obtained in descriptive statistics, followed by the results of inferential tests. After the results for each research question comes a brief summary of the results for that particular question. To facilitate reading, the summary of results for each research question is illustrated with line graphs, but for the sake of brevity, only selected findings are presented in this way. The chapter is concluded with a summary of all results.

7.2 Organization of the research questions

The order of answering the research questions in the current dissertation goes from general to specific. That is, first the influence of increasing cognitive complexity on performance is dealt with, then the effect of sequencing on performance, and finally the effect of proficiency on performance. Although the main objective of this work was to contribute knowledge about task-based performance as mediated by sequencing, a decision was taken to first deal with the more general issue of how task complexity affected performance on the three tasks employed in the study, without taking into consideration the effects of sequencing or proficiency.

7.3 Abbreviations

For the sake of brevity, throughout this chapter, the tasks employed in this dissertation are referred to as task 1 (="simple task"), task 2 (="complex task"), and task 3 (= "the most complex task"). Where it was considered adequate, "simple-complex sequencing condition" was abbreviated to "SC", and "randomized sequencing condition" was abbreviated to "RAN". The terms "individual (isolated) task performance" and "baseline data" were used interchangeably throughout the chapter.

7.4 Results of the affective variables questionnaire: difficulty perception

Prior to presenting the results for each research question, the results of the Affective Variables Questionnaire administered as part of the main study are presented. The answers were provided on a 9-point Likert scale, the scale points being: 1=easy and 9=difficult. This section reports only the results obtained for the Difficulty item of the questionnaire. The results are displayed in table 15.

Table 15. Difficulty perception: sequencing condition, proficiency level, and individual task performance

Condition		Task complexity level					
		Simple		***Complex***		***+Complex***	
		M	***SD***	***M***	***SD***	***M***	***SD***
Sequencing	*Simple-complex*	4.90	1.66	6.13	1.52	6.83	1.31
	Randomized	4.13	1.63	5.03	1.47	5.86	1.56
Proficiency	*Low*	4.00	1.33	5.13	1.38	5.86	1.40
	High	5.03	1.84	6.03	1.67	6.83	1.48
Individual task performance		4.78	1.80	4.83	1.58	6.15	2.00

As can be observed, in all the conditions subject to analysis (i.e., sequencing conditions, proficiency levels, and individual task performance), the trend in the data was for *the simple task to be perceived as the least difficult one*, and for *the most complex task to be perceived as the most difficult one*. Regarding the magnitude of difference between the different tasks, the greatest difference can be observed in the simple-complex sequencing condition: the difference between the means in the simple and the most complex task equaled 1.93. By comparison, the participants in the individual task performance condition did not perceive the tasks to be as different in cognitive complexity level: in this group, the difference between the simple and the most complex task was 1.37. On the other hand, the lowest mean obtained was 4 (simple task in low proficiency participants), and the highest mean obtained was 6.83 (the most complex task in the simple-complex sequencing condition and in high proficiency participants).

Beyond a purely quantitative inspection of the data, what the findings show is that overall no extreme values were detected on either end of the scale, for example 2 or 3 in the case of the simple task versus 8 or 9 in the case of the most complex task. The fact that the subjective ratings concentrated roughly on the center of the scale suggests that these tasks did not represent extremes of complexity (very simple versus very complex tasks), but rather a continuum of cognitive complexity levels: the differences were salient enough for the participants to discern them, but at the same time they did not represent extreme opposites.

Regarding the perception of difficulty showed by the different conditions, a closer inspection reveals two interesting patterns: 1) the participants in the simple-complex condition generally assessed the tasks as *more difficult than the randomized sequencing condition*, and 2) high proficiency participants assessed the tasks as *more difficult than the low proficiency participan*ts. It is also noteworthy that in the individual task performance condition, the difference between the simple and the complex task was rather marginal (M=4.78 and M=4.83, respectively), unlike in the case of the speakers belonging to different sequencing conditions or proficiency levels: both these groups perceived the complex task to be at least one scale point more difficult than its simple counterpart.

The findings obtained from the affective variables questionnaire are consistent with the predictions of the Cognition Hypothesis, which postulates that tasks designed to be complex should be perceived as difficult. More importantly, however, the designed cognitive complexity differences hypothesized by the researcher were matched by the participants' subjective ratings. These results further reiterate the findings obtained in the second pilot study carried out prior to the main experiment.

7.5 The effect of cognitive complexity on performance

7.5.1 The effect of cognitive task complexity on performance in the sequencing conditions (simple-complex vs. randomized)

The first research question enquired about the overall impact of increasing cognitive task complexity in the sequencing conditions, and was formulated in the following way:

Research question 1a. How do tasks of different cognitive complexity levels, as manipulated along the variables ±reasoning demands and ±number of elements, affect task performance in the condition of sequences, as measured by fluency, accuracy, and complexity?

Descriptive statistics for this research question are displayed in table 16.

Table 16. Descriptive statistics of tasks: means and standard deviations (sequencing conditions)

Dependent variable		**Task 1**		**Task 2**		**Task 3**	
		M	***SD***	***M***	***SD***	***M***	***SD***
Fluency	Rate A	122.44	26.42	125.69	28.99	120.65	28.00
	Rate B	113.84	25.82	114.55	28.02	110.22	27.30
	Dysfluency	3.94	1.78	7.86	3.97	4.27	1.83
Accuracy	Errors / ASU	1.08	0.51	1.07	0.66	0.97	0.56
	Errors / 100w	11.53	5.90	11.36	6.00	10.48	6.28
	TLU prep	65.04	18.10	66.64	15.88	76.02	11.06
Structural complexity	Words /ASU	9.28	1.43	9.18	1.75	9.20	1.31
	Words/ Clause	7.37	1.13	6.78	1.24	7.38	1.58
	Clause/ ASU	1.31	0.49	1.37	0.22	1.28	0.23
Lexical complexity	Guiraud's Index	5.40	0.74	5.18	0.71	5.28	0.64
	D	44.87	12.31	48.95	12.45	54.37	12.47

7.5.1.1 Fluency

As can be observed in table 16, the participants' performance in terms of speed fluency, as measured by Rate A and Rate B, was quite stable across the three tasks. Both in Rate A and Rate B, task 2 was the one which led to the most fluent behavior, triggering the score of *M*=125.69 in Rate A and *M*=114.55 in Rate B. In both measures the most notable difference can be observed between tasks 2 and 3, with a difference of 5.04 syllables per second in Rate A, and a difference of 4.33 syllables per second in Rate B.

Task 3 turned out to trigger the lowest speech rate on both measures (*M*=120.65 in Rate A, and *M*=110.22 in Rate B). However, this result stands in contrast to the picture displayed by Dysfluency ratio. Whereas tasks 1 and 3 displayed a similar pattern (*M*=3.94 and *M*=4.27, respectively), a considerable drop in fluency can be observed on task 2 (*M*=7.86). If we compare the performance on task 1 and 2 on Dysfluency ratio, it is noteworthy that the participants displayed almost twice as many hesitations on task 2 than they did on task 1. The picture obtained from the three fluency measures leads to the conclusion that the faster the rate at which the participants delivered the task, the more hesitations occurred. Speaking faster therefore might have taken place at the expense of significantly more pausing.

As shown in table 17, speed fluency, as measured by Rate A and Rate B, did not turn out to be significantly affected. However, a statistically significant main effect was found for Dysfluency ratio.

Table 17. Repeated measures ANOVA for the effect of cognitive task complexity on performance: fluency

Fluency	**Sum of squares**	**Wilk's Lambda**	***F***	***Df***	***p***	**Effect size**
Rate A	782.679	.926	2.290	2	.111	.074
Rate B	646.836	.938	1.881	2	.162	.062
Dysfluency	566.393	.333	56.961	2	.000**	.667

* α significant at $p<.05$
** α significant at $p<.016$

Post-hoc analyses of Dysfluency variable (table 18) revealed that statistically significant differences were located between tasks 1 and 2, and between tasks 2 and 3.

Table 18. Pairwise comparisons for fluency (Dysfluency ratio)

Fluency	**Task 1-2**	**Task 2-3**	**Task 1-3**
Dysfluency	.000**	.000**	.270

* α significant at $p<.05$
** α significant at $p<.016$

7.5.1.2 Lexical complexity

An analysis of descriptive statistics (table 16) shows that the measure *D* displayed a pattern of gradual increase in lexical complexity across the three tasks, with the most complex task (task 3) triggering the most lexically complex speech (*M*=44.87 in task 1, *M*=48.95 in task 2, *M*=54.37 in task 3). However, the same did not hold true for Guiraud's Index. In this measure there were only slight differences between the three tasks, the most lexically complex speech being generated by task 1 (*M*=5.40), and the lowest by task 2 (*M*=5.18). Therefore in D an increase in the diversity of lexical forms produced was proportionate to increases in cognitive complexity, but according to the other measure, the generated lexis was only marginally affected by increasing cognitive demands.

These patterns were confirmed by subsequent inferential tests. A one-way repeated measures ANOVA (table 19) detected a statistically significant difference in performance between the three tasks in the case of D, but not in the case of Guiraud's Index.

Table 19. Repeated measures ANOVA for the effect of cognitive task complexity on performance: lexical complexity

Lexical complexity	**Sum of squares**	**Wilk's Lambda**	***F***	***Df***	***p***	**Effect size**
Guiraud's Index	1.487	.910	2.835	2	.067	.090
D	2.730	.513	27.005	2	.000**	.487

* α significant at p<.05
** α significant at p<.016

As identified through post-hoc analyses (table 20), all pairwise comparisons for D turned out to be statistically significant.

Table 20. Pairwise comparisons for lexical complexity (D)

Lexical complexity	**Task 1-2**	**Task 2-3**	**Task 1-3**
D	.017*	.000**	.000**

* α significant at p<.05
** α significant at p<.016

7.5.1.3 Accuracy

The scores obtained for two global measures of accuracy showed a tendency for a slight but steady decrease in the number of non target-like forms across the three tasks (table 16). In the case of Errors/AS-units, task 1 led to an average of *M*=1.08 non-target like forms per AS-unit, while in task 3, *M*=0.97. The same holds true for Errors/100 words; task 1 yielded a score of *M*=11.53, and in the case of task 3 *M*=10.48. This pattern of decrease in deviation from non-target like forms was accompanied by a simultaneous increase in target-like use of prepositions. The results obtained for this measure confirmed the facilitative role of the most complex task, which was the most accurately delivered one (*M*=65.04 in task 1, *M*=66.64 in task 2, and *M*=76.02 in task 3). Table 21 shows the exact location of statistically significant differences.

Table 21. Impact of cognitive complexity on performance (accuracy): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Errors/ 100 words	-.479	.632	.000	-1.675	.094	.08	-2.061	.039*	.09
TLU_prep	-.611	.541	-.09	-5.145	.000*	-.68	-4.807	.000**	-.73

* α significant at p<.05
** α significant at p<.016

In the global measure Errors/100 words, statistically significant differences were found only between the simple and the most complex task. In target-like use of prepositions, the differences turned out to be significant for two pairs of tasks: 2 and 3 and 1 and 3, with p<.016 and large effect sizes[27] in both cases. The participants therefore displayed a tendency in which accuracy increased proportionately to increases in cognitive task complexity. Although the differences in scores obtained in the case of global production measures of accuracy were very modest, they showed a trend towards less deviation from norms as tasks placed increasingly higher demands on the speakers. On the other hand, the differences displayed by the specific measure (target-like use of prepositions) were quite prominent, which was confirmed both by descriptive statistics and inferential tests.

7.5.1.4 Structural complexity

Descriptive statistics showed in general very slight differences in performance regarding the range of syntactic structures produced (table 16). The participants' overall structural complexity of speech, as measured by Words/AS-units, showed virtually identical performance on tasks 1 and 3 (*M*=9.28 and *M*=9.20, respectively).

[27] Cohen's (1988) benchmarks were employed when assessing effect sizes, and therefore: *d*=.10 to .29 or *d*=-.10 to -.29 (small effect size); *d*=.30 to .49 or *d*=-.30 to -.49 (medium effect size); *d*=.50 to 1.0 or *d*=-.50 to -1.0 (large effect size).

The same holds true for the amount of subordination as measured by Clauses/AS-units (in task 1, *M*=1.31 and in task 3, *M*=1.28), with task 2 leading to slightly more subordination (*M*=1.37) than the other two tasks. Regarding clause length, as measured by Words/clause, almost identical values can be observed in tasks 1 and 3 (*M*=7.37 and *M*=7.38, respectively), with task 2 triggering a slightly lower score (*M*=6.78).

As can be observed in pairwise comparisons carried out by means of Wilcoxon signed ranks test (table 22), Words/clause yielded statistically significant differences between tasks 1 and 2, and between tasks 1 and 3, with task 3 leading to the most structurally complex speech (*M*=7.38). In the case of Clauses/AS-unit, differences were detected between tasks 1 and 2, and in this measure it was task 2 which rendered the production the most complex (*M*=1.37).

Table 22. Impact of cognitive complexity on performance (structural complexity): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ Clause	-3.825	.000**	.49	- 1.914	.056	-.42	-2.183	.029*	-.007
Clause/ ASU	-4.126	.000**	-.15	- 1.406	.160	.39	-1.760	.078	.07

* α significant at p<.05
** α significant at p<.016

The picture obtained from the different structural complexity measures is inconsistent. Depending on the measure, it was either the simple, complex, or the most complex task, which led the participants to produce the most structurally complex speech. While significant differences were obtained for a total of three comparisons, these results should be treated with caution given the negligible differences displayed by the means. This holds particularly true for the result obtained for Words/clause (*M*=7.37 for task 1, *M*=7.38 for task 3; *p*=.000**). This state of affairs, and the fact that the differences between the three tasks were generally very small in magnitude, suggests that there was not any perceptible or consistent effect.

7.5.1.5 Summary of results: the effect of cognitive task complexity on performance in the sequencing conditions

Increasing cognitive task complexity had the following effect on performance in the sequencing conditions. The participants' speech rate was quite stable across the three tasks, with the highest values being generated on task 2. On the same task, the speech contained the largest number of hesitations, as displayed by the repair fluency measure. Accuracy showed an overall pattern of gradual improvement on all three measures: the production on the most complex task contained less deviation from norms, with a corresponding increase in the correct use of prepositions. As for lexical complexity, the two measures employed showed different patterns, with D detecting a considerable increase in the diversity of lexical forms, whereas only minor variations were found for Guiraud's Index. Structural complexity could perhaps be best described as non-affected as a result of increasing task complexity, given that the differences obtained across the three tasks were only marginal.

7.5.2 The effect of cognitive task complexity on performance in the baseline data

In what follows, research question 1b, enquiring about the effect of cognitive complexity on performance, will be explored in relation to individual, isolated task performance. Research question 1b was formulated in the following way:
Research question 1b: How do tasks of different cognitive complexity levels, as manipulated along the variables ±reasoning demands and ±number of elements,

affect individual, immediate task performance, as measured by fluency, accuracy, and complexity?

In contrast to the sequencing condition described above, in which the same participant performed the three tasks, in individual task performance, each of the different tasks (task 1, task 2, and task 3) was performed by a set of randomly assigned speakers. Table 23 presents the descriptive statistics of tasks as obtained in individual task performance condition.

Table 23. Descriptive statistics of tasks (individual task performance): means and standard deviations

Dependent variable		Task 1		Task 2		Task 3	
		M	***SD***	***M***	***SD***	***M***	***SD***
Fluency	Rate A	95.22	43.35	113.27	35.01	98.38	34.99
	Rate B	89.46	42.39	102.98	32.88	90.76	34.77
	Dysfluency	2.50	1.21	4.10	2.56	2.99	1.83
Accuracy	Errors / ASU	1.60	1.17	1.14	0.56	1.10	0.57
	Errors / 100w	17.32	10.17	12.94	6.84	12.56	7.05
	TLU prep	51.45	22.84	55.72	13.59	69.31	18.18
Structural complexity	Words /ASU	9.14	1.36	8.98	1.27	9.14	1.37
	Words/ Clause	7.68	1.40	7.17	1.98	6.57	0.58
	Clause/ ASU	1.21	0.18	1.31	0.28	1.40	0.22
Lexical complexity	Guiraud's Index	5.06	0.83	5.22	0.47	5.33	0.74
	D	39.31	10.81	46.43	11.06	49.44	11.87

7.5.2.1 Fluency

As can be observed in table 23 (descriptive statistics), the participants who performed the simple and the most complex task revealed very similar patterns in terms of speed of speech delivery; in Rate B, *M*=89.46 for task 1, and *M*=90.76 for task 3. The highest speech rate was obtained by the participants who did task 2, with *M*=102.98. However, at the same time as delivering their speech faster, these participants were the ones who generated most repair fluency when delivering the task (*M*=4.10 on task 2; on task 1 *M*=2.50, and on task 3 *M*=2.99). The participants in task

2 therefore spoke faster at the expense of more repetitions, false starts, and self-corrections. The only difference which turned out to be statistically significant, as detected by the Mann-Whitney U-test (table 24) was the one between tasks 1 and 2 on Dysfluency ratio.

Table 24. Impact of cognitive complexity on individual task performance: fluency (Mann-Whitney U-test)

	Task 1-2			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Rate A	-1.762	.078	-.45	-1.686	.092	.42	-.439	.661	-.08
Rate B	-1.459	.145	-.35	-1.461	.144	.36	-.161	.872	-.03
Dysfluency	-2.067	.039*	-1.21	-1.545	.122	1.26	-.643	.520	-.31

* α significant at $p<.05$
** α significant at $p<.016$

7.5.2.2 Lexical complexity

In both Guiraud's Index and D, a pattern of gradual increase in lexical complexity across tasks can be observed, as displayed in descriptive statistics (table 23). Although the differences obtained for Guiraud's Index are modest (M=5.06 in task 1 and M=5.33 in task 3), they are very prominent in the case of D: M=39.31 in task 1, and M=49.44 in task 3. Performing the task of the most demanding cognitive level was therefore conducive to the most lexically rich output, and vice versa for the simple task. As detected by Mann Whitney U-test (table 25), a statistically significant difference was yielded for the measure D between tasks 1 and 3, with $p<.016$ and a large effect size.

Table 25. Impact of cognitive complexity on individual task performance: lexical complexity (Mann-Whitney U-test)

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Guiraud's Index	-.517	.605	-.23	-.422	.673	-.23	-.994	.320	-.34
D	-1.854	.064	-.65	-.506	.613	-.27	-2.456	.014**	-.89

* α significant at $p<.05$
** α significant at $p<.016$

7.5.2.3 Accuracy

As descriptive statistics show (table 23), overall, the simple task led to the highest occurrence of non-target like forms, whereas the most complex task triggered the opposite pattern. Regarding global accuracy measures, the differences observed in Errors/AS-units were small in magnitude (*M*=1.60 in task 1 and *M*=1.10 in task 3), but a more marked difference was displayed in the case of Errors/100 words: in task 1, *M*=17.32 and in task 3, *M*=12.56. In the latter measure, error rate was almost identical in the case of tasks 2 and 3 (*M*=12.94 vs. *M*=12.56). The specific measure, target-like use of prepositions, captured the most prominent differences between the tasks: in the most complex task, *M*=69.31, whereas task 1 generated a score of *M*=51.45. It can therefore be concluded that overall the high task demands of the most complex task fostered more target-like output than the other two tasks.

Mann-Whitney U-test (table 26) revealed significant differences on target-like use of prepositions between tasks 2 and 3 and between tasks 1 and 3. While no significant differences were found for the global accuracy measures, the results for the specific measure show a strong effect of complexity on performance, as illustrated by the p value and the effect size.

Table 26. Impact of cognitive complexity on individual task performance: accuracy (Mann-Whitney U-test)

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Error/ASU	-1.337	.181	.5	-.379	.704	.09	-1.827	.068	.54
Errors/ 100words	-1.231	.218	.5	-.365	.715	.05	-1.666	.096	.54
TLU_prep	-.334	.738	-.22	-3.203	.001**	-.84	-2.690	.007**	-.86

* α significant at p<.05
** α significant at p<.016

7.5.2.4 Structural complexity

As revealed by descriptive statistics (table 23), Words/Clause and Clauses/AS-units captured different syntactic behaviors. In the former, the structural complexity of speech decreased as a function of increasing cognitive complexity (*M*=7.68 in task 1 and *M*=6.57 in task 3). At the same time, an increase can be observed in the amount of subordination from task 1 to task 3 (*M*=1.21 and *M*=1.40, respectively). The participants maintained the same overall level of complexity, measured by Words/AS-units, as exemplified by identical scores obtained for tasks 1 and 3 (*M*=9.14), with a negligible difference between these two tasks and task 2 (*M*=8.98), However, producing simultaneously fewer words (Words/clause) and more clauses (Clauses/AS-units) suggests that the participants decomplexified their speech at the clause-internal level, but complexified it at the subordination level.

In the case of Words/Clause, Mann-Whitney U-test (table 27) showed that task 1 was performed with greater structural complexity compared to task 2, and compared to task 3. As regards Clauses/AS-unit, the only significant difference was detected between task 1 and task 3, with task 3 triggering more subordination.

Table 27. Impact of cognitive complexity on individual task performance: structural complexity (Mann-Whitney U-test)

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ ASU	-.091	.927	.87	-.211	.833	-.12	-.175	.861	.000
Words/ Clause	-2.553	.011**	.29	-1.194	.232	.41	-3.714	.000**	1.03
Clause/ ASU	-1.748	.080	-.42	-.857	.391	-.35	-2.735	.006**	-.94

* α significant at $p<.05$
** α significant at $p<.016$

7.5.2.5 Summary of results: the effect of cognitive task complexity on performance in individual task performance condition

In the individual task performance condition, the different dimensions of performance showed the following patterns as a result of increasing cognitive demands across the three tasks. Task 2 led to most speech rate, at the same time as triggering the highest rate of occurrence of hesitations. Tasks 1 and 3 displayed similar patterns on all three measures. Regarding accuracy, there was an overall pattern of increase in the use of target-like forms as the cognitive complexity of tasks increased, and the differences turned out to be most marked in the specific measure, target-like use of prepositions. As cognitive complexity increased, so did the diversity of the lexical forms used, and it was particularly prominent in the case of the measure D, although Guiraud's Index showed the same tendency. The participants' syntactic behavior displayed a pattern in which more subordination occurred simultaneously with a lower mean length of clause, as measured by Clauses/AS-units and Words/clause, respectively.

7.5.2.6 Summary of the effect of cognitive task complexity on performance in sequencing conditions and baseline data

When comparing the effect of cognitive task complexity in the sequencing conditions and in baseline data, several observations can be made. Speech rate received the highest score on task 2, and in both conditions it was accompanied by a concurrent increase in the amount of dysfluency. In the sequencing conditions task 3 led to the lowest speech rate, and in baseline data the same held true for task 1 (however, the differences were very small in magnitude). The results obtained for both conditions are illustrated in figure 2 (Rate B) and figure 3 (Dysfluency).

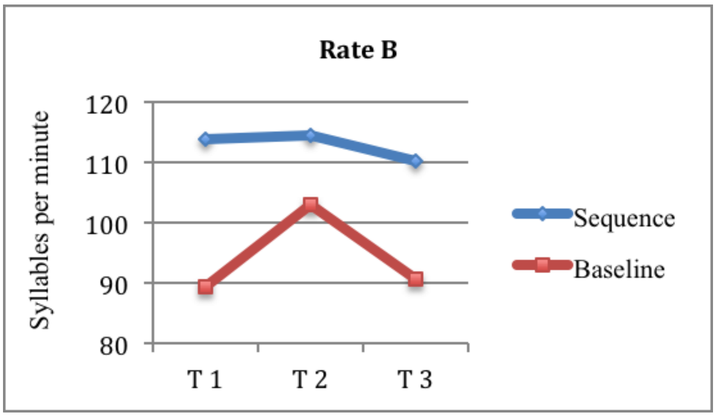


Figure 2. Rate B in sequencing and baseline condition

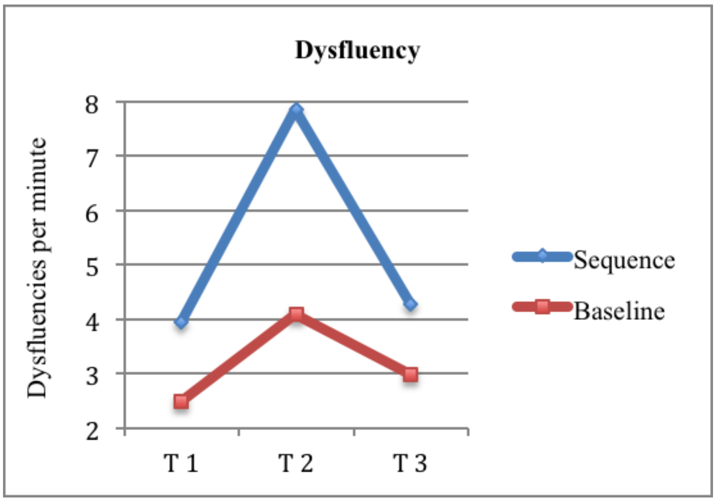


Figure 3. Dysfluency in sequencing and baseline condition

In both conditions an overall tendency for fewer non target-like forms can be observed as cognitive complexity increased; however, this pattern was particularly marked in the specific measure of target-like use of prepositions (figure 4), with the global measures revealing only minor differences.

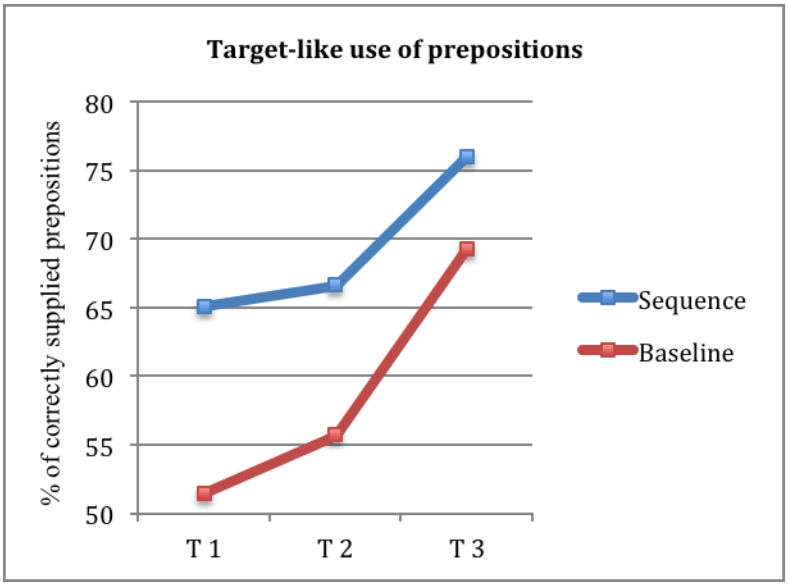


Figure 4. Target-like use of preposition in sequencing and baseline condition

In both conditions the increase in accuracy took place simultaneously with a wider range of lexical forms used. In the individual task performance condition, both of the employed measures revealed this pattern, but only D did so in the sequencing conditions, as displayed in figure 5 (Guiraud's Index) and figure 6 (D).

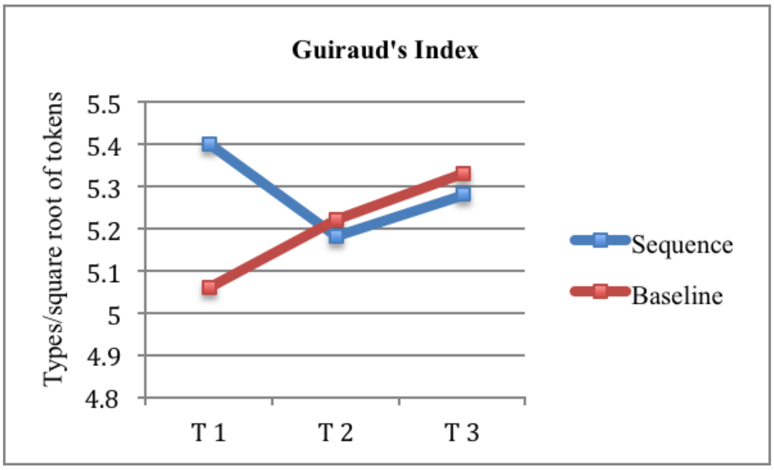


Figure 5. Guiraud's Index in sequencing and baseline condition

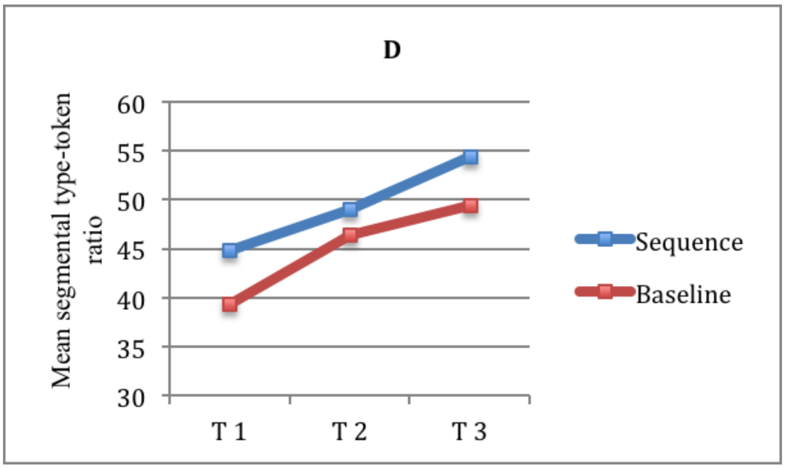


Figure 6. *D in sequencing and baseline condition*

The picture obtained in the three dimensions described above is a fairly homogenous one, but it is at the level of syntactic complexity that interesting differences emerged. The speech of participants in the sequencing condition turned out to be more syntactically complex at the clausal level (as measured by Words/clause, figure 7), whereas the participants in the individual task performance complexified their speech at the level of subordination (as measured by Clauses/AS-units, figure 8). The latter suggests that in the most complex task the effect of sequencing accumulated, leading to deeper-level complexification in the case of the participants who performed three tasks.

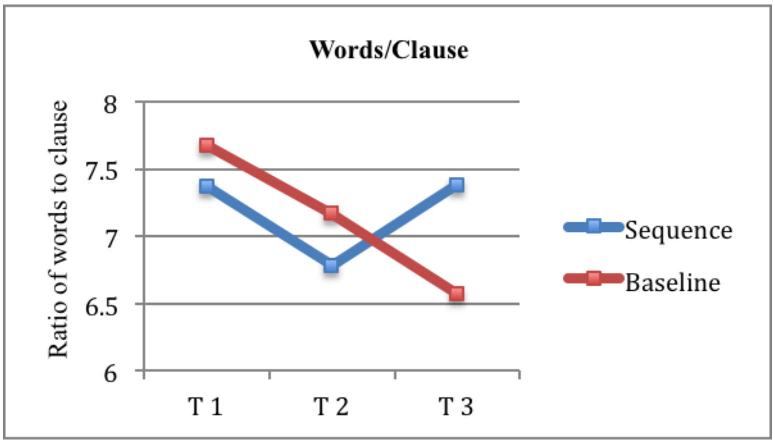


Figure 7. *Words/clause in sequencing and baseline condition*

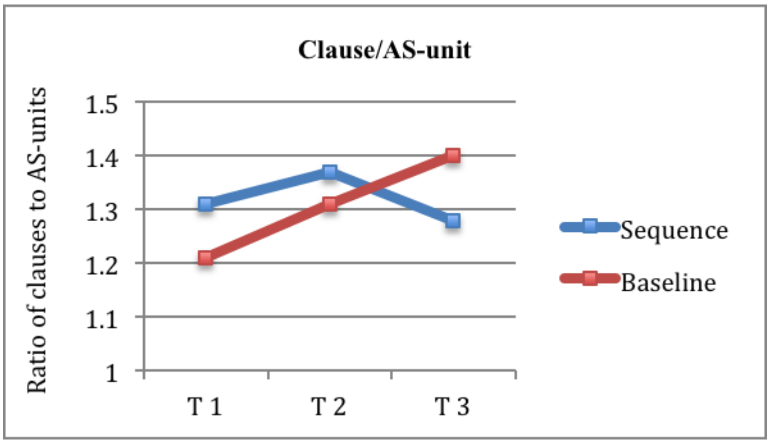


Figure 8. *Clauses/AS-unit in sequencing and baseline condition*

7.6 The effect of sequencing on performance

In this section the results related to the effect of sequencing on performance are presented. As was already mentioned in the Methodology chapter, the issue of sequencing was explored from two perspectives. The first approach to analyzing sequencing consisted in comparing the overall differences in performance between two conditions: simple-complex and randomized. The question pertinent to this exploration was that of whether the simple-complex sequence led to an optimal performance, which is the fundamental claim of the SSARC model of pedagogic task sequencing. The second approach concerned the performance on each of the individual tasks performed in isolation versus in a sequence of tasks. In order to explore the latter issue, the performance of the simple, complex, and the most complex task in the simple-complex sequence was compared to the same task performed in the absence of any other task.[28]

[28] Although the initial idea was for task 3 to be only object of this analysis, a close inspection of patterns present in descriptive statistics prompted a decision to include also tasks 1 and 2 in this analysis. This decision is further explained later in this chapter.

7.6.1 Sequencing tasks from simple to complex versus randomized sequencing

The research question pertaining to the first of the above-mentioned issues was formulated in the following way:

Research question 2a: How do different sequencing orders (simple-complex vs. randomized) affect performance in terms of CAF?

Table 28 presents the descriptive statistics obtained for the comparison of the simple-complex versus the randomized sequence.

7.6.1.1 Fluency

As can be observed in table 28 (descriptive statistics), a common tendency in both sequencing conditions was that speech rate decreased as cognitive complexity increased: task 3, compared to the other two tasks, triggered the lowest speech rate in both groups (for Rate B, *M*=107.26 in SC, and *M*=113.18 in RAN). At the same time, in the simple-complex condition it was task 1, which led to the highest speech rate (*M*=111.22), and the same was true for task 2 in the randomized condition (*M*=118.76). A noteworthy finding is that, across tasks, it was the randomized sequence, and not the simple-complex one, that triggered higher overall values for both Rate A and Rate B. This pattern of results was further reiterated in Dysfluency ratio: the randomized condition displayed a pattern of fewer hesitations across tasks, compared to the simple-complex sequence. In both conditions tasks 1 and 3 generated a very similar number of hesitations (*M*=4.23 and *M*=4.50 in SC, and *M*=3.66 and 4.04 in RAN, on tasks 1 and 3, respectively). However, in both conditions there was a considerably higher amount of repair fluency in task 2 (*M*=8.41 in SC and *M*=7.31 in RAN).

Table 28. Descriptive statistics of tasks (simple-complex versus randomized sequencing): means and standard deviations

Dependent variable		**Task 1**		**Task 2**		**Task 3**	
		Simple-complex	***Randomized***	***Simple-complex***	***Randomized***	***Simple-complex***	***Randomized***
		M (SD)	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***
Fluency	Rate A	120.5 (26.38)	124.39 (26.77)	122.50 (25.66)	128.87 (32.11)	118.75 (26.63)	122.54 (29.64)
	Rate B	111.22 (25.88)	116.45 (25.92)	110.35 (24.33)	118.76 (31.12)	107.26 (24.49)	113.18 (29.97)
	Dysfluency	4.23 (1.94)	3.66 (1.59)	8.41 (3.78)	7.31 (4.14)	4.50 (1.96)	4.04 (1.68)
Accuracy	Errors / ASU	0.97 (0.49)	1.19 (0.52)	0.96 (0.71)	1.18 (0.60)	0.91 (0.62)	1.03 (0.50)
	Errors / 100w	10.62 (6.17)	12.44 (5.58)	10.17 (5.52)	12.55 (6.32)	9.66 (6.16)	11.31 (6.40)
	TLU prep	68.12 (18.70)	61.96 (17.24)	71.75 (17.16)	61.52 (12.82)	76.39 (13.15)	75.64 (8.69)
Structural complexity	Words /ASU	9.02 (1.18)	9.53 (1.63)	9.05 (1.89)	9.31 (1.63)	9.02 (1.23)	9.37 (1.39)
	Words/ Clause	7.22 (1.36)	7.52 (0.83)	6.77 (1.42)	6.79 (1.06)	7.60 (1.84)	7.16 (1.25)
	Clause/ ASU	1.34 (0.66)	1.28 (0.24)	1.36 (0.24)	1.38 (0.21)	1.23 (0.24)	1.33 (0.22)
Lexical complexity	Guiraud's Index	5.36 (0.63)	5.44 (0.84)	5.08 (0.67)	5.28 (0.76)	5.12 (0.59)	5.44 (0.67)
	D	44.32 (11.30)	45.41 (13.41)	47.07 (12.10)	50.82 (12.70)	52.32 (12.31)	56.43 (12.50)

The different measures employed therefore showed that speech rate decreased as a consequence of increasing cognitive complexity, with the lowest speech rate in the case of the most complex task. Task 2, on the other hand, led to an escalation of repair behavior in both conditions.

The results of the repeated measures ANOVA (table 29) revealed that neither speed fluency nor repair fluency were statistically significantly affected by the sequencing condition.

Table 29. The effect of sequencing on performance (fluency): Repeated-measures ANOVA

Dependent variable	**Sum of squares**	**Wilk's Lambda**	***F***	***Df***	***p***	**Effect Size**
Rate A	64.110	.994	.186	2	.831	.006
Rate B	83.972	.991	.256	2	.775	.009
Dysfluency	3.526	.986	.418	2	.660	.014

* α significant at p<.05
** α significant at p<.016

7.6.1.2 Lexical complexity

Different patterns of lexical behavior were displayed by the two measures employed to capture performance patterns in lexical complexity. Guiraud's Index detected only minor variation between the three tasks, and this was particularly prominent in the randomized condition, where tasks 1 and 3 triggered the exact same value (*M*=5.44), and task 2 generated a slightly lower score (*M*=5.28). The participants in the simple-complex condition displayed a tendency for most lexical complexity on task 1 (*M*=5.36), with tasks 2 and 3 leading to somewhat less complex behavior (*M*=5.08 and *M*=5.12, respectively). The groups therefore displayed different patterns. By contrast, the results obtained for D revealed that the participants in both conditions systematically used more varied lexis from task 1 to task 3 (*M*=44.32 and *M*=52.32 in SC; *M*=45.41 and *M*=56.43 in RAN, respectively).

The results obtained for D indicated that, irrespective of the order in which the tasks were performed, it was the most complex task that triggered the most lexically rich production. It can therefore be speculated that it was the cognitive complexity of tasks, rather than the sequencing order in which they were performed, that brought about qualitative changes in performance. As was the case with fluency, in lexical complexity on all tasks and measures, the participants in the randomized condition obtained overall higher values than their simple-complex counterparts.

A repeated measures ANOVA (table 30) showed no main effect of the sequencing condition on lexical complexity, as measured by Guiraud's Index or D.

Table 30. The effect of sequencing on performance (lexical complexity): Repeated-measures ANOVA

Dependent variable	**Sum of squares**	**Wilk's Lambda**	***F***	***Df***	***p***	**Effect Size**
Guiraud's Index	.417	.966	1.012	2	.370	.090
D	87.215	.980	.575	2	.566	.020

* α significant at p<.05
** α significant at p<.016

7.6.1.3 Accuracy

The results obtained for general accuracy measures showed a trend in which sequencing tasks from simple to complex led to the generation of fewer non-target like forms than randomized sequencing did, as displayed in table 28 with descriptive statistics. However, the differences obtained for these measures were very small: in Errors/AS-unit, task 1 yielded a difference of .022 errors to the advantage of the simple complex condition (*M*=0.97 in SC and *M*=1.19 in RAN), and in Errors/100 a difference of 2.38 can be observed in task 2 (*M*=10.17 in SC and *M*=12.55 in RAN).

The specific accuracy measure revealed a peculiar and somewhat counterintuitive pattern between the two groups. The line of performance in the simple task showed that the participants in the simple-complex condition produced significantly fewer non-target like forms, with *M*=68.12 and *M*=61.96 in SC and RAN, respectively. The same held true for the results obtained for the complex task: *M*=71.75 and *M*=61.52, respectively. However, the differences disappeared in the case of the most complex task, with *M*=76.39 in SC and *M*=75.64 in RAN on this task, the difference being only marginally in favor of the simple-complex condition.

In the simple-complex condition, Wilcoxon signed-ranks tests (table 31) revealed the differences to be significant in the case of target-like use of prepositions. Their exact location was observed between tasks 2 and 3, and between tasks 1 and 3, indicating that a stronger pattern was found for the latter pair of tasks. The global measures did not display significant differences.

Table 31. Pairwise comparisons of tasks in the simple-complex sequencing condition (accuracy): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Errors/ASU	-0.761	.447	.01	-0.401	.688	.07	-1.399	.162	.10
Errors/ 100words	-0.483	.629	.07	-0.093	.926	.08	-0.895	.371	.15
TLU_prep	-1.460	.144	-.2	-1.979	.048*	-.3	-3.198	.001**	-.51

* α significant at $p<.05$
** α significant at $p<.016$

Table 32 shows the results obtained for accuracy in the randomized sequencing condition, as detected by Wilcoxon signed-ranks tests.

Table 32. Pairwise comparisons of tasks in the randomized sequencing condition (accuracy): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Errors/ ASU	-.357	.721	.01	-2.118	.034*	.27	-2.638	.008**	.31
Errors/ 100words	-.184	.854	-.01	-2.458	.014**	.19	-2.047	.041*	.18
TLU_prep	-.442	.658	.02	-4.573	.000**	-1.28	-3.569	.000**	-1

* α significant at p<.05
** α significant at p<.016

In the case of target-like use of prepositions, there was a statistically significant difference between tasks 2 and 3 and between tasks 1 and 3, in both cases task 3 yielding much more accurate performance. The same line of behavior was observed in the simple-complex condition. This pattern was further confirmed by global accuracy measures. As measured by Errors/100 words, performance was significantly more accurate on task 3 than on task 2, and it was also more accurate on task 3 than on task 1. In the case of Errors/AS-units, significant differences were detected between tasks 2 and 3 and between tasks 1 and 3. Although in the case of these speakers statistically significant differences were detected in both specific and global accuracy measures, the effect of cognitive complexity on production was overall greater in the case of the specific measure, which is evidenced in the *p* values and detected effect sizes.

To sum up the results for accuracy, in the case of the simple-complex condition significant differences were detected only for the specific measure, and in the randomized condition they were found for all measures in the case of two comparisons: tasks 2 and 3, and tasks 1 and 3. No differences were found between tasks 1 and 2 in either of the two conditions. Overall, however, both groups showed a trend for systematically more use of target-like forms as cognitive task complexity increased. Unlike in the previously discussed dimensions (fluency and lexical complexity), in accuracy the participants in the simple-complex condition showed an advantage over the other group when comparing each of the tasks, which may be an indication of a potential role of sequencing in generating more target-like forms. At the same time, regardless of the sequencing condition, both groups displayed increases in overall accuracy levels, which, in turn, may suggest a greater role of cognitive task complexity than sequencing.

7.6.1.4 Structural complexity

As can be observed in table 28 with descriptive statistics, overall syntactic complexity, as measured by Words/AS-units, was very stable in the simple-complex condition (*M*=9.02 on tasks 1 and 3, and *M*=9.05 on task 2), and quite stable in the randomized condition, with task 1 rendering the most structural complexity (*M*=9.53), and task 2 the least (*M*=9.31). However, the differences were generally very small in magnitude. Opposite trends in both groups can be observed when it comes to clause-internal expansion as measured by Words/Clause. In the simple-complex group, it was task 3 that triggered the highest score (*M*=7.60); however, in the randomized condition it was task 1 (*M*=7.52). In both groups, task 2 led to the most subordination (Clauses/AS-units) (*M*=1.36 in SC and *M*=1.38 in RAN). At the same time, task 3 triggered the lowest score for subordination in the simple-complex sequence (*M*=1.23), and in the other group it was task 1 (*M*=1.28). Wilcoxon signed-ranks test (table 33) was employed in order to investigate the location of significant differences.

Table 33. Pairwise comparisons of tasks in the simple-complex sequencing condition (structural complexity): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ASU	-0.679	.497	-.01	-0.679	.497	.01	-0.267	.789	.000
Words/Clause	-2.303	.021*	.32	-1.604	.109	-.5	-.668	.504	-.23
Clause/ASU	-3.017	.003**	-.04	-1.017	.309	.54	-1.121	.262	.22

* α significant at *p*<.05
** α significant at *p*<.016

In Words/Clause, task 1 generated significantly more complex speech than task 2. In Clauses/AS-units significant differences were also observed between tasks 1 and 2, but here task 2 was the one which led to the most structurally complex output compared to task 1. No effects were found for Words/AS-units.

Similar patterns can be observed in the randomized sequencing condition (table 34): significant differences were detected for Words/Clause and Clause/AS-units.

Table 34. Pairwise comparisons of tasks in the randomized sequencing condition (structural complexity): Wilcoxon signed-ranks test

Dependent variable	**Task 1-2**			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ ASU	-0.576	.565	.13	-0.278	.781	-.03	-0.508	0.611	.10
Words/ Clause	-3.319	.001**	.76	-1.018	.309	-.31	-2.499	.012**	.33
Clause/ ASU	-2.841	.005**	-.44	-.941	.347	.23	-1.440	.150	-.21

* α significant at *p*<.05
** α significant at *p*<.016

In Words/Clause, Wilcoxon signed-ranks test revealed a significant difference between tasks 1 and 2 and between tasks 1 and 3. In the case of subordination index, a significant difference was displayed between tasks 1 and 2, with task 1 leading to more structurally complex speech in both of the measures. In all reported pairwise comparisons *p*<.016, and the reported effect sizes were medium to large. No significant effects were found for the third of the employed measures, Words/AS units, but, as was the case in the other measures of structural complexity, task 1 was the one which triggered the most structurally complex speech.

To sum up, performing tasks in the order from simple to complex led to complexifying speech at the level of clause (Words/Clause), which was not the case in the randomized condition. This finding may suggest a potential role of simple-complex sequencing in generating more elaborated clauses as cognitive task demands increase. The linguistic behavior captured by the other measures did not reveal fundamental differences, with an overall stable level of syntactic complexity and a comparable amount of subordination produced.

7.6.1.5 Summary of results: simple-complex versus randomized sequencing

The results presented in this section dealt with the issue of the effect of sequencing of pedagogic tasks (simple-complex versus randomized sequencing) on performance. In the case of fluency and lexical complexity, the participants in the randomized condition delivered their speech faster, with fewer dysfluencies, and with a greater variety of lexical forms compared to the participants in the other condition. Both groups showed a tendency for decreased speech rate in the case of the most complex task, but irrespective of the order of performing the tasks, the greatest number of dysfluencies was produced in task 2. Regarding lexical complexity, as measured by D, in both groups there was an increase in the diversity of lexis, whereas Guiraud's Index showed little overall variation between the three tasks. The results obtained for accuracy are generally in favor of the simple-complex sequencing condition: there is some evidence to suggest that performing tasks from simple to complex is conducive to more target-like production than when tasks are performed in a randomized order. However, both conditions showed a tendency for improvement in the accuracy of their speech, which is evidenced in a gradual decrease in errors, and a concurrent increase in target-like use of prepositions. The results obtained for accuracy in both groups (Errors/100 words and target-like use of prepositions) are graphically represented in figures 9 and 10.

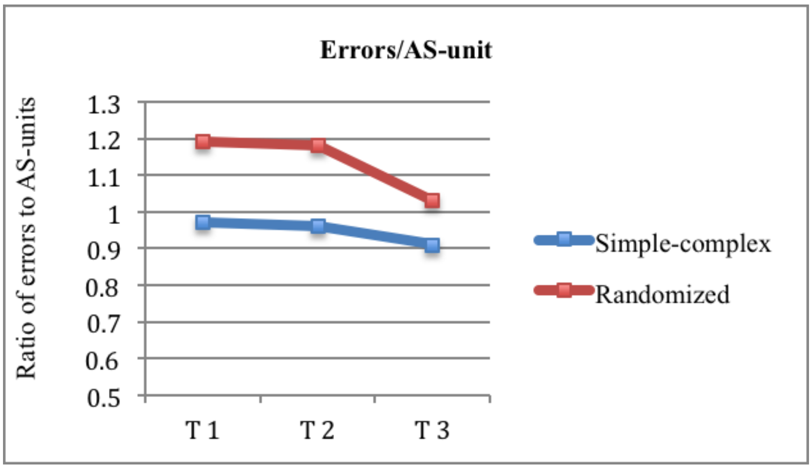


Figure 9. Errors/AS-unit in simple-complex and randomized sequencing

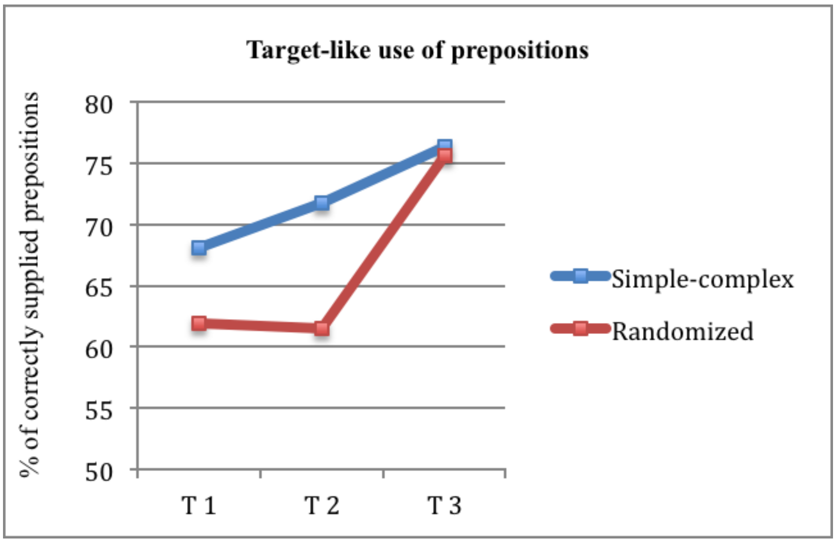


Figure 10. Target-like use of prepositions in simple-complex and randomized sequencing

Both groups showed an overall stable level of holistic syntactic complexity (Words/AS-units), but the other two measures revealed that they were affected by cognitive complexity in opposite ways: in the simple-complex sequencing condition, task 3 led to greatest clause length and lowest subordination, and the same was found for the randomized sequencing group for task 1. These results are illustrated in figures 11 and 12.

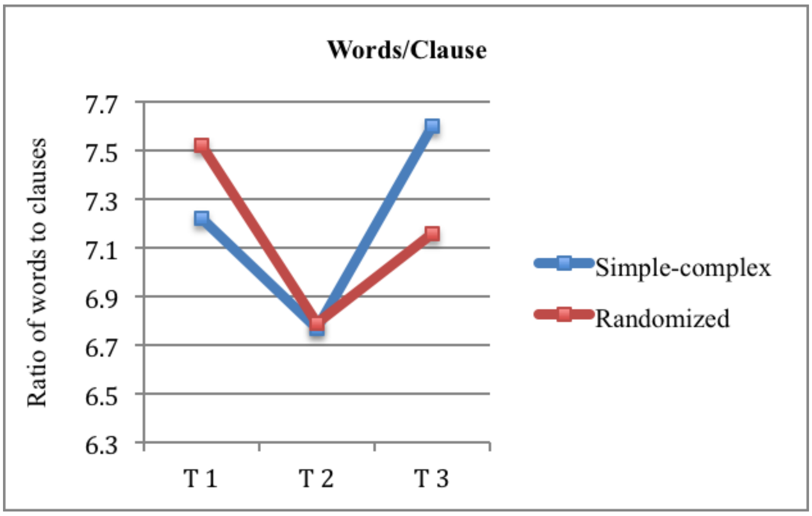


Figure 11. *Words/clause in simple-complex and randomized sequencing*

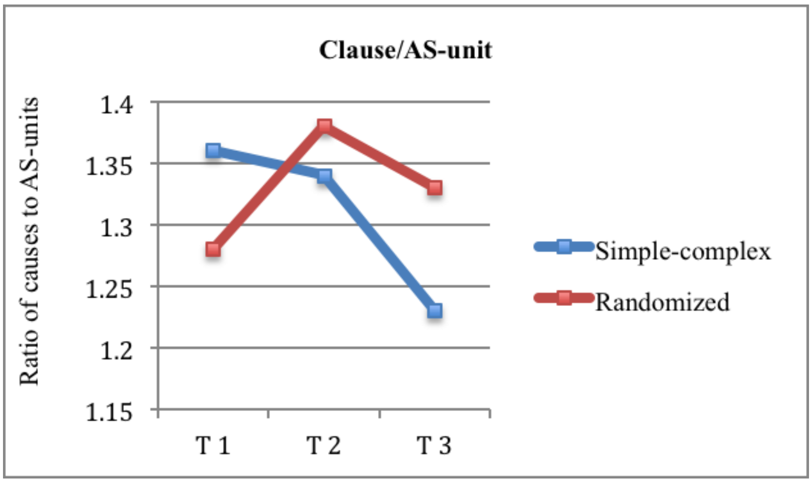


Figure 12. *Clauses/AS-unit in simple-complex and randomized sequencing*

The patterns observed for all the dimensions, both the statistically significant and non-significant ones, suggest that overall the sequencing condition did not seem to exert a substantial influence on performance, except for accuracy. Irrespective of the sequence in which the tasks were performed, the output produced by participants in both sequencing conditions was affected in comparable ways. At the same time, the observable changes in the quality of performance displayed by both sequencing conditions indicate that designed cognitive task complexity levels had an impact on performance. This finding leads to a preliminary interpretation that it was the cognitive task complexity, rather than sequencing, that had an impact on the speakers' oral production.

7.6.2 Effect of sequencing on performance: tasks as performed in a simple-complex sequence versus in isolation

This section deals with the second of the previously mentioned perspectives on the issue of sequencing. Its objective was to explore the linguistic behavior of participants in two conditions: isolated, individual task performance and the simple-complex sequence. The fundamental issue related to this research question was that of whether the performance on the most complex task (task 3) benefitted from prior exposure to the simple and the complex task, as compared to the most complex task being the only performed task, (i.e., without previously performing task 1 or task 2). The research question pertinent to this exploration was formulated in the following way:

Research question 2b: Does performing a task in a sequence lead to qualitatively different output compared to performing a task in isolation?

Although the initial idea was to explore the linguistic behavior exclusively on the final task as performed in a sequence and in isolation, in order to obtain a complete insight into the issue of sequencing, a decision was taken to analyze performance on each of the corresponding tasks in the two conditions (sequencing and isolation): task 3, task 2, and task 1. Although one would expect differences to emerge primarily, if not exclusively, in the case of the most complex task, a close inspection of descriptive statistics revealed that intricate patterns of performance emerged in all three tasks, including the simple one. It may seem counterintuitive insofar as the set up for this task was exactly the same in the sequencing group and the other one (i.e., it was the only task performed by the isolated performance group and the first one in a sequence performed by the simple-complex group). Given the findings obtained in raw data, a comprehensive approach was adopted whereby each of the three tasks performed in a sequence and in isolation were subjected to analyses. The order of presenting the results is from the most complex one to the simple one, following the original idea of including only the most complex task in the analysis. Descriptive statistics for all three tasks and dimensions are displayed in table 35.

Table 35. Descriptive statistics of tasks: simple-complex sequencing and individual task performance

Dependent variable		**Task 1**		**Task 2**		**Task 3**	
		Simple-complex	***Baseline***	***Simple-complex***	***Baseline***	***Simple-complex***	***Baseline***
		M (SD)	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***
Fluency	Rate A	120.50 (26.38)	95.22 (43.35)	122.50 (25.66)	113.27 (35.01)	118.75 (26.63)	98.38 (34.99)
	Rate B	111.22 (25.88)	89.46 (42.39)	110.35 (24.33)	102.98 (32.88)	107.26 (24.49)	90.76 (34.77)
	Dysfluency	4.23 (1.94)	2.50 (1.21)	8.41 (3.78)	4.10 (2.56)	4.50 (1.96)	2.99 (1.83)
Accuracy	Errors / ASU	0.97 (0.49)	1.60 (1.17)	0.96 (0.71)	1.14 (0.56)	0.91 (0.62)	1.10 (0.57)
	Errors / 100w	10.62 (6.17)	17.32 (10.17)	10.17 (5.52)	12.94 (6.84)	9.66 (6.16)	12.56 (7.05)
	TLU prep	68.12 (18.70)	51.45 (22.84)	71.75 (17.16)	55.72 (13.59)	76.39 (13.15)	69.31 (18.18)
Structural complexity	Words /ASU	9.02 (1.18)	9.14 (1.36)	9.05 (1.89)	8.98 (1.27)	9.02 (1.23)	9.14 (1.37)
	Words/ Clause	7.22 (1.36)	7.68 (1.40)	6.77 (1.42)	7.17 (1.98)	7.60 (1.84)	6.57 (0.58)
	Clause/ ASU	1.34 (0.66)	1.21 (0.18)	1.36 (0.24)	1.31 (0.28)	1.23 (0.24)	1.40 (0.22)
Lexical complexity	Guiraud's Index	5.36 (0.63)	5.06 (0.83)	5.08 (0.67)	5.22 (0.47)	5.12 (0.59)	5.33 (0.74)
	D	44.32 (11.30)	39.31 (10.81)	47.07 (12.10)	46.43 (11.06)	52.32 (12.31)	49.44 (11.87)

7.6.2.1 Task 3 as performed in simple-complex sequence and in isolation

As can be observed in descriptive statistics (table 35), performing the most complex task in a sequence presented several advantages over performing it in the absence of the other tasks.

Regarding fluency, the speech rate of the participants in the simple-complex condition was higher than of those who performed it in the absence of any other task (for Rate B, *M*=107.26 and *M*=90.76, respectively). A notable difference between the two groups can be observed in repair fluency, which displayed a different pattern to speech rate. Surprisingly, task 3 performed in isolation from the other tasks led to considerably fewer hesitations, with *M*=2.99 for this group, and *M*=4.50 for the sequencing condition.

When it comes to accuracy, performing the third task after prior exposure to tasks 1 and 2 led to more accurate linguistic behavior than carrying it out in isolation. The differences were the most visible in the case of target-like use of prepositions, with *M*=76.39 in the simple-complex condition, and *M*=69.31 in the individual task performance condition. The trend towards more accuracy to the advantage of the sequencing condition was also present in the global measures, but to a lesser extent than in the specific measure. In Errors/AS-unit, *M*=0.91 in SC and *M*=1.10 in the other group, and for Errors/100 words *M*=9.66 in SC and *M*=12.56 in the other group.

At the level of lexis, the performance of both groups was quite similar, but the different measures employed showed different trends. Guiraud's Index was higher for the participants in the baseline condition (*M*=5.33 vs. *M*=5.12 in the simple-complex condition), but the score for D was higher in the simple-complex condition (*M*=52.32, vs. *M*=49.44 in the baseline condition).

Regarding structural complexity, those participants who only performed task 3 produced more subordination than their counterparts in the other condition (*M*=1.40 and *M*=1.23, respectively). However, the opposite was true for mean length of clause: the participants in the sequencing condition complexified their speech at the clausal level (Words/clause) more than the other group (for sequencing condition, *M*=7.60, and for the other group, *M*=6.57). However, both groups' overall level of complexity was quite similar, as shown by Words/AS-units (*M*=9.02 in SC vs. *M*=9.14 in baseline). Mann Whitney U-test (table 36) was employed to detect significant differences.

Table 36. Effect of simple-complex sequencing compared to individual task performance (task 3): Mann Whitney U-test

Dependent variable		**Task 3 Simple-complex - individual task**	
		p	**Effect size**
Fluency	RateA	.019*	.65
	RateB	.019*	.54
	Dysfluency	.007**	.79
Accuracy	Error/ASU	.148	-.31
	Error/100words	.109	-.43
	TLU_prep	.104	.44
Structural complexity	Words/ASU	.656	-.09
	Words/Clause	.047*	.75
	Clause/ASU	.022*	-.73
Lexical complexity	Guiraud's Index	.440	-.31
	D	.357	.23

* α significant at p<.05
** α significant at p<.016

When comparing task 3 in both conditions, it can be observed that statistically significant differences were yielded in the area of fluency and structural complexity; more specifically, in Rate A, Rate B, Dysfluency ratio, Words/Clause, and Clauses/AS-units. Out of the five measures for which significant differences were detected, performing the final task in a sequence had a positive influence on performance in the case of three measures: Rate A, Rate B, and Words/clause. The significant differences obtained for the other two measures were to the advantage of the individual task performance condition, and these were: Dysfluency ratio and Clauses/AS-unit. On the basis of the obtained *p* values and effect sizes, the effects can be described as moderate to large, except for the value obtained for the repair fluency measure.

To sum up the results for this section, performing the third task in a sequence led to a higher speech rate, fewer errors, and more grammatical complexity, compared to the same task performed without prior exposure to any other task. There were no fundamental differences between the two groups' performance on structural complexity. The findings obtained provide evidence to suggest that simple-complex sequencing was facilitative to language production, rendering the performance speedier, closer to native-speaker norms, and richer at the syntactic level.

7.6.2.2 Task 2 as performed in simple-complex sequence and in isolation

The trends obtained for fluency in task 2 were overall similar to those obtained in task 1 (table 35): higher speech rate can be observed in the simple-complex condition (for Rate B, M=110.35) than in baseline condition (M=102.98). In the former there was a concurrent considerable intensification of the number of hesitations, with the participants in this group displaying more than double the total amount of repair behavior (M=8.41) than their counterparts in the baseline condition (M=4.10).

Higher scores for the simple-complex condition compared to the baseline were revealed in accuracy across all the measures employed. Target-like use of prepositions prompted the most prominent difference (M=71.75 in SC and M=55.72 in baseline), but this pattern was also confirmed by global measures. Participants in the simple-complex condition produced fewer errors per AS-unit (in this group *M*=0.96, versus *M*=1.14 in baseline), and fewer errors per 100 words (*M*=10.17 for this group, versus *M*=12.94 in baseline).

As far as the structural complexity of linguistic output is concerned, the participants in the simple-complex condition complexified their speech at the level of subordination and those in the other group did so at the clausal level. This is evidenced in the figures obtained for Clauses/AS-units and Words/clause. Regarding the former, *M*=1.36 in SC, and *M*=1.31 in baseline; in the latter, higher overall syntactic complexity was produced by the participants in the baseline condition (*M*=7.17), and *M*=6.77 in the simple-complex one. Finally, no fundamental differences were detected for the richness of lexical output, with contrary results and marginal differences obtained for Guiraud's Index (*M*=5.08 in simple-complex vs. *M*=5.22 in baseline) and D (*M*=47.07 in simple-complex and *M*=46.43 in baseline).

As displayed in table 37 (Mann-Whitney U-test), statistically significant differences were detected in the case of Dysfluency ratio and target-like use of prepositions, the results for both of which suggest a strong effect. No significant effects were found in the case of the other measures.

To sum up, in task 2 very clear effects were observed for dysfluency and accuracy. The participants in the simple-complex produced a larger total number of hesitations than the other group, but at the same time the accuracy of their linguistic output showed less deviation from norms. The findings obtained for lexical complexity and structural complexity did not present major differences between the two groups.

Table 37. Effect of simple-complex sequencing compared to individual task performance (task 2): Mann Whitney U-test

Dependent variable		**Task 2** Simple-complex - individual task **p**	**Effect size**
Fluency	RateA	.460	.3
	RateB	.538	.25
	Dysfluency	.000**	1.33
Accuracy	Error/ASU	.121	-.28
	Error/100words	.094	-.44
	TLU_prep	.001**	1.03
Structural complexity	Words/ASU	.652	.04
	Words/Clause	.460	-.23
	Clause/ASU	.918	.19
Lexical complexity	Guiraud's Index	.330	-.24
	D	.943	.05

* α significant at $p<.05$
** α significant at $p<.016$

7.6.2.3 Task 1 as performed in simple-complex sequence and in isolation

As displayed in table 35, when it comes to fluency, the same pattern as in the other two tasks can be observed. The participants who performed task 1 in a sequence showed a speedier rate of delivery than the other group (for Rate B, *M*=111.22 in SC, and *M*=89.46 in baseline), with a simultaneous occurrence of greater repair fluency (*M*=4.23 in the simple-complex condition, and *M*=2.50 in baseline).

The participants in the simple-complex condition showed considerably less deviation from target-like norms in all three measures employed. They displayed a more correct use of prepositions (*M*=68.12 vs. *M*=51.45 in baseline), lower error rate per AS-unit (*M*=0.97 vs. *M*=1.60 in baseline), and per 100 words (*M*=10.62 vs. *M*=17.32 in baseline). This pattern is consistent with the one previously described in tasks 2 and 3.

Regarding structural complexity, simple-complex sequencing led to greater subordination (*M*=1.34 for these participants, and *M*=1.21 in baseline) and less clause-internal expansion, as measured by Words/clause (*M*=7.22 in the simple complex-condition, and *M*=7.68 in baseline). Compared with the results obtained for the other two tasks, the same syntactic complexity pattern was observed in task 2, but task 3 triggered the opposite trend. Slightly higher values were obtained for lexical complexity in the simple-complex group, with *M*=5.36 for Guiraud's Index (*M*=5.06 in baseline), and *M*=44.32 for D (*M*=39.31 in baseline). Table 38 shows the statistically significant differences obtained in the Mann-Whitney U-test in task 1.

Table 38. Effect of simple-complex sequencing compared to individual task performance (task 1): Mann Whitney U-test

Dependent variable		**Task 1**	
		Simple-complex - individual task	
		p	**Effect size**
Fluency	RateA	.017*	.69
	RateB	.023*	.61
	Dysfluency	.001**	1.07
Accuracy	Error/ASU	.011**	-.72
	Error/100words	.010**	-.79
	TLU_prep	.008**	.79
Structural complexity	Words/ASU	.924	-.09
	Words/Clause	.647	-.33
	Clause/ASU	.536	.26
Lexical complexity	Guiraud's Index	.246	.40
	D	.268	.45

*   α significant at $p<.05$
** α significant at $p<.016$

Significant differences were yielded in all measures of fluency and accuracy. As was the case in tasks 2 and 3, participants in the individual task performance condition exhibited less dysfluency behavior than their counterparts in the sequencing condition on Dysfluency ratio. In all the other measures, which yielded statistical significance, the sequencing condition generated a higher speech rate and more accurate linguistic behavior. This was the case in Rate A, Rate B, Error/AS-units, Error/100 words, and target-like use of prepositions. The dysfluency measure and all accuracy measures both showed large effect sizes and $p<.016$. In the other areas (structural complexity and lexical complexity) no significant differences were detected.

7.6.2.4 Summary of results: all tasks performed in a sequence versus in isolation

This section explored the linguistic behavior of two groups of speakers: those who performed a pedagogic task in the absence of exposure to any other task, and those who performed the same pedagogic task in a sequence from simple to complex. The analysis of the obtained findings revealed several interesting patterns in the data.

The results obtained in this section provided evidence to suggest that performing sequences of tasks, rather than individual tasks, may be potentially facilitative to speech production in the case of some dimensions. The performance on the most complex task benefitted from previous exposure to the other tasks in three aspects: the speed of speech delivery (figure 13), the conformity to language norms (figures 14 and 15), and syntactic complexity at clausal level (figure 16). Regarding the latter, a close look at descriptive statistics shows that the participants in the simple-complex condition produced the most complex speech at the clausal level on the most complex task, whereas the same task performed on its own generated the lowest figure of all three tasks. This finding may suggest that performing sequences of tasks is conducive to producing more elaborate clauses. These potential advantages of simple-complex sequencing were concurrent with a very clear disadvantage for dysfluency behavior: greater repair fluency was observed in the simple-complex condition than in the isolated performance.

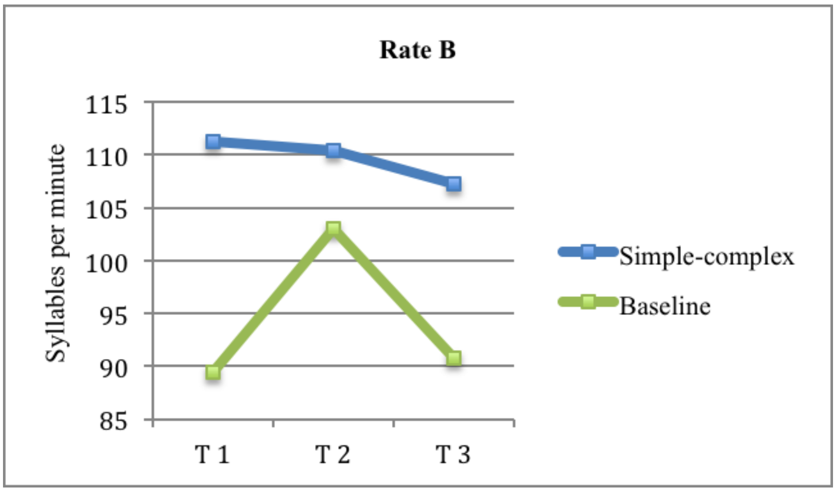


Figure 13. *Rate B in simple-complex sequencing and individual task performance*

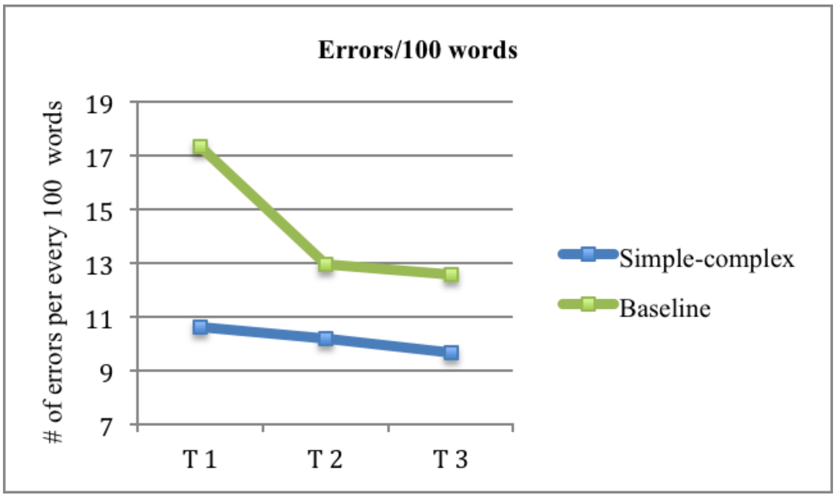


Figure 14. *Errors/100 words in simple-complex sequencing and individual task performance*

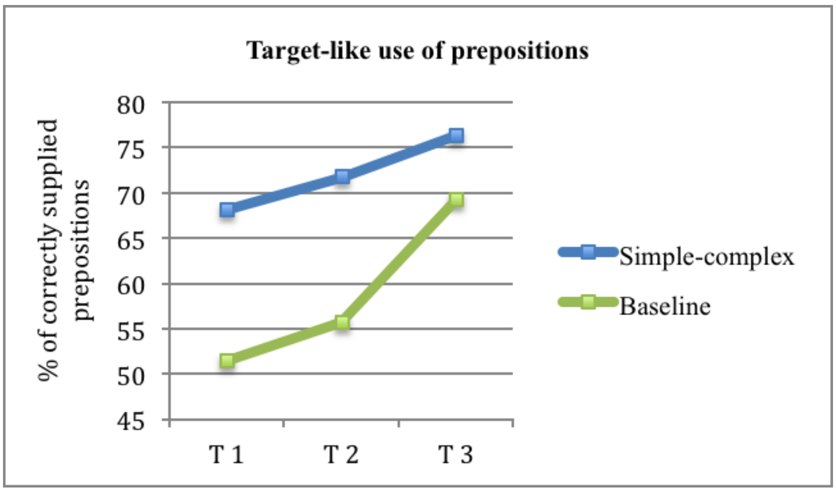


Figure 15. *Target-like use of prepositions in simple-complex sequencing and individual task performance*

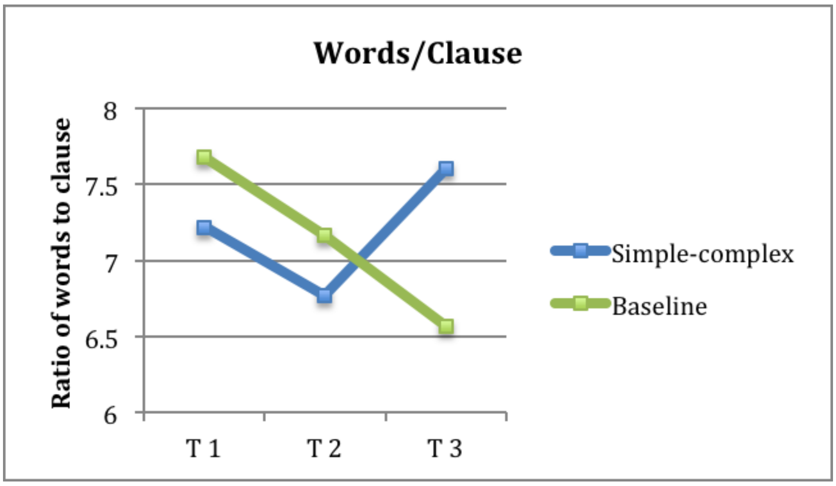


Figure 16. *Words/clause in simple-complex sequencing and individual task performance*

At the same time, it is quite intriguing that sequencing tasks from simple to complex had an overall very similar effect across tasks. For example, those participants who performed all three tasks systematically delivered their speech faster and produced fewer non target-like forms. This is particularly surprising in the case of the simple task, given that the circumstances of performing this task were the same for both groups (i.e., it was the first task in the simple-complex condition, and the only one delivered by the participants in the baseline condition). It may be the case that performing a sequence of tasks was in itself a circumstance which predisposed the participants in this condition to specific linguistic behaviors.[29]

[29] The participants were informed that they would perform a sequence of tasks, and not a single task, directly before engaging in the performance of the first task.

7.7 The effect of proficiency on performance

This section deals with the role of proficiency in task-based performance. As was the case with the effect of cognitive task complexity and the effect of sequencing, the role of proficiency was also explored from two perspectives. The first perspective concerns the way task complexity affected speakers of different proficiency levels. The second one explores a potentially facilitative role of the simple-complex sequence in low-proficiency speakers as compared to the randomized sequence.

7.7.1 The effect of cognitive task complexity in speakers of different proficiency levels

The research question, which explored the first of the aforementioned perspectives on proficiency, was formulated in the following way:

Research question 3a: How does task complexity affect the performance of speakers of different proficiency levels, as measured by CAF?

Table 39 presents the descriptive statistics obtained for both proficiency groups.

7.7.1.1 Fluency

Regarding speech rate, as shown by descriptive statistics (table 39), cognitive complexity led to a decrease from task 1 to task 3 in high proficiency speakers (for Rate B, *M*=126.13 and *M*=120.35, respectively). The performance of the low proficiency speakers was very stable across the three tasks, with very similar scores yielded by the simple and the most complex task (for Rate B, *M*=101.54, and *M*=100.10, respectively).

Table 39. Descriptive statistics of tasks for low and high proficiency speakers: means and standard deviations

Dependent variable		**Task 1**		**Task 2**		**Task 3**	
		Low proficiency	***High proficiency***	***Low proficiency***	***High proficiency***	***Low proficiency***	***High proficiency***
		M (SD)	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***
Fluency	Rate A	109.41 (27.40)	135.48 (17.83)	114.43 (29.56)	136.95 (23.96)	110.14 (29.28)	131.15 (22.57)
	Rate B	101.54 (26.84)	126.13 (17.98)	103.83 (27.49)	125.28 (24.56)	100.10 (28.82)	120.35 (21.75)
	Dysfluency	3.69 (1.57)	4.19 (1.97)	7.34 (4.21)	8.38 (3.72)	4.26 (1.78)	4.27 (1.91)
Accuracy	Errors / ASU	1.40 (0.47)	0.75 (0.30)	1.41 (0.76)	0.74 (0.27)	1.31 (0.60)	0.63 (0.21)
	Errors / 100w	15.42 (5.77)	7.64 (2.49)	15.23 (6.01)	7.49 (2.46)	14.45 (6.57)	6.52 (2.14)
	TLU prep	56.43 (18.24)	73.66 (13.44)	59.68 (16.42)	73.59 (11.97)	68.99 (9.78)	83.04 (7.14)
Structural complexity	Words /ASU	8.92 (1.11)	9.64 (1.63)	8.76 (1.89)	9.60 (1.52)	9.02 (1.22)	9.37 (1.40)
	Words/ Clause	7.34 (1.06)	7.39 (1.20)	6.88 (1.39)	6.67 (1.09)	7.43 (1.27)	7.33 (1.85)
	Clause/ ASU	1.24 (0.23)	1.38 (0.66)	1.28 (0.20)	1.45 (0.21)	1.23 (0.17)	1.33 (0.27)
Lexical complexity	Guiraud's Index	5.12 (0.67)	5.69 (0.69)	4.87 (0.67)	5.50 (0.62)	4.99 (0.60)	5.58 (0.55)
	D	41.37 (11.42)	48.37 (12.34)	44.29 (12.35)	53.60 (10.86)	48.61 (12.44)	60.14 (9.64)

The third of the employed measures, Dysfluency ratio, displayed the pattern in which the performance in both groups was very similar for tasks 1 and 3, but task 2 triggered a considerable increase in the amount of hesitations produced (*M*=7.34 for low proficiency speakers, and *M*=8.38 for high proficiency speakers). According to this measure, low proficiency speakers displayed a more stable fluency behavior across the three tasks, obtaining systematically slightly lower scores than their counterparts in the high proficiency group. A one-way repeated measures ANOVA (table 40) revealed significant differences for proficiency (speech rate) and for task complexity in the case of Dysfluency. No interaction effects were found.

Table 40. The role of task complexity and L2 proficiency in performance (fluency): Repeated-measures ANOVA

Dependent variable		***F***	***Df***	***p***	**Effect size**
Rate A	Task	2.277	2	.112	.074
	Proficiency	15.284	1	.000**	.209
	Task*Proficiency	.527	2	.593	.018
Rate B	Task	1.885	2	.161	.062
	Proficiency	14.535	1	.000**	.200
	Task*Proficiency	.408	2	.667	.014
Dysfluency	Task	15.424	2	.000**	.668
	Proficiency	.732	1	.399	.012
	Task*Proficiency	1.731	2	.186	.057

* α significant at $p<.05$
** α significant at $p<.016$

7.7.1.2 Lexical complexity

The two measures employed for lexical complexity revealed opposite patterns of performance (table 39). The results obtained for Guiraud's Index displayed no major differences between the three tasks in either of the proficiency groups. According to this measure, the most lexically rich performance was triggered by task 1 in both low and high proficiency speakers (*M*=5.12 vs. and *M*=5.69, respectively). The opposite pattern was revealed by D, which detected a gradual progression in the variety of lexis used from task 1 to task 3 (*M*=41.37 and *M*=48.61 in the low proficiency group, and in the high proficiency group *M*=48.37 and *M*=60.14, respectively).

As displayed in table 41, there was an effect of cognitive task complexity in the case of D, and an effect of proficiency in the case of both measures. No interaction effect was found in either of the variables. As the two measures used present different scenarios, the picture of lexical complexity is a mixed one. According to Guiraud's Index, the groups were affected by cognitive complexity in different ways, while D showed that both groups were positively affected by increasing cognitive demands.

Table 41. The role of task complexity and L2 proficiency in performance (lexical complexity): Repeated-measures ANOVA

Dependent variable		***F***	***Df***	***p***	**Effect size**
D	Task	28.013	2	.000**	.496
	Proficiency	12.867	1	.000**	.182
	Task*Proficiency	1.453	2	.242	.049
Guiraud's Index	Task	2.837	2	.067	.091
	Proficiency	20.518	1	.000**	.261
	Task*Proficiency	.065	2	.937	.002

* α significant at $p<.05$
** α significant at $p<.016$

7.7.1.3 Accuracy

According to descriptive statistics (table 39), cognitive task complexity had a beneficial effect on the accuracy of the participants in both proficiency groups. In the global accuracy measures, a substantial difference between task 3 and the other two tasks can be observed in the measure Errors/AS-units: in low proficiency speakers, *M*=1.31 on task 3 vs. *M*=1.40 on task 1, and in the other group, *M*=0.63 on task 3 vs. *M*=0.75 on task 1. No major differences were detected between tasks 1 and 2 in either of the groups. Errors/100 words showed a gradual pattern of decrease in errors as cognitive complexity increased, yielding a difference of 0.97 between tasks 1 and 3 in the low proficiency group, and a difference of 1.12 in the high proficiency speakers. Both groups also showed a pattern of increasingly more correct use of prepositions from the simple to the most complex task (*M*=56.43 vs. 68.99 in the low proficiency group, and *M*=73.66 vs. *M*=83.04 in the high proficiency group). Wilcoxon signed-rank test was employed to detect significant differences in accuracy (table 42).

Table 42. Impact of task complexity on high-proficiency learners (accuracy): Wilcoxon signed-ranks test

Dependent variable	***High-proficiency speakers***								
	Task 1-2			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Errors/ASU	-0.195	.846	.03	-1.78	.075	.45	-2.376	.018*	.46
Errors/ 100words	-.216	.829	.06	-2.293	.022*	.42	-2.551	.011**	.48
TLU_prep	-.134	.894	.000	-4.076	.000**	-.95	-3.425	.001**	-.87

* α significant at $p<.05$
** α significant at $p<.016$

As detected by Errors/AS-units, high proficiency learners displayed statistically more target-like linguistic behavior on task 3 than on task 1. This pattern was also present in Errors/100 words. According to this measure, high proficiency participants benefitted from increases in cognitive complexity, displaying more accuracy on task 3 than on task 2 and on task 1. In the same group measured on target-like use of prepositions, task 3 led to statistically most accurate performance compared to task 2 and compared to task 1.The results obtained by the low proficiency speakers for accuracy are presented in table 43.

Table 43. Impact of task complexity on low-proficiency learners (accuracy): Wilcoxon signed-ranks test

Dependent variable	***Low-proficiency speakers***								
	Task 1-2			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Errors/ASU	-.895	.371	-.01	-0.763	.445	.13	1.557	.119	.16
Errors/ 100words	-.339	.734	.03	-.432	.666	.12	-.607	.544	.15
TLU_prep	-.887	.375	.18	-3.254	.001**	-.68	3.322	.001**	-.85

* α significant at $p<.05$
** α significant at $p<.016$

The pattern of accuracy in low proficiency speakers was the same for target-like use of prepositions as in high proficiency speakers: task 3 led to a more accurate performance compared to task 2 and compared to task 1. Although the trend observed in the data for the global measures was towards more accurate performance as cognitive complexity increased, no significant differences were detected for Errors/AS-units and Errors/100 words.

To sum up, both groups benefitted from increases in cognitive complexity, and their speech was influenced in similar ways. However, cognitive complexity seems to have had a more marked impact in the case of high proficiency speakers, as this group was found to be significantly affected on all three measures: they produced statistically fewer Errors/100 words and Errors/AS-units, and their use of prepositions was more target-like as the cognitive demands of tasks increased. On the other hand, in the low proficiency group a significant effect was found only in the case of the specific measure. In both groups of speakers, it was in the case of the specific accuracy measure where the greatest effect of cognitive complexity was found ($p<.016$ and large effect sizes).

7.7.1.4 Structural complexity

As shown by descriptive statistics (table 39), both proficiency groups maintained a quite stable level of overall complexity across the three tasks, as measured by Words/AS-units. However, the trends found for the two proficiency groups were opposite. For the low proficiency speakers, task 3 led to the most syntactically complex output (M=9.02), whereas in the other group it was task 1 (M=9.64).

The different proficiency groups were also found to have been differentially affected by increases in cognitive complexity at the clausal level (Words/clause). High proficiency speakers complexified their speech the most on task 1 (M=7.39), but in the case of low proficiency speakers it was task 3 (M=7.43). This finding may suggest that in low proficiency speakers, increases in cognitive complexity were conducive to the production of more elaborate L2 clauses, expanding in this way the repertoire of grammatical structures.

Whereas opposite patterns were detected for the above-described measures, the same tendency in both groups was detected in the case of the amount of subordination (Clauses/AS-unit). In both groups task 2 led to the greatest amount of subordination (M=1.28 in the low proficiency speakers, and M=1.45 in the high proficiency speakers), and the most complex task prompted the lowest score for subordination for the low and high proficiency group (M=1.23 and M=1.33, respectively). According to this measure, the amount of cognitive load was therefore not proportionate to the amount of subordination generated by the speakers.

Pairwise comparisons with Wilcoxon signed-ranks tests for low and high proficiency learners are displayed in tables 44 and 45, respectively.

Table 44. Impact of task complexity on low-proficiency learners (structural complexity): Wilcoxon signed-ranks test

Dependent variable	***Low-proficiency speakers***								
	Task 1-2			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ ASU	-1.687	.092	.1	-0.905	.365	-.16	-0.119	.905	-.08
Words/ Clause	-3.060	.002**	.37	-2.232	.026*	-.41	-1.028	.304	-.07
Clause/ ASU	-2.078	.038*	-.18	-.930	.352	.26	-.514	.607	.04

* α significant at p<.05
** α significant at p<.016

Table 45. Impact of task complexity on high-proficiency learners (structural complexity): Wilcoxon signed-ranks test

Dependent variable	***High-proficiency speakers***								
	Task 1-2			**Task 2-3**			**Task 1-3**		
	Z	***p***	**Effect size**	***Z***	***p***	**Effect size**	***Z***	***p***	**Effect size**
Words/ ASU	-0.165	.869	.02	-0.435	.651	.15	-0.309	.758	.17
Words/ Clause	-2.541	.011**	.62	-.494	.622	-.43	-2.026	.043*	.03
Clause/ ASU	-3.497	.000**	-.14	-1.103	.270	.49	-1.502	.133	.09

* α significant at p<.05
** α significant at p<.016

In the measure Words/Clause, pairwise comparisons revealed that in the low proficiency group (table 44) significant differences were yielded between tasks 1 and 2 to the advantage of task 1, and between tasks 2 and 3 to the advantage of task 3. In the high proficiency group (table 45), the difference for Words/Clause turned out to be statistically significant between tasks 1 and 2, and between tasks 1 and 3, with task 1 triggering the most structural complexity.

In the case of the measure Clauses/AS-units, both proficiency groups produced the most structurally complex speech on task 2, with significant differences found between tasks 1 and 2 in low-proficiency participants, and the same pair of tasks yielded a significant difference in high-proficiency participants. No significant differences were found in the case of Words/AS-units.

To sum up, the measures of syntactic complexity revealed different ways in which cognitive complexity affected speakers of different proficiency levels, with the low proficiency speakers' production being most complex on the most complex task, and this held true for this groups' overall structural complexity level, and the clausal level of structural complexity. No such effects of increasing cognitive complexity on performance were found in the case of the other group. High proficiency speakers revealed the opposite pattern, but to the advantage of the simple task. However, in both groups the differences turned out to be statistically significant for Words/Clause, but not for Words/AS-unit. An interesting pattern observed in the case of subordination ratio is that both groups produced more subordination from task 1 to task 2, but the figures obtained for tasks 1 and 3 were strikingly similar, particularly so in the low proficiency speakers.

Both the statistically significant patterns (in the case of Words/clause) and the non-significant ones (Words/AS-units) suggest a potentially facilitative role of the complex task in promoting structurally more complex behavior in low proficiency speakers. At the same time the results demonstrated that the manipulation of cognitive complexity had opposite effects on the high proficiency speakers.

7.7.1.5 Summary of results: the role of cognitive complexity in speakers of different proficiencies in the L2

The results reported in this section described the potential role of proficiency in task-based performance. The two proficiency groups reported here showed a generally similar pattern of performance on the employed fluency measures, with the high proficiency group reducing the pace of speech delivery in a somewhat marked way compared with the low proficiency group. Regarding lexical complexity, we could observe a tendency for more lexically rich forms to emerge as cognitive task complexity increased; however, this pattern was only confirmed for D. Both groups showed clear benefits of increases in cognitive complexity on accuracy, which was evidenced in a lower error rate per AS-unit and per 100 words, and a simultaneous increase in target-like use of prepositions. However, high proficiency participants' accuracy was found to be positively affected by cognitive complexity manipulation on all measures, whereas for the low proficiency speakers this held true only for the specific measure. Finally, the findings obtained for structural complexity suggested that cognitive task complexity may affect speakers of different proficiency levels in different ways, taking into consideration the overall level of syntactic complexity and the clausal level of speech production. There was a clear effect of increases in cognitive complexity on low-proficiency speakers in that they generated more clause-internal expansion in the most complex task, but such an effect was not observed in the case of the high proficiency speakers.

Structural complexity showed the greatest variation among the two groups, and the results for the three employed measures are presented in figure 17 (Words/AS-unit), figure 18 (Words/clause), and figure 19 (Clauses/AS-units).

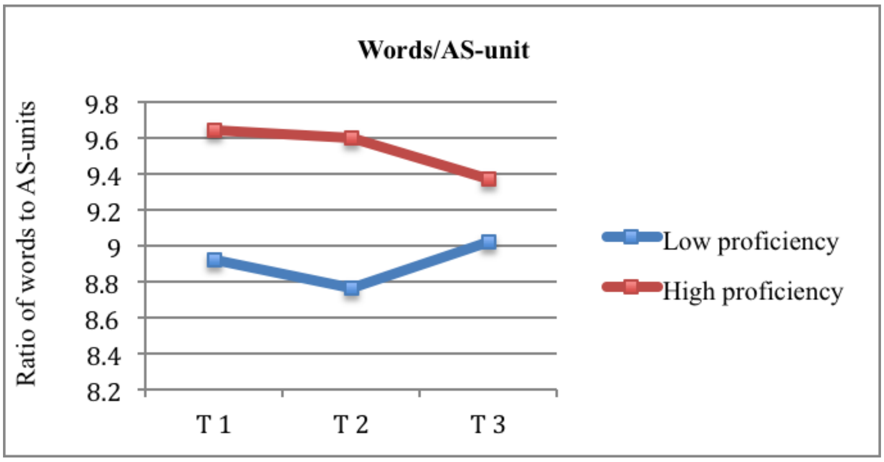


Figure 17. Words/AS-unit in low and high proficiency participants

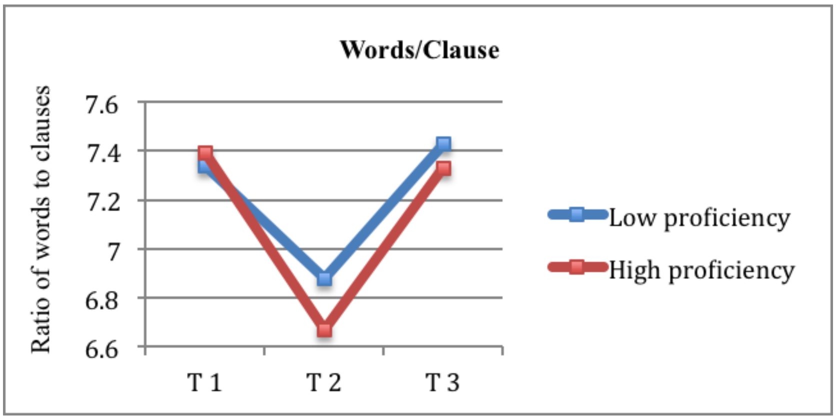


Figure 18. *Words/clause in low and high proficiency participants*

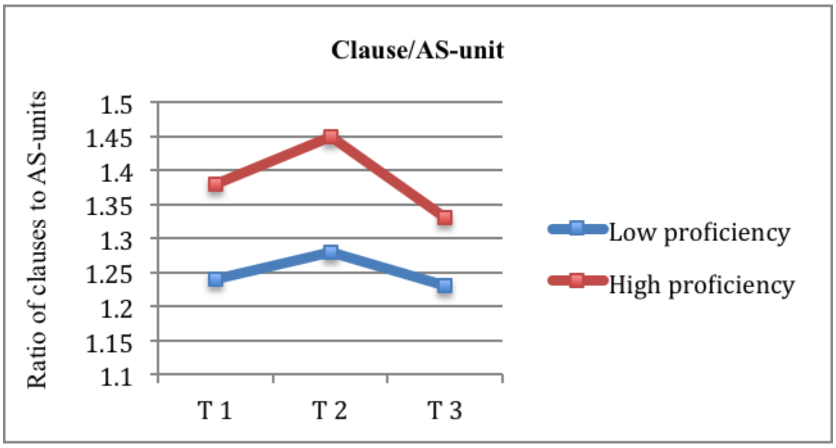


Figure 19. *Clause/AS-unit in low and high proficiency participants*

When interpreting the results obtained for the role of proficiency, it must be borne in mind that the participants involved in the current study did not represent two extremes of proficiency, but rather a continuum of proficiency levels.

7.7.2 The role of sequencing in low-proficiency participants

The last issue addressed in the current dissertation enquired about the facilitative role of the simple-complex sequence in low-proficiency speakers' performance. The research question pertinent to this exploration was formulated in the following way:

Research question 3b: Do low-proficiency learners benefit more from the simple-complex sequence than from the randomized one?

Table 46 presents the descriptive statistics obtained for this research question.

7.7.2.1 Fluency

As illustrated by descriptive statistics (table 46), in the simple-complex condition, speech rate clearly dropped as cognitive complexity increased (in Rate B, *M*=93.24 on task 1, and *M*=88.82 on task 3). Such a pattern was not detected in the randomized condition, in which the rate of speech delivery was maintained at approximately the same level across the three tasks, with the lowest score, as measured by Rate B, triggered in the case of task 1 (*M*=109.84), and the highest in task 2 (*M*=113.97). At the same time, on both Rate A and Rate B the participants in the randomized condition obtained higher scores than their counterparts in the simple-complex one. Regarding dysfluency, both groups displayed a pattern in which performance on tasks 1 and 3 was very similar (on task 3, *M*=4.19 in SC and *M*=4.34 in RAN). However, task 2 led the participants in both groups to commit numerous dysfluencies compared to the other two tasks; for this task, *M*=7.51 in SC, and *M*=7.17 in RAN. When comparing both groups' performance on the three tasks it can be observed that on tasks 1 and 2 the total number of hesitations was greater in the simple-complex condition than in the randomized one, but the opposite held true for the most complex task, in which the randomized group's performance was less fluent.

Table 46. Descriptive statistics of tasks: means and standard deviations (low-proficiency participants in the simple-complex versus randomized condition)

Dependent variable		**Low proficiency speakers**					
		Task 1		**Task 2**		**Task 3**	
		Simple-complex	***Randomized***	***Simple-complex***	***Randomized***	***Simple-complex***	***Randomized***
		M (SD)	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***
Fluency	Rate A	101.85 (21.6)	116.98 (31.09)	104.84 (18.38)	124.01 (35.71)	99.70 (15.66)	120.58 (36.02)
	Rate B	93.24 (21.15)	109.84 (29.95)	93.68 (15.61)	113.97 (33.19)	88.82 (12.38)	111.38 (35.99)
	Dysfluency	3.76 (1.20)	3.62 (1.91)	7.51 (3.05)	7.17 (5.22)	4.19 (1.51)	4.34 (2.07)
Accuracy	Errors / ASU	1.28 (0.48)	1.53 (0.45)	1.26 (0.89)	1.56 (0.60)	1.27 (0.70)	1.36 (0.50)
	Errors / 100w	14.55 (6.41)	16.30 (5.13)	13.50 (5.88)	16.97 (5.82)	13.52 (6.40)	15.38 (6.83)
	TLU_prep	55.84 (18.53)	57.02 (18.57)	63.65 (18.61)	55.70 (13.35)	67.91 (12.57)	70.07 (6.14)
Structural complexity	Words /ASU	8.58 (1.05)	9.26 (1.11)	8.67 (1.88)	8.86 (1.96)	8.78 (0.85)	9.26 (1.49)
	Words/ Clause	7.22 (1.25)	7.46 (0.85)	7.06 (1.68)	6.70 (1.05)	7.46 (1.20)	7.41 (1.38)
	Clause/ ASU	1.22 (0.26)	1.26 (0.22)	1.24 (0.14)	1.33 (0.25)	1.19 (0.16)	1.26 (0.19)
Lexical complexity	Guiraud's Index	5.01 (0.51)	5.23 (0.80)	4.72 (0.51)	5.01 (0.80)	4.76 (0.48)	5.22 (0.64)
	D	40.56 (10.58)	42.17 (12.53)	41.46 (11.43)	47.13 (12.96)	45.62 (9.60)	51.60 (14.47)

As can be observed in table 47, there was a statistically significant effect of sequence on speech rate, with the participants in the randomized condition delivering their speech significantly faster. An effect for task complexity was detected in the case of Dysfluency.

Table 47. Repeated measures ANOVA for low proficiency: fluency

Dependent variable		***F***	***Df***	***p***	**Effect size**
Rate A	Task	1.114	2	.343	.076
	Sequence	3.860	1	.059	.121
	Task*Sequence	.968	2	.646	.032
Rate B	Task	.685	2	.512	.048
	Sequence	5.002	1	.033*	.152
	Task*Sequence	.489	2	.618	.035
Dysfluency	Task	21.166	2	.000**	.611
	Sequence	.016	1	.901	.001
	Task*Sequence	.971	2	.675	.029

* α significant at $p<.05$
** α significant at $p<.016$

7.7.2.2 Lexical complexity

When comparing the performance of both groups on the two lexical complexity measures (table 46), it can be observed that the speech of the participants in the randomized condition was generally lexically richer than of those in the simple-complex sequence across tasks.

According to D, both groups produced gradually more lexically complex speech from the simple to the most complex task (M=42.17 vs. M=51.60 in RAN, and M=40.56 vs. M=45.62 in SC). As can be observed, the participants in the randomized condition benefited from increases in cognitive complexity more than the other group, which is evidenced in the difference in scores obtained on task 1 and task 3 in the two groups. When it comes to Guiraud's Index, the picture of performance was rather inconclusive, with task 1 triggering the most lexical complexity in the simple-complex condition (*M*=5.01), and a very similar lexical behavior emerged in tasks 2 and 3 (*M*=4.72 and *M*=4.76, respectively). In the randomized condition the performance on tasks 1 and 3 was virtually identical (*M*=5.23 vs. *M*=5.22, respectively), and the least lexical complexity was generated by task 2 (*M*=5.01). A repeated-measures ANOVA (table 48) showed no effect for sequencing condition, but a significant difference was detected for task complexity in the case of D.

Table 48. Repeated measures ANOVA for low proficiency: lexical complexity

Dependent variable		***F***	***Df***	***p***	**Effect size**
D	Task	9.247	2	.001**	.407
	Sequence	1.250	1	.273	.043
	Task*Sequence	.948	2	.400	.066
Guiraud's Index	Task	1.859	2	.175	.121
	Sequence	3.008	1	.094	.097
	Task*Sequence	.537	2	.591	.038

* α significant at $p<.05$
** α significant at $p<.016$

7.7.2.3 Accuracy

As displayed in table 46, the participants in the simple-complex sequence produced overall fewer non target-like forms than the other group, as measured by global performance measures. A closer look at the different measures revealed intricate patterns.

According to the figures obtained for Errors/AS-units, the participants in the simple-complex group showed remarkably similar patterns of error rate across the three tasks (*M*=1.28, *M*=1.26, and *M*=1.27 on tasks 1, 2 and 3, respectively). A greater range of scores was observed in the other group, with task 3 rendering these participants' speech most error-free (*M*=1.36) compared to task 2 (*M*=1.56) and task 1 (*M*=1.53).

A similar tendency was observed in the other global measure, Errors/100 words, with task 3 leading to the most accurate linguistic output in the randomized condition (*M*=15.38) compared to the other two tasks (*M*=16.30 in task 1 and *M*=16.97 in task 2). In the simple-complex group on the same measure there was a slight increase in accuracy from task 1 (*M*=14.55) to task 2 (*M*=13.50), with task 3 rendering marginally less accurate behavior (*M*=13.52).

The patterns detected by global accuracy measures, in which the simple-complex group displayed a more error-free behavior than the other one, were not confirmed by the specific measure. The randomized condition produced more target-like use of prepositions than the simple-complex one on task 1 (*M*=57.02 vs. *M*=55.84 in SC, respectively) and on task 3 (*M*=70.7 vs. *M*=67.91, respectively), but not on task 2; however, the reported differences were very small in magnitude. Mann-Whitney U-test (table 49) was employed to explore inferentially the impact of simple-complex versus randomized sequencing on accuracy in low-proficiency participants.

Table 49. Mann-Whitney U-test for low proficiency: accuracy

Dependent variable	**Task 1**		**Task 2**		**Task 3**	
	Simple-complex - randomized		**Simple-complex - randomized**		**Simple-complex - randomized**	
	p	**Effect size**	***p***	**Effect size**	***p***	**Effect size**
Errors/ASU	.116	-.53	.026*	-.39	.325	-.1
Errors/100words	.161	-.3	.074	-.59	.412	-.28
TLU_prep	.775	-.06	.061	.49	.967	0.21

* α significant at $p<.05$
** α significant at $p<.016$

As displayed in table 49, sequencing played a statistically significant role in performance only in the case of task 2 on the measure Errors/AS-units, the effect size being medium.

7.7.2.4 Structural complexity

The results obtained for structural complexity (table 46) show that both groups generally maintained the same level of syntactic complexity, as measured by Words/AS-units, and the randomized condition obtained higher overall scores than the simple-complex one. In the latter there was a pattern of gradual increase in complexity from task 1 to task 3 (M=8.58 vs. M=8.78, respectively), but the same did not hold true for the randomized condition, in which M=9.26 in tasks 1 and 3, with a lower score obtained in task 2 (M=8.86). Interesting patterns emerged in the case of the other two measures. The participants in the simple-complex condition produced the highest mean length of clause in the most complex task (M=7.46), whereas in the case of the randomized condition it was task 1, with the exact same mean. Both groups produced shortest clauses on task 2 (M=7.06 in SC and M=6.70 in RAN). At the same time, task 2 led to the highest index of subordination (Clauses/AS-units) in both groups (M=1.24 in SC and M=1.33 in RAN). In the simple-complex condition the least amount of subordination was produced on task 3 (M=1.19), whereas in the randomized condition the performance on the simple and the most complex task triggered the same value (M=1.26). The differences observed in descriptive statistics did not turn out to be statistically significant (table 50).

Table 50. Mann-Whitney U-test for low proficiency: structural complexity

Dependent variable	**Task 1**		**Task 2**		**Task 3**	
	Simple-complex - randomized		**Simple-complex - randomized**		**Simple-complex - randomized**	
	p	**Effect size**	***p***	**Effect size**	***p***	**Effect size**
Words/ASU	.061	-.62	.967	-.09	.486	-.39
Words/Clause	.567	-.22	.744	.25	1.0	.03
Clause/ASU	.486	-.04	.305	-.44	0.33	-.39

* α significant at p<.05
** α significant at p<.016

7.7.2.5 Summary of results: the role of sequencing from simple to complex in low-proficiency speakers

This section explored the role of simple-complex versus randomized sequencing in low-proficiency speakers. The participants in the randomized condition produced output characterized by a higher speech rate and more lexical diversity compared to the participants in the simple-complex condition. Performing tasks in the order from simple to complex proved marginally beneficial for accuracy, as measured by Errors/AS-unit, suggesting that it may potentially be an area in which gradual progression of cognitive demands from low to high is conducive to more target-like output. The picture of structural complexity is an intriguing one, with the simple task leading to the most clause-internal expansion in the randomized condition, and the same holding true for the most complex task in the simple-complex condition. However, no statistically significant differences were detected for this dimension of performance.

7.8 Summary of the chapter

This chapter presented the results obtained in the current dissertation. The objective of this section is to briefly summarize the main ideas which have been explored, followed by an overview of the principal findings.

The concepts pertinent to the current dissertation have been cognitive task complexity, task sequencing, and L2 proficiency. Each of these concepts has been explored from two perspectives. Regarding cognitive task complexity, the first approach enquired about the effect of cognitive complexity manipulation on tasks as performed in a sequence (the same speaker performing three tasks in a subsequent manner) versus tasks performed in isolation (each of the three tasks performed by a different speaker). Within the issue of sequencing, the first of the adopted approaches explored the ostensibly facilitative role of the simple-complex sequence as compared to the randomized sequence. The second one enquired about the simple, the complex, and the most complex task, as performed in a sequence versus in isolation, in the absence of any other task. Finally, regarding L2 proficiency, the general goal of the current dissertation was to research the effect of cognitive task complexity on speakers of different proficiency levels in the L2, and the specific goal was to focus on the role of simple-complex sequencing in low proficiency participants.

The results obtained in the current dissertation provide further evidence to the compelling role of cognitive task complexity in task-based performance. A major finding in this area is that the two analyzed groups (i.e., sequencing conditions and individual task performance condition) were found to be similarly affected by increases in cognitive task complexity in the dimensions of lexical complexity and accuracy, but the impact of cognitive task complexity was different in the case of structural complexity. While the range of lexical forms and target-like behavior increased in both groups as a function of increasing cognitive demands, the participants in the sequencing condition produced more structural complexity at clausal level.

Regarding the role of sequencing in production, the findings generally did not demonstrate short-term effects of sequencing of tasks from simple to complex versus sequencing them in any alternative order, except for some evidence to the contrary found in the case of accuracy. In this area, the speech of the participants in the simple-complex condition contained less deviation from norms. This finding was observed both when comparing the two sequences, and when comparing the role of simple-complex versus randomized sequencing in the low-proficiency participants. At the same time, when comparing the simple-complex sequence with individual tasks (in the case of the complex and the most complex task), there was a very clear effect of sequence versus isolated task performance, with the sequence triggering a faster speech rate and more accurate linguistic behavior, and a simultaneous increase in repair fluency.

Two major findings regarding the role of L2 proficiency were that: 1) increasing task demands yielded benefits in the area of accuracy in both low and high proficiency speakers, with the high proficiency speakers surprisingly being somewhat more affected by increases in complexity than the other group, and 2) the two groups displayed opposite trends in performance in the area of structural complexity in that there was a clear effect of increases in cognitive complexity on low-proficiency speakers who generated more clause-internal expansion in the most complex task, but such effect was not observed in the case of the high proficiency speakers. This last finding suggests that cognitive task complexity may affect speakers of different proficiency levels in different ways in terms of structural complexity.

CHAPTER 8
DISCUSSION

8.1 Introduction

The objective of this chapter is to provide interpretations to the results obtained in this study. The discussion of findings related to the first two research questions, apart from making reference to the two task complexity models this work draws on, the Cognition Hypothesis and the Trade-off Hypothesis, is juxtaposed with the results obtained in TBLT research on cognitive task complexity. A critical view of findings in this study and beyond is offered in an attempt to understand how the different dimensions of performance behave under simple and complex task conditions.

The discussion of research questions enquiring about the role of simple-complex sequencing focuses primarily on how the findings obtained in the current dissertation inform the task sequencing proposal this work draws on, the SSARC model of pedagogic task sequencing. This perspective is considered in relation with pertinent cognitive, psycholinguistic, and performance-based accounts.

Finally, the role of proficiency is discussed mainly from the point of view of attention allocation during task performance, and speech production models. A comparison with other studies, which set out to investigate performance, is followed by an evaluation of the role of proficiency as a mediating variable in second language speech production, focusing on the advantages speakers of different proficiency levels seem to exhibit when exposed to tasks placing different levels of cognitive demands. The discussion of all research questions is followed by theoretical, methodological, and pedagogical implications of the current dissertation.

8.2 Research questions 1a and 1b: the effect of cognitive task complexity on performance: sequencing conditions and baseline data

The current dissertation investigated the effect of cognitive task complexity on performance from two angles. The first perspective enquired about the effect of cognitive task complexity on performance understood as performing three tasks in a sequence (same speaker). The other investigated the effect of complexity when tasks were performed in isolation (different speakers). In the Results chapter, the effect of cognitive complexity on performance was treated as two independent research questions, one for the sequencing group, and the other one for the baseline group. In the chapter at hand, the results obtained from these two groups are collapsed into a single analysis. I believe such organization makes it possible to obtain a comprehensive picture of patterns.

Let us commence by summarizing the findings obtained for both research questions. The results broadly revealed that cognitive task complexity:

1) in the case of both groups *promoted greater accuracy* and *forced the participants to stretch their lexical repertoire* to meet the demands of the most complex task;
2) led to intricate patterns in structural complexity, with *baseline and sequencing group displaying different types of structural complexification*;
3) triggered *different fluency behaviors in baseline and sequencing groups*, with one group characterized by a stable behavior with a slight tendency towards a decrease in the most complex task (sequencing conditions), and more of a variation between tasks in the case of baseline data.

From a psycholinguistic perspective, as well as from the attention allocation perspective, it is noteworthy that an increase in accuracy levels (particularly so in the case of the specific measure) was accompanied by more elaborate lexis as cognitive task complexity increased. A pattern of linear progression in these two dimensions was observed in both sequencing and baseline conditions. Regarding accuracy, such a finding in the sequencing condition confirms the prediction of the Cognition Hypothesis, that faced with cognitively more demanding input, speakers stretch their linguistic resources, the final product being a message delivered in a more target-like manner. In functionalist terms, increasing cognitive task load caused a shift from the pragmatic to the syntactic mode and performance was pushed beyond the "basic learner variety" (Klein & Perdue, 1992, 1997, in Robinson, 2003, p. 62). Complex conceptualization associated with complex tasks triggered the learners to stretch their linguistic resources. When dealing with a sequence of tasks, the exposure to the simple and the complex task prior to the most complex one, served the function of consolidating previous knowledge (with its familiar target like and non-target like forms), alleviating therefore the cognitive demands imposed by the most complex task.

However, it must be borne in mind that when analyzing the effect of cognitive complexity in the current dissertation, the tasks were administered in different orders. For instance, in the simple-complex sequence the speakers performed the most complex task as the last one, while for others it was performed in all the possible positions (randomized sequencing), and finally, in the baseline condition it was performed in isolation from the other tasks. While performed as the last one in a sequence, it may indeed be expected to trigger more monitoring, it is an intriguing question why such a scenario should hold true for any alternative order, or a task performed in the absence of prior input, and output.

Less accurate linguistic behavior than the one revealed by the findings could be expected in the individual task performance condition, given that these participants did not have available the scaffolding mechanism in the form of prior exposure to the simple and the complex tasks before engaging in the most complex task. Analogous lines of accuracy behavior in the two groups suggest that cognitive task complexity exerted a profound impact on linguistic performance and, in the case of accuracy, was conducive to more error-free production, regardless of whether previous exposure took place or not, or how intense this exposure was (i.e., how many tasks were performed prior to the most complex task). Enhanced complexity did not, therefore, have an adverse effect on the channeling of attention resources to controlling their output. On the contrary, when task demands were high, they encouraged a more efficient control over the quality of one's performance.

This finding leads to a possible conclusion that monitoring one's output in a task designed along resource-directing dimensions is a self-governing, independent mechanism. From the speech production point of view, it can be viewed as an "inherent feature of the perception and production processes" (Kormos, 2006, p. 30), and in the findings obtained in the current experiment it got activated regardless of the availability or intensity of prior exposure to tasks of different cognitive complexity levels. Following Gilabert (2007c), "when task demands were increased along resource-directing variables learners' attention was drawn to the way they were encoding their messages, with positive effects on accuracy" (p. 236).

Turning to the other dimensions of performance, in addition to close monitoring of one's linguistic output, the complex task condition led to the production of lexically more varied speech. This is evidenced in the fact that independently from the presence or absence of task-based performance prior to performing the most complex task, both groups employed the greatest variety of lexical forms in the most complex task. A simultaneous increase in these dimensions – accuracy and lexical complexity

– means that learners had sufficient attentional resources available in order to control their output, and concurrently retrieve from the mental lexicon items of greater diversity on the most complex task.

The findings obtained for accuracy and lexical complexity are in line with previous research, which investigated the effects of resource-directing dimensions on performance, both when these variables were manipulated in separation and simultaneously (e.g., Kuiken et al., 2005; Michel et al., 2007; Révész, 2011; Malicka & Levkina, 2012; Gilabert, 2007c). In these studies, which explored the effect of resource-directing variables on production, it was found that increases in the degree of conformity to language norms were proportionate to increases in cognitive task complexity, suggesting a facilitative role of increased demands in promoting target-like output. What is noteworthy about previous findings is that, although statistically significant differences were not always found for accuracy, in the vast majority of studies a trend was detected towards more accuracy as cognitive task complexity increased. A recent research synthesis and meta-analysis of task complexity studies (Jackson, 2013) revealed a positive effect size for accuracy (d=0.28) on the basis of 39 contrasts from 9 publications. Such was the case in the three scenarios identified in literature (i.e., studies which manipulated the variable ±elements, studies which manipulated the variable ±reasoning demands, and studies which manipulated both of these variables within a single task design).

The picture obtained for lexical complexity across studies was not as consistent as the one obtained for accuracy, but there was a general tendency towards more lexical complexity as task demands increased. However, several studies yielded the opposite result: a greater variety of lexis was observed in the case of the simple task. In the current dissertation, in the section at hand and throughout the results, generally different patterns were found for the two measures employed – Guiraud's Index and D. The latter systematically exhibited a consistent trend towards greater complexity as a function of increasing task demands, and the former showed quite incoherent patterns of behavior (i.e., a drop, an increase, or overall same or comparable lexical complexity levels across tasks). Typically CAF studies investigating the impact of these variables (that is, elements and reasoning demands) on performance have used only one measure of lexical complexity, or two measures, which captured the same phenomenon, but at different stringency levels (e.g., type-token ratio and type-token ratio corrected for text length). The current study used two different measures, and to my knowledge the measure D was previously used in only two studies which explored the impact of ±elements and ±reasoning demands on performance. One study yielded no significant differences (Kormos & Trebits, 2012), and another found a significant difference in favor of the complex task condition (Révész, 2011). However, to my knowledge, Guiraud's Index and D have not been used jointly in the same study.

Considering the nature of these two measures, it can be concluded that D turned out to be a more sensitive measure in the case of the data set in the current dissertation. In Jarvis' (2013) review of available lexical complexity measures, Guiraud's Index was mentioned alongside other measures which supposedly correct for text length (e.g., Herdan's Index or Uber Index), but they were criticized because "none of these transformations of TTR adjusts accurately for text length" (Jarvis, 2002, p. 59), the conclusion being that all these measures, including Guiraud's Index, only partially overcome the problem of sample length (Jarvis, 2013). By contrast, the measure D "represents the shape and position of the entire TTR curve" (Jarvis, 2002, p. 59), and consequently, it is predicted to eliminate the problem of sensitivity to text length. It is on these grounds that the measure D can be speculated to have been a more reliable predictor of lexical complexity in the current study. In addition, as could be observed in the previous chapter, and will be pointed out on several occasions in this chapter, the measure D showed a stable, systematic pattern of performance (generally an increase in lexical complexity being equally proportionate to an increase in cognitive task complexity), whereas the results for Guiraud's Index have been mixed, and therefore likely less trustworthy.

Returning to the discussed dimensions of performance, the clear line of behaviors demonstrated in accuracy and lexical complexity stands in stark contrast to that detected in structural complexity, which presented the most ambiguous behavior of all involved dimensions. Across tasks, the participants in the baseline condition were found to complexify their speech at the level of subordination and decomplexify it at the clausal level, with the results obtained for both measures visually representing straight lines. The participants in the sequencing group displayed the opposite pattern: a V-shape behavior in the case of clausal complexity (words/clause) and a reverse V-shape behavior in the case of subordination (clauses/AS-unit). Opposite patterns can therefore be observed, suggesting that structural complexity may potentially be more responsive to the amount of previous task input and its intensity, than accuracy and lexical complexity. At first sight such performance patterns may be reflective of the participants' current L2 competence level, insofar as generating a higher subordination ratio, as opposed to longer clauses, is characteristic of less advanced L2 speakers. However, as was previously mentioned, there were no a priori differences in proficiency between the different groups involved in the study (in the case of the discussed research question, on the proficiency measure, *M*=32.32 for the sequencing group and *M*=31.6 for the baseline group), which calls for alternative explanations. In order to explore them, let us present the results obtained for structural complexity in light of previous work.

The lack of clear patterns of findings in the case of structural complexity is not uncommon in those CAF studies which manipulated resource-directing factors. Across the literature, structural complexity has presented intricate, often times inconsistent patterns, compared to the other dimensions. In the literature pertinent to the current dissertation and reviewed in the literature review chapter, the effect of individually and jointly manipulating the variables ±elements and ±reasoning demands generated from insubstantial to very limited effects on structural complexity. More specifically, the studies pertinent to the current dissertation revealed four scenarios: 1) negligible descriptive differences between task complexity levels *in favor of the complex task condition*, as opposed to negligible descriptive differences between task complexity levels *in favor of the simple task condition*, 2) a statistically significant difference *in favor of the simple task condition*, and 3) significant differences, both descriptive and inferential*, in favor of the complex writing tasks* as compared to their simple counterparts. A noteworthy insight from several studies investigating elements and reasoning demands is that when elements were manipulated simultaneously with providing planning time, a significant effect was found for the "many elements, +planning time" condition (e.g., Ishikawa, 2008).

In my view, there is a series of accounts offering explanations to the question of the picture of results yielded for structural complexity, both in this study and in the literature in general: the multifaceted nature of structural complexity, its lack of straightforwardness as a construct, and the multiple ways in which it has been measured across studies (Housen & Bulté, 2012) or individual differences mediating performance (such as L2 proficiency). It could also be speculated that structural complexity is a longer-term, more profound kind of developmental phenomenon that does not change qualitatively as a result of engaging in a single task, but requires substantial time and effort to change in the interlanguage of the learner. Given that there are different levels at which speech can get complexified in a learner's output, and the fact that specific types of complexification are associated with different stages of competence in the L2 (so individual differences come into play), structural complexity is the least linear construct, which may be the justification for the a relative arbitrariness revealed by study findings found across literature. However, it must also be borne in mind that both this current dissertation and most previous research explored structural complexity from the point of view of grammatical *complexity* (e.g., subordination ratio, clausal complexity, and overall complexity). Recent conceptual research on measures (Housen & Bulté, 2012) advocates, in addition to these measures, exploring syntactic complexity from the point of view of grammatical *variety*, which can manifest itself for example in the number of verbs forms used, the range of tensed verbs, or the number of infinitival constructions. Employing the latter category of measures which focus on diversity, in addition to the already extensively used complexity measures, could be potentially illuminative of the nature of syntactic complexity and shed a new light on this multifaceted construct.

However, these claims are only partially applicable to the extensive array of studies carried out which measured the effect of planning time (a resource-dispersing variable) on structural complexity. The studies which manipulated pre-task planning time systematically found an effect on performance: when time to plan the message was available, speakers produced structurally more complex output, and this is evidenced for example in the work of Crookes (1989), Wigglesworth (1997a), Skehan and Foster (1997), Foster and Skehan (1996), Ortega (1999), Yuan and Ellis (2003), Ellis and Yuan (2004), Mehnert (1998), Mochizuki and Ortega (2008), Taguchi (2007), and Meraji (2011).[30] At the same time, in the planning studies mentioned above, the findings obtained in the area of accuracy have been largely inconsistent, unlike in those studies which manipulated resource-directing dimensions. In the experiment reported in this dissertation planning time was not provided (and therefore the tasks may be described as complex on a resource-dispersing dimension), and had planning time been made available, structural complexity would have possibly revealed a more stable pattern.

This state of affairs quite clearly raises the possibility that different types of dimensions (i.e., resource-directing vs. resource-dispersing) are responsible for generating different types of transformations in the learners' interlanguage. In other words, rather than having analogous influence regardless of the kind of task design manipulation, there seem to be quite strong associations between the type of variable (resource-directing or dispersing) and the sort of qualitative change it brings about. While this is not to say that each variable has a unique effect on performance, it becomes increasingly transparent in research findings that planning time affects structural complexity and that heavy resource-directing demands affect accuracy. In light of this, I believe that the results obtained for structural complexity in the current study further reiterate the tendency observed in previous research, that the manipulation of the two variables in question - elements and reasoning demands - is overall not conducive to more structurally complex output. It is probably a far-fetched scenario to expect that all four dimensions of production (fluency, accuracy, lexical complexity, and structural complexity), will be equally responsive to every kind of manipulation. For instance it is debatable why engaging in narrating a story in the past

[30] It must be borne in mind, however, that not all planning time studies showed a trend towards more structural complexity, and therefore we can talk at best about a strong trend rather than common knowledge.

as opposed to the present (so, here-and-now versus there-and-then) should trigger the occurrence of more complex syntactic forms *and* greater lexical complexity. Unless the more cognitively complex task condition is designed to direct learner's attention to specific lexical forms, a qualitative change beyond past time reference is speculative.

As was the case with structural complexity, fluency (more specifically, speech rate) also revealed an interesting pattern. While the participants in the sequencing condition exhibited a linear behavior, in the baseline condition more variation could be observed between the different tasks. Roughly the same speed of delivery was detected in the simple and the most complex task, but the speech was faster in the task of medium cognitive complexity level. This finding presents an interesting contrast to the linear pattern observed in the sequencing condition. Comparing the two groups it can be speculated that performing a series of tasks, versus performing them in isolation, is in itself a condition which helps to maintain an overall stable fluency level, with only minimal variation across tasks. This issue will be further explored in the next section which focuses specifically on the comparison between the baseline condition and the simple-complex condition. Despite the different patterns, or the different routes taken by the speakers in both conditions, a fact worth emphasizing is that, particularly in the sequencing condition, the participants maintained an overall *stable speech rate,* rather than displayed *a decrease in speech rate*. While it is true that the most complex task triggered slower output than the other two tasks, the differences can be described as marginal.

The nature of the results obtained for speech rate is probably better understood if the other fluency measure, Dysfluency ratio, is analyzed. In both conditions (sequencing and baseline data) it could be observed that the rate of speech delivery was fastest in the case of the complex task. The same task triggered a substantial amount of repair behavior in both conditions, but the mean value obtained by the speakers in the sequencing condition was double of that obtained in the baseline condition. It is particularly true for the complex task, but a similar trend was also displayed in the other two tasks. One explanation for this finding could be that the number of dysfluencies is proportionate to speech rate: speaking faster takes place at the expense of a build-up of dysfluency phenomena: repetitions, false starts, and repairs. In fact, repairs have been shown to increase significantly in more complex tasks (Gilabert, 2007c). Conceived of in this way, the findings suggest a possible trade-off effect between speech rate and repair fluency, with an increase in one of them occurring concurrently with a decrease in the other.

Contextualizing the findings obtained in the current dissertation in light of previous research, the results for speech rate are largely in line with what the literature has shown. The two most common scenarios regarding speed fluency behavior encountered in literature have been that fluency either dropped as the result of increases in cognitive complexity, or it remained virtually intact, with only a marginal trend in favor of the simple task condition (however, some studies also found increased speech rate in the complex task condition). A quite modest effect of cognitive complexity on speech rate is also evidenced in the general lack of significant differences detected for this dimension across pertinent studies. A few aspects of these studies deserve close attention. A significant difference in favor of the simple task was yielded in one study (Robinson, 2001), but it employed words per c-unit as a speed fluency measure, whereas in later studies researchers employed unpruned and pruned speech rate instead (syllables per minute or second). In Levkina and Gilabert (2012), a significant difference was detected in favor of the simple task condition, but under the condition where the resource-directing variable (number of elements) was manipulated simultaneously with a resource-dispersing variable (planning time). It is therefore possible that the resource-dispersing dimension was the one responsible for the observed quantitative change. Finally, in the study by Malicka and Levkina (2012), the effect of manipulating cognitive complexity on performance was detected in the reverse direction to that predicted by the Cognition Hypotheses: the complex task led to more fluent speech.

Another possibility is that speech rate may reach a threshold beyond which speech is simply not delivered faster. It could be related to proficiency level: while the speech of learners at a low level of L2 competence is not automatized, and therefore the speed of delivery is typically slow (so, dysfluent), at higher levels of proficiency in a second language speakers maintain an overall stable rate of speech delivery, rather than a marked increase.

Regarding dysfluency ratio, in the literature relevant to the current dissertation, this measure was used only on one occasion (Ishikawa, 2008), and a similar pattern to the one observed in the current study was found: as cognitive complexity of tasks increased, so did the amount of dysfluency. It must be recalled that dysfluency ratio was a multifaceted construct in this study and it included a range of phenomena, including self-repairs, operationalized as rectifying lexical and morphosyntactic errors. This means that in the current study dysfluency ratio captured what some researchers would, in fact, claim to be a measure of self-repair behavior, and it can therefore be understood as instances of focus on form, rather than fluency.

When looking at the different dimensions of performance from a global perspective, it can be broadly concluded that both groups – sequencing and baseline conditions - were influenced by increases in cognitive task complexity. In the sequencing condition, producing output at a stable rate was concurrent with an increase in the level of compliance with norms, and greater lexical complexity. In the baseline condition, delivering the message in what could be described as a rather unstable speech rate, took place simultaneously with greater accuracy, and higher lexical complexity.

Given that opposite and quite unclear patterns were detected in the case of structural complexity, one could speculate that there was a trade-off effect between this dimension of performance and one of the two other dimensions, either accuracy or lexical complexity, as posited in the Trade-off Hypothesis. What merits attention however, is that while that attention was clearly geared towards accuracy in more complex tasks, syntactic complexity did not increase, but *it did not decrease either*; rather, it was maintained at approximately the same level. It is also worth reminding that in the current dissertation pre-task planning time was not made available to the speakers, which rendered the research tasks investigated in this experiment *complex on the resource-dispersing dimension*, planning time. The Cognition Hypothesis posits that tasks be simple along these dimensions. Therefore clearer lines of behavior in the case of structural complexity (e.g., a possible increase), could be expected to emerge had planning time been provided. However, the different scenarios discussed above, which speculated about a relative insensitivity of structural complexity to the manipulation of resource-directing factors, as opposed to its clear responsiveness to the manipulation of resource-dispersing factors, casts a doubt on a trade-off effect as a viable explanation to the results obtained for structural complexity. I am more inclined towards considering the results of structural complexity, both in this study and in similar research, as an effect of a lack of an adequate stimulus necessary for structural complexity to be altered from one task to another, or it is simply not so responsive to minor task design manipulations. As was mentioned before, in planning time studies structural complexity was systematically shown to be affected by the manipulation of this resource-dispersing factor. On the basis of previous research, I would therefore suggest that it is only reasonable that structural complexity 1) was affected in both conditions in different ways, and 2) that the manipulation of cognitive task complexity did not reveal clear patterns. These plausible explanations notwithstanding, it cannot be discarded that the measures employed thus far in the case of structural complexity, both in this dissertation and related research, have simply not been sensitive enough to capture qualitative differences between tasks of different cognitive complexity levels. Given unclear patterns, it may be the case that structural complexity could be more precisely measured using specific rather than global performance measures, theory-driven on one hand (possibly based on the emergence of different forms in the L1) and task-relevant on the other hand (measuring linguistic phenomena identified by the researcher as relevant to task performance). Effort should be made to identify those features of tasks without which task transaction is incomplete or considerably compromised, and use these features as devices to capture interlanguage change.

To conclude this section, the results obtained were confirmatory of previous research, which investigated the impact of the variables ±reasoning demands and ±elements on production. The findings for research questions 1a and 1b added support to the claim that as cognitive task complexity increases, so does the lexical complexity and accuracy, with a concurrent unclear picture of structural complexity. I believe that these results, and the results obtained across the different studies on cognitive task complexity, rather than providing evidence to any of the hypotheses this work draws on, call for a potential fine-tuning or revisiting of certain claims of the hypotheses, which would render them more complete and precise

8.3 Research question 2a: the effect of task sequencing on performance: simple-complex versus randomized sequencing

The second research question enquired about the ostensibly facilitative role of the simple-complex sequence on performance, as compared to any other alternative order of performing a sequence of pedagogic tasks. In order to better understand the findings obtained for this research question and the next one, which enquired about tasks performed in a sequence versus in isolation, let me recall the model these two research questions draw upon.

The exploration of the issue of sequencing was inspired by the claims of the SSARC model of pedagogic task sequencing. This model is based on, and further reinforces, the fundamental claim of the Cognition Hypothesis, which states that pedagogic tasks should be sequenced in an order of increasing cognitive complexity, as this promotes L2 development and improvements in the ability to perform target tasks in the L2 (Robinson, 2007). An important advancement the SSARC model presents over early models of sequencing, and probably its most unique component, is that it assigns specific functions to tasks of different cognitive complexity levels: in simple tasks performance is based on the simple state of the interlanguage, slightly more complex tasks render the performance more automatized, and still more complex tasks lead to the destabilization and restructuring of the current interlanguage system (Robinson, 2010).

The current dissertation set out to test these claims by having a group of participants perform pedagogic tasks in the simple-complex sequence, following the premises of the SSARC model, and having the other group of participants – the randomized sequencing condition - perform the tasks in all alternative orders in which three tasks can be sequenced. Four major observations for this research question detected in the findings were:

1) the participants in *the randomized sequence delivered their tasks faster*, *with more lexical complexity, and with somewhat less dysfluent behavior*, than their counterparts in the simple-complex sequence;
2) the participants in the *simple-complex sequence displayed overall greater accuracy* than those in the randomized sequence;
3) both groups displayed an *overall trend towards greater accuracy and lexical complexity*, and a minor drop in fluency, as cognitive task load increased;
4) in structural complexity, a pattern of *a trade-off effect was observed between average clause length and subordination ratio*.

The major conclusion from the findings summarized above is that, depending on the dimension of language production, it was either the simple-complex condition or the randomized condition, which presented relative benefits for performance. Given such a picture of findings, this section is divided into two parts, the first dealing with benefits encountered in the simple-complex condition and the randomized one, followed by the patterns identified in each sequence. These findings are explained in relation to the SSARC model of pedagogic task sequencing, speech production and attention allocation models, as well as learner factors (i.e., affective variables questionnaire). The section concludes with an attempt to evaluate the potential role of simple-complex sequencing in promoting optimal output.

As was briefly advanced at the beginning of this section, the participants in the simple-complex condition exhibited an overall more target-like output than the speakers in the other group. More specifically, despite the trend towards more quality output in both conditions as cognitive complexity increased, on all three tasks the participants in the simple-complex sequence showed a lower error rate on both general measures, and more target-like behavior in terms of the specific measure, target-like use of prepositions. At the same time, these participants performed their tasks slightly slower than the other group, and also they displayed not as wide a range of lexis.

The participants in the simple-complex sequencing condition, prior to engaging in the most complex task, received input and produced output in the form of two tasks of cognitively less complex levels than the most complex one. It may be the case that previous exposure to these tasks proved facilitative to performance because it clearly activated the mechanism of monitoring in these speakers. Simple-complex sequencing perhaps triggered a greater allocation of attentional resources to the area of accuracy. Augmented complexity drew these speakers' attention to monitoring their speech and controlling the quality of their output, and as a result, less deviation from norms took place. From the speech production perspective, it can be speculated that in these speakers, the most complex task did not require as great attention devoted to planning their speech (the stage of message conceptualization) as for example in the case of the simple task, and consequently more attention could be devoted to the accuracy of message delivery. On the basis of these results, it can be stated tentatively that simple-complex sequencing led to more target-like output, which suggests a potential role of this type of sequencing in promoting more error-free behavior, compared to alternative orders of performing tasks.

It needs to be taken into account, however, that in the case of the specific accuracy measure, target-like use of prepositions, the difference obtained between the participants in the two conditions on the most complex task was only marginally in favor of the simple-complex sequencing condition (*M*=76.39 in SC and *M*=75.64 in RAN), with substantial differences to the advantage of the SC sequencing observed in the other tasks, the simple and the complex one. To my mind, this finding highlights the role of cognitive task complexity in generating qualitative changes in a speaker's output, in this case manifesting itself as a factor contributing to more error-free linguistic behavior.

A further observation about accuracy is that the participants in the simple-complex sequence displayed a more linear pattern of behavior than their counterparts in the randomized condition. Given that increases in accuracy levels were proportionate to increases in cognitive task complexity, from a psycholinguistic perspective it may be the case that simple-complex sequencing triggered changes in self-monitoring behavior, which were of a more gradual or systematic nature. Gentle yet progressive increases in cognitive task load pushed the learners to producing qualitatively more target-like output in a step-by-step fashion, with a moderate but clear tendency towards less deviation from norms as they were faced with heavier cognitive demands. The participants in the other group displayed a markedly more non-linear behavior, which was particularly prominent in the case of target-like use of prepositions. The different points in these speakers' performance do not visually represent a straight line, as is the case in simple-complex sequencing. It may be the case that when tasks are performed in the order of increasing complexity (so, from low to high task demands), such sequencing promotes optimal distribution of resources; on the other hand, sequencing tasks in any alternative order triggered worse coordination of resources. As was mentioned at the beginning of this section, the heightened attention to language forms in the case of the participants in the simple-complex condition was accompanied by a lower speech rate and less lexical complexity compared to their counterparts in the other condition.

The potential advantages for the simple-complex condition in the area of accuracy described above are contrasted with the findings obtained for the randomized sequence: the speakers in the latter condition were found to deliver their speech faster, with more lexical complexity, with a slightly lower dysfluency ratio than the speakers in the simple-complex sequence. Let us first explore the different results obtained by the two groups from the point of view of attention allocation.

The fact that the randomized sequencing participants displayed higher speech rate and more lexis, and the participants in the other group showed a greater trend towards accuracy, may suggest that different orders in which tasks were performed may have led to different patterns of attention allocation. Sequencing tasks from simple to complex, by posing minor yet detectable increases in cognitive load, may have led the speakers to channeling their attention towards their output's quality. At the same time, the speakers in the other condition might have prioritized the other two dimensions of performance. In other words, the attentional resources devoted to delivering their speech in a timely manner and with a greater variety of lexis took the attention slightly away from monitoring, with a possible trade-off between the aforementioned meaning and form. Interpreted in this way, the findings obtained violate the core prediction of the Cognition Hypothesis, that simultaneous attention to accuracy and complexity occurs as a result of engagement in tasks of increasingly higher levels of cognitive complexity.

Up until now the discussion of findings for the simple-complex versus randomized sequencing has focused on the comparison of the relative advantages of one type of sequencing over the other. Throughout this section I have claimed that simple-complex sequencing led to greater accuracy, whereas randomized sequencing triggered greater fluency and lexical complexity. Let us now analyze the trends encountered *within* each of the sequencing conditions, which will shed new light on the findings.

In both simple-complex and randomized sequencing conditions, the participants were found to stretch their interlanguage to meet the demands of the most cognitively complex task irrespective of its location within a sequence. In both conditions, *in response to the most complex task, there was a tendency towards: 1) more lexical complexity and more accuracy, and 2) less fluency*. This pattern implies that a task characterized by a high internal cognitive load predetermined certain linguistic behaviors and, in the case of accuracy and lexical complexity, it forced the learners to mobilize their available linguistic resources to successfully complete the most complex task. The fact that speakers in both conditions attended to both form and meaning is indicative of their optimal distribution of attentional resources and control over different aspects of producing a message. These findings are largely in line with the predictions of the Cognition Hypothesis: simultaneous attention was paid to complexity (at the lexical level) and accuracy, to the detriment of fluency.

The picture of performance would not be complete without the patterns observed in structural complexity. As cognitive task complexity increased, the speakers were found to reach a threshold in subordination ratio beyond which they did not complexify their speech any more: in the simple-complex condition the speakers peaked in task 3, after which a decrease in structural complexity was observed, and the other group generated the highest structural complexity in task 2. The order of task performance also led to different patterns emerging in terms of the mean length of clause. I consider a performance-based perspective, as well as an individual differences perspective, as relevant to this finding.

From the task performance perspective, it could be argued that in the simple-complex condition prior exposure to the simple and the complex task served the function of a preparatory mechanism to complexifying this group's output at a deeper level: they produced more elaborate clauses on the most complex task. It can be stated that these participants' generation of output in tasks of a lower cognitive complexity level made these learners stretch their syntactic resources in the most complex task compared to the simple and the complex one. This is particularly so if this group's performance is juxtaposed with that of the participants in the randomized condition: in these speakers it was the simple task which triggered the most clause-internal expansion. Simple-complex sequencing could therefore be responsible for the generation of longer clauses.

These differences notwithstanding, both groups displayed a similar pattern of linguistic behavior on the most complex task, in which a trade-off effect was observed: greater mean length of clause took place concurrently with a decrease in subordination ratio. The most complex task therefore prompted the speakers to complexify their output at a more advanced clausal level, with a simultaneous detriment to subordination. This trend was particularly marked in the simple-complex condition, but to a lesser extent it could also be observed in the randomized sequence.

To sum up, irrespective of the position of the most complex task within a sequence, it still triggered the speakers to push the boundaries of their interlanguage to the most accurate production (according to three measures) and lexically most complex production (according to one measure). The picture obtained for structural complexity is an intricate one, with the most complex task leading to a high-level complexification, typically associated with more competence in the second language: as speakers move towards more advanced stages of L2 competence, they move from a basic syntactic complexification (subordination) to a more elaborate complexification manifested by expanding the syntax at the level of clause, for example through pre- and post-modification.

What do these findings reveal about the functions of tasks of different cognitive complexity levels, as postulated by the SSARC model of pedagogic task sequencing? The stabilization function of simple tasks and automatization function of complex tasks may not be necessary prerequisites for the subsequent restructuring or complexification of one's output. To my mind the results obtained in this section of the dissertation show that a second language learner is able to retrieve from their resources, and to successfully incorporate in their production, advanced L2 forms, at the same time as monitoring their output to increasingly higher standards. It therefore seems that the restructuring/complexifying function of very complex tasks applies even when the order of performing the tasks is any variation of the simple-complex sequence. Considered from this angle, the findings possibly undermine the idea that it is in response to a specific sequence that qualitative changes are brought about.

Given the results obtained for all dimensions and both sequencing conditions, a tentative conclusion is that the qualitative changes exhibited by both groups were primarily due to the deliberate manipulation of task-internal cognitive load (so, the independent variable of cognitive task complexity), rather than in response to the order in which the participants were required to deal with the cognitive load (so, the independent variable of task sequencing). However, the picture of task sequencing would not be complete without an inspection of the patterns observed when performing a simple-complex sequence is compared with performing a single, isolated task. This issue is the object of the next section. Therefore, before going on to the implications, which can be drawn from the current piece of research about task sequencing, let us explore and interpret the patterns observed in the next research question.

8.4 Research question 2b: Does performing a task in a sequence lead to qualitatively different output compared to performing a task in isolation?

The second perspective on task sequencing, which this dissertation aimed to explore, addressed the potential qualitative changes in production when a task was performed in a simple-complex sequence versus in the absence of any other task. The role of sequencing was investigated by having the same speaker perform a sequence of three tasks of differing cognitive complexity levels on the one hand, and on the other hand by having each of these three tasks performed by different speakers. Major findings pertinent to the comparison of the tasks performed in a sequence and in isolation were the following:

1) sequencing tasks *from simple to complex led to enhanced accuracy* (task 1 - all measures and task 2 - target-like use of prepositions), *to a higher speech rate* (tasks 1 and 3), and to *a greater structural complexity in terms of Words/clause* (task 3);
2) isolated task performance led to *a lower dysfluency ratio* (tasks 1, 2, and 3), and to *greater structural complexity in terms of Clauses/AS-units* (task 3);
3) the two conditions exhibited *different patterns in structural complexity*;
4) no differences between the two conditions were detected in the case of lexical complexity.

Overall, two important implications from the findings obtained for the research question at hand are that 1) the participants' speech was influenced differently by cognitive complexity when it was performed as part of a simple-complex sequence and when performed in isolation, and 2) the nature of this influence, rather than being systematic across tasks, changed from task to task. Looking at the patterns of results, it can be stated that the greatest influence of cognitive complexity was observed in the case of task 1 (all fluency and accuracy measures) and task 3 (all fluency measures and two structural complexity measures). By comparison, in task 2 only two significant differences were found (dysfluency ratio and target-like use of prepositions). Focusing specifically on the most complex task performed in the simple-complex sequence, as opposed to the baseline condition, it led to a higher speech rate, greater dysfluency ratio, and greater structural complexity understood as words per clause. It can therefore be stated that performing a sequence of tasks, compared to performing tasks in isolation, exerted an influence on certain aspects of two dimensions of performance (fluency and structural complexity). In what follows, I will present how the different dimensions of performance were specifically affected under the two conditions.

Regarding structural complexity, the speakers in each condition complexified their speech at different levels. While in the baseline condition subordination ratio increased and words/clause decreased, the exact opposite pattern could be observed in the simple-complex condition. These speakers produced longer clauses, with a simultaneous decrease in clauses/AS-units. In the most complex task, significant differences were yielded between the two groups on both measures.

The emergence of such a pattern of performance is surprising insofar as complexifying at the level of clause, rather than at the subordination level, is characteristic of greater proficiency in the L2: as interlanguage develops, in terms of syntactic complexity learners may make a shift from complexifying at the subordination level and move on to a higher-level complexification, that is, at the clausal level. In the data analyzed here, complexification at the clausal level probably emerged as a result of the interaction of two factors: the speakers' proficiency level, and increasingly higher cognitive demands. Performing a sequence of tasks therefore led to a deeper-level, more advanced structural complexification of speech, typically associated with different levels of L2 competence.

Relating this result to the theoretical underpinnings of the SSARC model of pedagogic task sequencing, the first two tasks may have served the function of a scaffolding mechanism. In these tasks, which were cognitively less demanding than their most complex counterpart, the speakers most probably relied on old knowledge and familiar structures, as a way of consolidating their current interlanguage state. Understood in this way, such linguistic behavior matches what Robinson (2010) labeled the "simple, stable <*attractor*> state of current interlanguage", associated with cognitively simple tasks. It was in response to the most complex task that the speakers stretched their syntactic repertoire beyond the simple subordination and moved on to deeper levels of structural complexification. Again, in Robinson's terms it can be stated that the most complex task promoted "restructuring of the current interlanguage system (…) introducing maximum complexity" (Robinson, 2010, p. 247).

As was previously mentioned, gradual increases in cognitive complexity had the opposite effect on those speakers who were only exposed to a single task; these participants complexified their speech at the level of subordination, and simultaneously decomplexifying it at the clausal level. As cognitive complexity increased, these speakers produced longer clauses and less subordination on the simple task. The opposite trend held true when these participants were faced with a more cognitively demanding task: they produced shorter clauses and more subordination. A noteworthy pattern, further adding to the differences in the way the two conditions were affected by cognitive complexity, is reflected in the visual representation of the findings. In the case of both measures, the pattern of findings in the baseline condition represented a perfectly straight line, demonstrating a linear and gradual nature of progression in structural complexity across tasks. The performance of the simple-complex group, on the other hand, was characterized by clearly discontinuous and irregular trends.

The tendencies observed in both groups show that simple-complex sequencing prompted a more profound difference at the syntactic level, pointing in the direction of a facilitative role of sequencing tasks from simple to complex in promoting quality output. This interpretation, however attractive and convincing, would not be complete without making reference to the actual means obtained in the simple task in the baseline condition and the most complex task in the simple-complex sequencing condition. These two data points received virtually the same values in words/clause (M=7.68 in the baseline and M=7.6 in simple-complex sequencing). Looked at from this angle, the findings can be interpreted as undermining the claims of the SSARC model insofar as the differences between the means obtained in both conditions can best be described as marginal, and the fact that the findings did not reveal a clear advantage for simple-complex sequencing; rather, the two groups took different routes to accomplishing the same goal: the simple task prompted greatest complexification in the baseline condition, and the most complex task did so in the simple-complex condition.

In the simple-complex sequencing condition, higher-level structural complexification of speech took place simultaneously with two other phenomena: a faster speech delivery than that displayed by the other group, and a considerably greater number of dysfluencies compared to their counterparts in the baseline condition. From the point of view of speech rate, the speakers in the simple-complex condition can therefore be described as more fluent than that of the other group, but from the point of view of repair fluency, they can be described as less fluent (at least in the way dysfluency was operationalized in the current study). It was in the area of dysfluency where the participants in both groups were found to be systematically affected in the same way. In this measure, the pattern observed across tasks was that simple-complex sequencing led to roughly double the amount of dysfluencies than performing tasks in isolation.

One of the explanations for this finding lies in the way dysfluency was operationalized in this study, a circumstance mentioned earlier in this chapter. Some researchers have suggested that dysfluency, and in particular self-repairs, should be conceived of as a measure of accuracy (as instances of focus on form) rather than as a measure of fluency (e.g., Gilabert 2007c; Lyster & Ranta, 1997; Swain, 1998). In the current dissertation, dysfluency encompassed a range of linguistic phenomena, including, among others, lexical and morphological self-repairs. In the analysis of the data, self-repair behavior was not disentangled from other dysfluency phenomena, such as repetitions and false stars; they were all considered jointyl as different components of dysfluency. Had the different phenomena been treated as separate objects of analysis, or had dysfluency been operationalized as an accuracy rather than fluency measure, the obtained results could potentially be interpreted as simple-complex sequencing leading to more accurate behavior. Since the actual accuracy measures employed in this dissertation pointed towards the speakers in the simple-complex condition delivering their output in a more target-like way (task 1 and partially task 2), such interpretation of dysfluency results would add evidence to the role of simple-task sequencing in promoting less deviation from language norms.

An alternative way of looking at these findings is that, given that across tasks the participants in the simple-complex condition delivered their speech faster than the other group, dysfluency behavior might be related to, or dependent on, speech rate: the faster the rate at which output is delivered, the greater the occurrence of dysfluency phenomena. Therefore, understood as a measure of fluency, dysfluency ratio showed that the participants in the simple-complex condition were more prone to reformulate and repair their speech, compared to their counterparts in the baseline condition.

It was suggested in previous research that it is the location of dysfluency phenomena, rather than the intensity of its occurrence, that distinguishes non-native from native-like production. More specifically, following Kormos (2006), there is a general agreement among researchers that in the case of non-fluent L2 speakers dysfluencies are located in clusters, whereas fluent speakers display dysfluent behavior at grammatical junctures (Kormos, 2006, p. 164). Along similar lines, following De Jong et al., (2013) dysfluencies may be conceived of "as solutions to the problems rather than problems because they are signals used by the speakers to inform their interlocutors of upcoming delays" (p. 912). These pieces of research illuminate the current work in two ways. First, the current study did not set out to explore the location of dysfluencies, but the frequency of their occurrence, which may have been an insufficient procedure. Second, in the current dissertation the occurrence of dysfluency phenomena was classified as symptomatic of a dysfluent linguistic behavior (so, as “a problem”) rather than “solutions to a problem”. Had the latter operationalization been adopted, the findings would call for alternative explanations.

Turning to speech rate, from a speech production perspective it could be argued that the participants in the simple-complex sequence produced their output faster because each subsequent task required systematically less engagement in the process of message conceptualization, given that the sequence of three tasks followed the same procedure and instructions, and analagous task input. After performing the first task, they were faced with a relatively similar scenario (task input). Therefore, the procedures encountered in the first task could be transferred to, or replicated in, new contexts (the complex and the most complex task). I believe that such transfer of knowledge to a new setting to some extent took away the need to conceptualize the message, and also, though to a lesser extent, required less planning at the micro and macro level. It is therefore possible that the available resources were channeled towards more automatized message delivery.

Along similar lines, from the perspective of instance theory (Logan, 1988), the observable advantage in automaticity displayed by the participants in the sequencing condition over those in the baseline condition could be attributed to memory. It could be the case that on subsequent tasks certain formulas were retrieved from memory and the solutions did not have to be searched for in the mental lexicon. As the participants engaged in more input reception and production of output, the formulas they had to use became gradually readily available. Rather than being selected from the mental lexicon, they were retrieved from memory in the form of ready-made, previously applied chunks, alleviating the mental burden associated with the search process. It can be speculated that in these participants, such relative availability of prefabricated, retrievable from memory chunks, resulted in less time necessary for message delivery.

In other words, in the case of the complex and the most complex task, what for the participants in the baseline condition was unfamiliar, new knowledge, in the case of the other group was relatively familiar; a copy of previous experience. In terms of Robinson's Triadic Componential Framework, it could be stated that in the simple-complex condition the complex and the most complex task were characterized by an additional dimension (prior knowledge and/ or task familiarity), which lessened overall complexity.

While these are viable explanations to the results obtained for speech rate in the case of the complex and the most complex task, it remains an intriguing question why the participants in the simple-complex condition delivered their tasks faster also in the case of the simple task. One could speculate that performing a whole sequence of tasks, rather than individual, isolated tasks, was in itself a condition which determined the emergence of specific linguistic behaviors in the speakers. However, in the set up of the current study, care was taken not to reveal to the participants that they would perform a sequence of tasks, so that the anticipation of multiple tasks would in itself not be a factor determining the speakers' subsequent linguistic behavior. The participants were informed that they would perform a sequence right before engaging in the performance of the first task. More specifically, once they received the input for the first task, they were informed that they would perform two more tasks following the one they were about to carry out.[31]

In the case of the participants in the simple-complex condition, faster speech delivery occurred concurrently with generating more target-like output, with most significant differences yielded in the case of the simple task. However, a circumstance hard to neglect is that in the other group, in which one participant performed a single

[31] It cannot be excluded, however, that even though the participants were informed of the sequence of tasks right prior to performance, such knowledge still had an influence on their performance.

task, more compliance with norms (and most target-like use of prepositions) was observed in the case of the most complex task, and the least in the simple task. This group's performance, characterized by approaching language norms as cognitive task complexity increased, was therefore not influenced by any additional exposure, let alone exposure to a specific sequence of tasks. I would therefore like to speculate that the agent responsible for the changes in interlanguage was the amount of cognitive load inherent in each of the individual tasks.

The core question pertinent to this section was whether performing tasks in a sequence presents benefits over performing tasks of differing levels of cognitive complexity in the absence of any other tasks. As postulated in the Cognition Hypothesis, and particularly in the SSARC model of pedagogic task sequencing associated with it, pedagogic tasks should be sequenced from simple to complex. The findings obtained for this research question, which compared three tasks performed by the same speaker and performed by different speakers, revealed that the manipulation of a task's cognitive load certainly has *a different effect when tasks are performed in a sequence versus in the absence of any other task*. As was seen throughout this section, all the areas except for lexical complexity were affected in one way or another, the major findings being that performing tasks in a simple-complex sequence led to greater speech fluency, increased accuracy, more clause-internal expansion, and considerably greater dysfluency. However, only in the case of dysfluency, the pattern was observed across all three tasks, with other dimensions representing rather irregular patterns and not all tasks being affected by them.

While one cannot ignore the different impact that the deliberate manipulation of cognitive demands had on the performance of tasks in a sequence versus in isolation, to my mind the nature of the encountered differences was not sufficiently homogenous to be able to state that performing a task in a sequence leads to qualitative changes in production compared to performing tasks in isolation. I believe that these findings, along with those reported in the previous section (simple-complex versus randomized sequencing) are confirmatory of a noticeable impact that cognitive task complexity exerts on a speaker, and a qualitative transformation it has the potential to bring about, possibly questioning the potential impact of task sequencing. This conclusion is evidenced in the fact that irrespective of the availability or not of other tasks, and the order in which the speakers performed the tasks, the most complex task, in randomized, simple-complex and baseline condition, triggered greatest accuracy, lexical complexity, and partially also structural complexity. The exploration of the issue of sequencing which this thesis attempted, rather than providing a definitive answer, opened a wide array of intriguing questions about the issue of sequencing and the impact of cognitive task complexity, which future research will hopefully address and further explore.

8.5 Research question 3a: How does task complexity affect the performance of speakers of different proficiency levels, as measured by CAF?

Research question 3a enquired about the effect of cognitive task complexity in speakers of different proficiency levels in the L2. The tasks in this dissertation were administered to two groups of participants, who were assessed as belonging to two different proficiency levels on the basis of their placement test results. As a preliminary analysis let us take a look at the patterns observed in both groups.

A common finding in both groups was that, as cognitive task complexity increased, all speakers' output was more lexically complex (according to one measure) and the output more target-like. However, regarding the results for accuracy, high proficiency speakers seemed to have taken greater advantage of increases in cognitive complexity. It is evidenced in the fact that in this group significant differences were located in all three measures of accuracy, whereas in low proficiency speakers only the specific measure, target-like use of prepositions, led to significant differences. The tendency towards more target-like output was therefore stronger, or more pronounced, in the case of high proficiency speakers.

At the same time, low proficiency speakers seemed to have taken a greater advantage of increases in cognitive complexity in the area of structural complexity, as demonstrated by the general measure, Words/AS-units. In these participants this measure demonstrated an increase in structural complexity, while a decrease was observed in the case of the high proficiency speakers. Additionally, low-proficiency speakers reached a peak in terms of clause-internal expansion in the most complex task, whereas in the other group the same held true for the simple task. A common trend in structural complexity in both groups was that there was a trade-off effect between two measures: Clauses/AS-units and Words/Clause. A decrease in one of them took place simultaneously with an increase in the other one, and vice versa. Regarding fluency, low proficiency participants maintained an overall stable pattern of speech rate across the three tasks, whereas high proficiency participants demonstrated a drop in fluency. These results demonstrate the following: 1) depending on the dimension and variable, speakers of different proficiency levels responded to increases in cognitive complexity in similar or different ways, and 2) both low and high proficiency speakers took advantage of increasing cognitive complexity, but they did so in different ways.

The patterns of findings, while undeniably related to each group's overall proficiency level in the L2, call for a deeper analysis of the architecture of attentional resources in speakers at different levels of competence in a foreign language, and the distribution of these resources at different stages of speech processing.

It is an accepted view that one aspect of becoming a more proficient L2 speaker is the conversion of declarative knowledge into procedural knowledge, a process related with developing automaticity in a foreign language (Kormos, 2006). As one becomes a more competent L2 speaker, certain processes associated with message generation and production get automatized, and what was once factual knowledge becomes a production rule in the course of L2 development. Given the results obtained for accuracy and structural complexity in the current experiment, it can be speculated that in the high proficiency speakers the speech encoding aspects of message generation and production – those associated with lexical, grammatical, and morpho-phonological aspects - were automatized, and so they required the allocation of a reduced amount of attentional resources. These speakers were therefore able to channel their attentional resources towards more advanced goals of message delivery (linguistic accuracy), and they consequently exerted an overall more efficient control over the quality of their linguistic output than the other group.

Low proficiency speakers, on the other hand, due to a more incomplete L2 system, perhaps had to allot a substantial amount of resources to lower-level aspects of message generation, such as selecting the desired lexical item from the mental lexicon, or assigning correct grammatical functions to the different components of a sentence. Consequently, the (limited) remainder of the resources was insufficient for as effective speech monitoring mechanism as in the case of high proficiency participants. Following Kormos (2000),

> "it can be assumed that due to the increased level of automaticity in lexical, grammatical, and phonological encoding, proficient speakers have more attention available for the other stages of message production. These remaining phases are the conceptualization of the message (…) and

> monitoring, which involves parsing both linguistic accuracy and contextual appropriacy" (p. 273).

While increases in cognitive complexity were beneficial to high proficiency speakers in the area of accuracy, low proficiency participants seemed to display greater gains in the case of structural complexity. The overall structural complexity measure, Words/AS-units, indicated a slight increase from the simple to the most complex task, and the most clause-internal expansion was observed in the case of the most complex task. Opposite trends in these two measures were observed in the high proficiency speakers. It can thus be speculated that in low proficiency speakers, enhanced cognitive complexity had a facilitative role in generating more complex syntactic structures.

These differences notwithstanding, it is curious that the results obtained for structural complexity were overall similar in both groups, unlike in the area of accuracy or lexical complexity, in which high proficiency speakers' performance was characterized by a markedly more advanced lexis and more error-free output. The difference between structural complexity and these two dimensions confirms the fact that structural complexity reaches a threshold beyond which output is not, or cannot be, complexified further. From a certain point onwards it represents a straight horizontal line, or even a U-shaped behavior, where, in terms of subordination, learners complexify their speech and later decomplexify it at more advanced stages of L2 competence. This stands in contrast to other dimensions, such as the size of the lexicon, characterized by a systematic expansion, or a greater compliance with language norms as interlanguage develops. The results obtained in the case of accuracy and structural complexity indicate that the level of competence in a second language determines the allocation of attentional resources to specific dimensions of performance.

Despite the different patterns of attention allocation observed and discussed in the case of structural complexity and accuracy, the two groups of speakers exhibited a surprising similarity in the trade-off observed between two measures of production: subordination ratio and clause-internal expansion. As was mentioned earlier, in both groups a decrease in one of them took place concurrently with an increase in the other. More specifically, both groups demonstrated shorter clause length from the simple to the complex task and a simultaneous increase in subordination index, but the opposite scenario emerged from the simple to the complex task (i.e., greater clause length and a lower subordination index). This result goes to show that, irrespective of proficiency level, a change in either direction in these variables takes place at the expense of the other variable, which suggests that competition occurs between these two aspects of structural complexity.

Having established potentially different ways in which attention is allocated in the case of accuracy and structural complexity, a counterintuitive finding is that low proficiency speakers maintained an overall stable level of speech rate across tasks, while their counterparts in the high proficiency level exhibited a drop in fluency. While both scenarios, stable speech rate and decreasing speech rate in response to increasing cognitive complexity, have been far from uncommon in literature, one wonders why a higher proficiency speaker should be more sensitive to increases in cognitive complexity than a lower level speaker. It could be the case that in high proficiency speakers, this group's careful control of the output evidenced in the accuracy measures, took a toll on the speed of message delivery. Focusing on producing output in compliance with language norms may have taken these speakers' attention away from delivering the message in a timely fashion. Beyond this cognitive account, it can also be speculated that in the case of high proficiency speakers risking one's fluency is not as face-threatening as it is at lower levels of L2 competence. From this point of view, the findings for accuracy and fluency could be understood in terms of priority: high proficiency speakers attached greater importance to accurate production than the rate at which they delivered their message.

To sum up the picture of performance obtained in this section, in the current dissertation increased cognitive task complexity had a positive impact on speakers in both groups: fewer errors were observed along with an increase in the target-like use of prepositions. However, in the high proficiency speakers more instances of significant differences were detected than in the low proficiency group; at the same time, in both groups there was a trend towards more target-like output as cognitive complexity increased. When drawing conclusions from the performance of these speakers, it must be borne in mind that on some occasions the detected differences were small in magnitude. It may be directly related to the fact that the low and high proficiency participants in the current study did not represent two extremes of proficiency, but rather a continuum of proficiency levels; the speakers involved were at different stages of competence in the L2 but they were not as different as a beginner and an advanced learner. Greater differences between the two groups would have been bound to emerge had the current study investigated the performance of speakers of more markedly different interlanguage levels. I speculate that further benefits of increasing cognitive complexity in the area of structural complexity could be expected in the case of low proficiency participants. Along similar lines, an exploration of a range of proficiency levels, rather than two proficiency levels, would give a more comprehensive insight into how the different dimensions of production respond to sequencing in a short-term fashion and how they develop in the long-term. In the case of the current experiment, a certain threshold in terms of syntactic structures and a range of lexicon were a prerequisite for performing these tasks, and they would be bound to pose too great a linguistic and cognitive challenge on speakers of a lower proficiency level than the one represented by the low proficiency speakers in the current study.

To my knowledge, in TBLT research only a handful of studies have investigated the impact of task complexity in relation to L2 proficiency differences. Moreover, none of the studies I am aware of investigated the effects of cognitive complexity as mediated by proficiency beyond a simple-complex dichotomy. Even so, let me compare the findings obtained in the current experiment with the picture of linguistic behavior obtained in five selected studies from the TBLT domain, which investigated L2 proficiency as an independent variable[32].

In the current dissertation, as a result of engaging in tasks of different cognitive complexity levels, both proficiency groups' accuracy and lexical complexity increased, with the high proficiency speakers taking greater advantage of increasing complexity in both. Low proficiency speakers' speech rate was stable across the three tasks, whereas in the other group it displayed a drop from the simple to the most complex task. Finally, somewhat greater gains could be observed in low proficiency speakers in the area of structural complexity.

In a task design similar to the one reported here, Malicka and Levkina (2012) investigated the impact of task complexity and proficiency on two oral tasks manipulated along ±elements and ±reasoning demands. In both studies, both proficiency groups took clear advantages of increasing cognitive complexity in the areas of accuracy and lexical complexity; however, the benefits were more pronounced in the case of the high proficiency group. Contrary to the findings

[32] Rather than being an exhaustive review of previous findings, this section serves the function of contextualizing the results obtained in the current experiment in light of selected relevant literature.

obtained in the current dissertation, in Malicka and Levkina (2012) the speech rate of participants in both groups *increased*, and, in fact, the difference reached statistical significance in the low proficiency group. The high proficiency participants, on the other hand, maintained their speech at an overall similar level in both tasks, with a slight tendency towards *more fluency in the complex task condition*. Along similar lines, a large discrepancy between both studies is that in the current experiment increases in cognitive complexity seemed to have proven beneficial to low proficiency speakers' speech, whereas structural complexity dropped according to the results of the other study. In the high proficiency group, both studies detected a drop in structural complexity.

In Ishikawa (2006), which used written tasks, there was an increase in all dimensions from the simple to the complex task condition, except for lexical complexity, which dropped in high proficiency participants. At the same time, it was precisely in the area of lexical complexity, and also in accuracy, where the low proficiency group seemed to have taken a greater advantage of increases in complexity. In a study by Kawauchi (2005), who investigated oral tasks in their unplanned and planned condition, high proficiency speakers received greatest benefits in structural and lexical complexity, and it was in the area of accuracy where low proficiency speakers displayed the greatest advantage. Finally, in a study by Kuiken, Mos, and Vedder (2005), which investigated written task performance in its simple and complex condition manipulated along ±number of elements, high proficiency speakers displayed more gains in terms of accuracy than the other group, with lexical and structural complexity not being significantly affected[33].

[33] Means and standard deviations for each proficiency group were not provided, so the actual trend in the data in the case of these two dimensions is unknown.

The fact that a very limited number of studies set out to investigate task complexity in relation with L2 proficiency does not enable conclusions to be drawn about the role of proficiency in performance. In addition to an insufficient body of research, making comparisons is particularly challenging given the operationalization of proficiency as an independent variable, which encompasses two aspects: different proficiency measurements across studies, and different benchmarks for classifying participants as "low" or "high" proficiency speakers. Also, the studies reported here used tasks in both oral and written modality, and they employed a different number and type of dependent variables. In spite of these shortcomings and limited evidence, the research carried out so far shows that *speakers at different stages of L2 competence take advantage of increasing cognitive complexity in different ways*. While it is premature to draw conclusions, the research reported here seems to imply that greater gains in some areas of performance may be associated with a certain proficiency level. For instance, it could be speculated that in high proficiency speakers, due to a more complete L2 system, and consequently, a better monitoring mechanism, increasing cognitive complexity could lead to greater gains in the area of accuracy, while such gains could be expected to a lesser extent in low proficiency speakers, who, given their incomplete L2 system, are not capable of controlling the quality of their output in such an efficient way. Independently from the exact nature of gains represented by speakers of different proficiency levels, which future research will hopefully address, proficiency is clearly one of the mediating factors in performance, and its impact cannot be neglected both when planning a study design and when drawing interpretations of the findings.

8.6 Research question 3b: Do low-proficiency learners benefit more from the simple-complex sequence than from the randomized one?

The issue which this section aimed to explore was whether sequencing tasks from simple to complex presented short-term advantages for low proficiency speakers. While this question was mainly inspired by the theoretical premises of the SSARC model of pedagogic sequencing, and in this sense it resembles the other perspectives on sequencing investigated in this dissertation, it additionally focused on speakers representing a specific, low L2 proficiency level. Main findings pertinent to this section were:

1) *randomized sequencing* led to statistically *faster speech delivery than simple-complex sequencing*, and to markedly *greater lexical complexity*;
2) *simple-complex sequencing* led to overall *greater accuracy*, particularly as demonstrated by global measures, with *randomized sequencing displaying a trend towards less error-ridden behavior* as cognitive complexity increased;
3) performing tasks in either order did not seem to considerably affect the structural complexity of the output, although *on several occasions randomized sequencing triggered greater gains*.

The picture of these findings is a complex and intriguing one. Quantitatively speaking, out of the four dimensions of performance this dissertation focused on – fluency, accuracy, lexical complexity, and structural complexity – sequencing tasks from simple to complex turned out to present benefits over randomized sequencing in one area: accuracy. In other words, on two global accuracy measures, randomized sequencing obtained lower means, indicating greater accuracy levels. It must be clarified, however, that the performance of the simple-complex group was stable across tasks as demonstrated by Errors/AS-units, and according to Errors/100 words, there was a drop from the simple to the complex task, with almost identical values obtained for the complex and the most complex task. On the other hand, the other group's error rate was generally higher than that detected in the simple-complex condition, but these participants still displayed a drop in error rate to the advantage of the most complex task. Randomized sequencing presented benefits in other areas – particularly so in the area of fluency (as demonstrated by speech rate) and in lexical complexity (as demonstrated by D). I believe that, from a psycholinguistic point of view there are two ways in which the obtained results can be interpreted.

First, it can be speculated that both types of sequencing, simple-complex and randomized, triggered lexical and grammatical encoding; however, in the data set analyzed in the current study, only one type of sequencing, simple-complex, led to better quality speech, as demonstrated through this group's performance being more error-free in general. The results obtained for accuracy demonstrate that systematic, gradual increases in cognitive complexity were conducive to more target-like output compared to alternative sequencing orders. Sequencing tasks from simple to complex perhaps alleviated the mental burden imposed on the speaker, so that attentional resources could be channeled towards more effective speech monitoring. When performing tasks in a random sequence, such an alleviative mechanism was not available, or not to such an extent as in the simple-complex sequence. In addition to accuracy, it is remarkable that in this group an increase in lexical complexity took place simultaneously with accuracy, and partially also with structural complexity (more specifically, overall structural complexity and mean length of clause). Considered from this angle, the findings potentially demonstrate that simple-complex sequencing in low proficiency participants triggered an optimal distribution of attentional resources between the areas of performance: lexical complexity, structural complexity, and accuracy. Such an account therefore confirms the facilitative role of simple-complex sequencing in generating what could be called “better quality output”.

From a different point of view, while the qualitative changes observed in accuracy in the simple-complex group are most likely attributable to the sequence in which these participants performed the tasks, one cannot neglect that randomized sequencing triggered a markedly higher speech rate and lexical complexity. A relative advantage of performing tasks in the sequence from simple to complex was therefore counterbalanced by this group’s performance on other measures. If the participants in the randomized sequence delivered speech characterized by a wider range of lexis and a higher speech rate, *at the same time as producing more error-free output in the most complex task,* this would probably confirm the conclusion drawn from the other perspectives on sequencing explored in this dissertation: the most complex task triggered the greatest lexical complexity and the greatest accuracy. In other words, irrespective of the sequence in which the tasks were performed, the simple task elicited overall less linguistically advanced output compared to its complex and most complex counterparts. It can therefore be speculated that, in terms of short-term effects of sequencing, qualitative changes in performance were not brought about by prior exposure to tasks of a lower cognitive load, but by internal cognitive design characteristics of a single pedagogic task. In response to a single task, lexical richness and accuracy of performance seemed to be proportional to the amount of cognitive load inherent in that task and imposed on the learner. An implication from these findings is that the quality of performance is determined by the mental load a task imposed, and an L2 speaker incorporates into their output just the number and variety of linguistic resources which the task demands from them, and not the actual resources they have available: these get activated only when an appropriate stimulus, in the form of higher cognitive load, is available.

To sum up, in findings obtained in the current experiment there is no compelling evidence that simple-complex sequencing led to *substantial* qualitative changes in performance. What the findings do show is that different sequencing orders may be responsible for different ways in which attentional resources are distributed during task performance, and the production of second language speakers involved in this experiment turned out to be sensitive to the type of sequencing they were exposed to. To my mind the results obtained provide powerful evidence of the role of cognitive task complexity on performance. The relative effect of task complexity versus that of task sequencing, considering all findings obtained, is dealt with in depth in the upcoming section.

8.7 Conclusion

There were three main objectives to the current study: to provide further evidence to the role of cognitive task complexity in performance (when tasks are performed in a sequence and in isolation), to investigate the ostensibly facilitative role of simple-complex sequencing, and to explore the role of proficiency as a mediating variable in L2 performance.

The findings obtained and the interpretations drawn lead to three main conclusions. First, the results showed that L2 speakers in this experiment were clearly sensitive to individual tasks of differing levels of cognitive complexity, as theoretically hypothesized *and empirically researched*. This is evidenced in the fact that the quality of their performance was determined by the cognitive load ostensibly imposed by the design of an individual task. As demonstrated by the results obtained for all research questions, the number and variety of linguistic resources employed during task performance was the function of the amount of cognitive load: the higher the cognitive load of a task, the more resources were invested in its execution, and consequently, the more complex the output became. To my mind, such susceptibility of speakers to the amount of cognitive load inherent in an individual task demonstrates a profound impact that cognitive task complexity exerts on L2 speakers.

While some of the obtained results suggest a potentially facilitative role of simple-complex sequencing in promoting more target-like output, such as when lower-level participants engaged in performing tasks in the simple-complex sequence, there is no *compelling* evidence to suggest that simple-complex sequencing presents benefits over any alternative order of sequencing tasks. In fact, on several occasions it was demonstrated that randomized sequencing presented advantages over simple-complex sequencing. Counterintuitive as this finding may seem, I think it goes to show that, at least in this experiment, task sequencing affected speakers in somewhat unpredictable and unsystematic ways. While this is by no means definitive, and is yet to be explored by further research, to me the results obtained for sequencing add further evidence to the intense impact of task complexity: if performance is qualitatively different in response to a cognitively demanding task irrespective of whether other tasks (of a lower cognitive load) are available or not, it means that cognitive task load alone is an instrument determining performance outcomes. At the same time, the evidence found for sequencing is simply not strong enough to draw conclusions about its role, and future research will hopefully shed more light on its affect on performance and interlanguage development.

Finally, there is evidence to suggest that the patterns of speech performance are related to the speaker's proficiency level. The findings obtained showed that speakers at different stages of competence take advantage of increases in cognitive task complexity in different ways. More specifically, it seemed to be the case that increasing cognitive load was beneficial to low proficiency participants, whereas high proficiency participants gained more in terms of accuracy.

8.8 Limitations

The current study presents a number of limitations, the most pertinent of which will be addressed here.

First, in its exploration of the role of sequencing, a narrow approach was adopted in the sense that only short-term effects of sequencing were explored. It is an interesting question and one worthy of empirical investigation, how participants' output would have been affected if long-term effects of tasks had been explored in a pre- post-test design. I feel, however, that given that the research on sequencing is only starting to emerge as a new area of investigation, exploring its short-term, rather than long-term effects was a natural exploratory first step.

In its predictions about the effects of sequencing on performance, the SSARC model of pedagogic task sequencing contemplates the influence of both resource-directing and resource-dispersing factors. In the task design pertinent to the experiment reported here, only resource-directing dimensions (±elements and ±reasoning demands) were manipulated, rendering the results illuminative of the SSARC model only up to a certain point, and thus caution is required when interpreting them. The inclusion and manipulation of only resource-directing factors was motivated by exploring only short-term effects of sequencing. I believe that a sequence of three tasks, which is the number of tasks included in the current experiment, allows for only a limited number of manipulations of different factors.

Not only does a manipulation of more factors require a larger number of tasks, but also, I believe, it leads to a much larger-scale study than the one reported here, with multiple tasks and procedures requiring substantially greater and more complex operations in terms of validating task complexity. I therefore considered a sequence of three tasks manipulated along two resource-directing variables a reasonable compromise between exploring only a simple-complex dichotomy versus a complete sequence of tasks (whatever number of tasks a “complete” or “entire” sequence would, or should, be comprised of), and between exploring only one variable versus multiple variables.

In the design of the current study, apart from measuring short-term effects of sequencing, a decision was taken that the three tasks would be performed in a single sitting, with short pauses between tasks, rather than at longer intervals, potentially operationalized in terms of hours, or days. While such a set-up potentially encouraged the transfer of skills from one task to another, the fundamental objective of this dissertation was precisely to explore how performance changed qualitatively when tasks were performed in a subsequent manner, and not at long intervals.

Moreover, both scenarios described in this dissertatoin (i.e., performing two or more tasks in a subsequent fashion versus at a longer time interval) are common practices in TBLT research, and I therefore considered administering three tasks in one sitting a procedure consistent with other studies in the TBLT domain. At the same time, it cannot be neglected that in the case of this particular study, given a substantial number of instructions and procedures inherent in each task, the speakers might have perceived the experiment as cognitively overwhelming, and carrying out a sequence of tasks, by the time they engaged in the third task, might have taken a toll on their overall performance because of sheer tiredness.

8.9 Practical implications: textbook writer, syllabus designer, and classroom teacher

The tasks employed in this dissertation, although they represent only a short-term sequence, as well as the results obtained from the analysis of speech samples of speakers involved in their performance, can serve as a starting point in designing a textbook unit, organizing a curriculum around tasks, or simply being instruments readily available to be used in the classroom context. In this sense, the findings obtained in the current dissertation have a number of implications for practitioners involved in the decision-making process at different stages of development of input to be subsequently processed by learners.

From the perspective of a syllabus designer, the designed tasks can serve as a starting point for the development of a task-based curriculum for tourism students, delivered over an extended period of time, such as a semester or an academic year. In this scenario, the tasks used in this dissertation could also be manipulated along resource-dispersing variables, such as planning time. The range of variables could be subsequently expanded to include different types of variables beyond the cognitive ones, such as participant variables (e.g., ±same gender) or participation variables (e.g., ±few participants). Finally, the tasks could also be performed in the written modality. The inclusion of such practices would naturally lead to a much longer-term sequence, with tasks being in the center of classroom proceedings.

From the perspective of a classroom teacher, the tasks designed for this dissertation can be conceived of as a readily available and easily implementable short-term sequence of tasks either as a free standing item, administered as complementary material in addition to a regular, not necessarily task-based course book, in which case it would serve as oral practice of lexis or syntax imposed by external material. In a different scenario, they could be seen as a slightly longer-term unit, with the tasks administered in this experiment representing task cycle, the performance of which could be complemented with pre- and post-task activities.

Beyond a readily-available, practical implication, the findings obtained in this study and similar research offer potentially important take-home messages which may help take informed decisions about teaching and learning foreign languages. Perhaps one of the most general, yet fundamental messages is that even slight modifications to the internal features of a task – be it a pedagogic task in the sense defined in TBLT literature, or any unit of classroom activity resembling such a task, makes a difference in performance. The body of research has shown that learners' speech is not only susceptible to such modifications, but, to some extent, these alterations seem to predetermine the outcome in terms of the different dimensions of production and target forms, an outcome likely impossible to accomplish by means of decontextualized formats typical of synthetic syllabi. The fundamental advantage of exposing learners to a range of tasks in different modalities (e.g., oral vs. written), types (e.g., description vs. decision-making), and with different manipulated variables or groups of variables (e.g., ±reasoning demands or ±planning time) has a potential of leading to a balanced development of different dimensions of performance over time.

Due to a relative lack of systematicity across findings in TBLT research, it is too early for definitive statements about the exact outcome of the different manipulations, with more questions left to answer than answered questions. However, the body of research carried out so far, and the accumulated knowledge, in the very least raise awareness of how performance is mediated by different factors.

8.10 Theoretical implications for available task complexity and task sequencing theories

The exploration of independent variables pertinent to this study drew on two cognitive complexity models: Robinson's Cognition Hypothesis and Skehan's Trade-off Hypothesis. In addition, two frameworks associated with the Cognition Hypothesis also informed this dissertation: the Triadic Componential Framework, and the SSARC model of pedagogic task sequencing.

The available cognitive task complexity models have guided and informed most of TBLT research. Taking as a starting point different approaches to how attentional resources are distributed, they make predictions as to how different groups of factors, including the mental effort required for task completion (*cognitive factors*), participation and participant variables (*interactive factors*), and the unique set of abilities a speaker brings to a task (*learner factors*) mediate a range of outcomes, such as the amount of between-participant interaction, variation within the interlanguage at a particular point in time (language performance), and transformation within the interlanguage over a period of time (language development). While these theories agree that performance is qualitatively different depending on the presence or absence of a feature inherent in task design, or its degree of intensity, the area in which they differ are the specific predictions about performance, particularly in relation to the trade-off effects in the areas of lexical complexity, structural complexity, and accuracy.

The experiment reported here, the research on cognitive task complexity referenced in this dissertation, and the remainder of TBLT literature (the reporting of which is beyond the scope of this dissertation), have accumulated sufficient evidence so as to call for a potential revisiting of some of the assumptions of the hypotheses. In their predictions, the hypotheses do not narrow down the scope of tasks, or variables, which have a potential to bring about qualitative changes in performance or interlanguage development. Therefore, it is often uncritically assumed that all sorts of manipulations of task design in any sort of task will elicit qualitative changes of some kind. It is perhaps overly optimistic to expect that all sorts of manipulations will necessarily trigger qualitative changes in performance in all dimensions of speech production. Following Pallotti (2009), "Why should all sorts of task complexification lead to higher complexity of any linguistic feature? Why should telling a story in the past stimulate the production of more rare or varied lexicon than the production of the same story in the present tense? And why should making a decision with more elements produce a higher subordination ratio than making one with fewer elements?" (p. 595).

I believe that the results obtained for structural complexity, both in this study and in the body of available TBLT literature, illustrate this point. Studies have demonstrated that this dimension of production does not undergo significant qualitative modifications in any systematic way in response to a single task in its simple and complex condition, at least in the way it has been have measured so far, and it has been shown to behave in oftentimes counterintuitive and unpredictable ways. At the same time, as reported earlier in this chapter, it has been shown to be affected, for example, by the availability or not of the time to plan the message. These results imply that the structural complexity of output is more responsive to task-external conditions (*resource-dispersing dimensions*) than to task-design internal characteristics (*resource-directing dimensions*). By contrast, other dimensions of performance, such as accuracy, have been shown to approach language norms under a cognitively complex task and not a simple one. Consequently, making predictions about the direction in which structural complexity changes in response to a single task manipulated along resource-directing direction is probably questionable.

Another characteristic of structural complexity, which potentially differentiates it from the other dimensions of production, is that it is sensitive to L2 proficiency: different types of complexification may be expected as a speaker reaches a higher level of competence in the L2. Therefore, low proficiency speakers may complexify at the level of subordination, whereas high proficiency speakers may do so at the level of clause-internal expansion. In light of this, the claim about "increasing structural complexity", common to both hypotheses, would be more precise if it specified the level at which such complexification takes place.

Along similar lines, the Cognition Hypothesis states that fluency, understood as speech rate, decreases as cognitive task complexity increases. While this claim has been confirmed by a number of empirical investigations referenced in the literature review section of this dissertation, it was an uncommon scenario to see the complex task delivered in a *substantially slower manner* than its simple counterpart. Moreover, alternative findings in fluency have been frequently observed: this dimension often remained intact as a result of increasing cognitive complexity, and on some occasions it increased from the simple to the complex task condition. These scenarios observed in literature naturally lead to the question of the specific conditions under which fluency drops, stays intact or increases, and whether the magnitude of the variation from a cognitively simple to a complex task is attributable to the manipulation of a single cognitive variable, or in relation to other task-internal manipulations and external conditions. The current study hinted at the possibility that the pattern represented by fluency may be related to proficiency level: stable or increasing behavior in the case of low proficiency speakers.

The fact that specific types of structural complexification, and potentially also fluency behavior, may be associated with certain proficiency benchmarks, takes me to the role of other factors than cognitive ones in performance, such as individual differences. While the available frameworks, particularly the Triadic Componential Framework, do acknowledge these factors' role in performance, the role of learner and individual factors has been largely at the periphery of TBLT exploration, and both the hypotheses and research carried out to date have focused primarily on whether, and how, a speaker's interlanguage changes as a result of a deliberate manipulation of task design characteristics, individual differences being tested not to such an extent. There is emerging evidence to suggest that performance, while clearly susceptible to modifications in cognitive load, is strongly influenced by other factors, such as L2 proficiency. As was concluded earlier in this chapter, speakers representing different stages of L2 development may present gains in different areas (although the picture of gains is by no means conclusive yet), and therefore a relative increase or decrease in a certain dimension may be related to this (or/ and other) individual differences. Therefore, the same two or more tasks at different cognitive complexity levels may generate markedly different speech samples depending on a variety of individual factors, which should be inherent both in theories attempting to make statements about the development of interlanguage, and empirical research putting these theories to test. In a narrow sense, this implies interpreting research findings bearing in mind the unique characteristics of the investigated population, and in a broad sense it implies the incorporation of these unique characteristics (individual differences) into study designs, potentially being more illuminative of complex processes underlying different phenomena in SLA than investigating single variables. At the same time, discovering universal task design factors, which mediate performance in predictable ways *irrespective of other conditions and circumstances*, has been precisely what much of TBLT theorizing has focused on and empirical investigations set out to explore. This dissertation is an example of such attempts.

Regardless of whether researchers take into account factors other than the cognitive ones, in the Triadic Componential Framework and in research investigating it, most authors have operationalized variables along the simple-complex dichotomy, with "plus or minus a feature" indicating a simple or a complex task. Whereas in the case of some variables there is general consensus about the relative complexity of a task associated with "plus" a feature (e.g., many elements typically equals a complex task), it seems that some of the variables defy a classification as complexifying or decomplexifying factors in task design. That is, not always more or less of a feature leads to a classification of a pedagogic task as "simple" or "complex". I believe that the condition monologic/dialogic is a case in point. In the first pilot study carried out prior to the main experiment in this dissertation, the tasks were originally designed to be dialogic, with one interlocutor playing the role of a receptionist, and the other one that of a potential client. In the process of analyzing the output obtained from the task-based interactions, as well as the results of the affective variables questionnaires and stimulated recall, it was discovered that the presence of an interlocutor (so, the variable ±monologic task) can be either a complexifying or a decomplexifying factor, depending on the unique characteristics the interlocutor brings to a task, which can either alleviate or build up the mental burden of any task. In the same way, it is not clear whether some of the participant variables, such as same gender or same proficiency level, are factors adding to the overall complexity of a task, or possibly reduce it.

Most importantly, however, none of the available theories stipulates just how much of a difference in cognitive load is sufficient to bring about a change in performance, although there is an effort being made in terms of the independent measurement of task complexity. A lack of theoretical guidelines has left researchers at great liberty to operationalize variables, and subsequently test their impact in practice. At the same time, the accumulated body of research has provided sufficient evidence to make somewhat more fine-tuned and precise predictions, ones which associate certain types of manipulations of task-internal and task-external conditions with more or less specific qualitative changes in performance.

Turning to the issue of sequencing, and more specifically, the SSARC model of pedagogic task sequencing, while representing a conceptual advancement over previous proposals, it leaves several key areas undefined. The first issue concerns task sequencing principle 1. According to this principle, all considerations which are *not* cognitive in nature – broadly speaking, those related to interactional and interactant demands – should be kept constant throughout subsequent, cognitively more demanding, task versions. Such an approach to sequencing inevitably raises the question of just how feasible such a scenario is in a pedagogic setting, and to what extent it is authentic. In a hypothetical practical scenario, it implies that, as cognitive complexity increases, the designed tasks should be performed under otherwise exactly the same conditions and circumstances: on the interactional side of demands, the task is repeatedly performed by the same number of participants and it has a convergent solution; on the interactant side of demands, the interlocutor repeatedly holds the same characteristics in terms of L2 proficiency, sex, and shared cultural knowledge. The authenticity and practical application of such a scenario is highly dubious.

This criticism about the SSARC model is not intended to suggest that it contemplates too few factors as the basis for sequencing; rather, the point of criticism is that, in addition to the factors already considered, in long-term task sequences and genuine pedagogic settings, manipulation via factors other than cognitive is not only bound to emerge as a natural step in classroom procedures, but can also be a desirable expansion of sequencing based only on cognitive factors.

The second issue concerns, in brief, the length of a sequence. Is there a minimum number of tasks, or an optimal number of tasks, which should constitute a sequence; in other words, *how long a sequence should be*? It is perhaps the most fundamental unaddressed issue in Robinson's proposal. More specifically, from a pedagogical perspective, does "sequence" refer to a short-term, one-time intervention, designed and performed within a designated time frame (such as one lesson), or a full-scale program-long curricular planning, delivered over an extended time period (e.g., one semester), in which multiple tasks of multiple cognitive complexity levels are involved? In the latter scenario, several relevant issues are the time period during which tasks actually stay simple on all dimensions, how many resource-dispersing variables are manipulated before the resource-directing variables are incorporated, and finally, how many resource-directing variables are manipulated. In other words, one might wonder whether "sequencing" should be conceived of as a *micro-level* short-term intervention not contextualized within a longer-term scheme, or as a *macro-level* systematic plan.

Independently from this quantitative consideration, and from the adopted operationalization of "sequencing", one also wonders at what intervals the tasks in a sequence should be performed: should sequences be conceived of as performing tasks of different cognitive complexity levels in a subsequent fashion, or should performance of two tasks be separated by a substantial amount of time. The decisions concerning these considerations perhaps depend on the general purpose for which tasks are designed (research vs. pedagogy). Further considerations to take into account are perhaps those of the measured outcome (e.g., between-participant interaction, language performance, and development of target forms).

8.11 Methodological and study design implications

The claims and predictions of the available cognitive complexity models have prompted an exhaustive and much-needed empirical investigation of the impact of a range of variables on performance. Each study is a unique contribution to the knowledge of whether and how performance changes under individual conditions (e.g., ±elements) or groups of conditions (e.g., ±elements and ±planning time), and the body of research carried out so far has deepened our understanding of the complex process underlying speaking a foreign language, and has considerably advanced the TBLT domain and the SLA field. However, a lack of common agenda across studies, including aspects such as frequently vague or inexistent operational definitions, assuming cognitive complexity levels rather than investigating them empirically, and the choice of type and number of dependent variables, have all led to studies yielding incomparable, and often contradictory findings.

While some studies have made an attempt to externally validate cognitive complexity involved in a task, and thus verify the hypothesized cognitive complexity differences, such evidence of emerging good practices is scarce and definitely far from a standard practice in our domain. Without minimal validation of cognitive load differences, the findings obtained in task-based research are imprecise at best, if not invalid altogether. Despite some authors' attempts to measure cognitive task complexity independently from task performance, the issue of how much difference is sufficient, or necessary, to distinguish between a simple and a complex task version, is, in itself, a complex task. When comparing the results observed in those studies which employed an affective variable questionnaire as a measure of validating complexity (see chapter 3), it could be observed that a complex task in one study is, in fact, a simple one in another study, and vice versa. Measuring cognitive complexity in itself, although a much-needed and illuminating procedure, is a necessary prerequisite and starting point for a subsequent exploration of a task's effect on language-related phenomena, but a critical assessment of the magnitude of difference between two tasks should precede any interpretations about the actual difference in cognitive load.

In the current piece of research, developing three tasks of varying levels of cognitive complexity was obtained through the use of three complementary methods: affective variables questionnaire, time judgment task, and stimulated recall. Although the results of the time judgment task proved rather inconclusive, the hypothesized differences in cognitive load were reflected in the speakers' quantitative and qualitative perceptions. Whether or not the procedures employed here were sufficient or optimal is a separate issue, but a correspondence between the hypothesized differences and the participants' subjective perception allowed me to attribute the qualitative changes observed in performance to each task's internal cognitive complexity, and not other factors.

In this dissertation care was taken also to select optimal measures of performance in terms of their number and variety. Using diverse measures for each of the dimensions of production – fluency, accuracy, structural complexity and lexical complexity - has rendered the findings sound in two ways: obtaining a comprehensive, complete picture of production and drawing reliable conclusions about speech production. Although employing several measures may result in them not correlating with each other, it is an advantage rather than a handicap in the sense that it either further illuminates the patterns of performance (such as in the case of structural complexity in the current study: subordination ratio and mean length of clause), or it provides insights into the reliability of a measure (Guiraud's Index vs. D in the current study). On the other hand, including a specific accuracy measure in addition to the global ones, target-like use of prepositions, confirmed and further reiterated the claim that enhanced task complexity promotes more target-like output.

Both in this study and in TBLT research in general, the number and diversity of dependent variables used, and their exact measurement, unavoidably influence study findings, given that each measure captures a specific, unique phenomenon, or at least a specific perspective on the same phenomenon. The multifaceted nature of speech performance calls for its comprehensive depiction in the instruments used to measure the different phenomena. Not less important than the number and variety of measures is their appropriateness to the overall study design, including factors such as task modality, the participants' proficiency level or age, given that the measures used inspire subsequent study results and their interpretation.

8.12 Concluding remarks

I believe that in order to advance the TBLT domain further, a series of actions on the part of researchers apply, such as careful, theory-driven operationalizations, validating hypothesized cognitive complexity differences, piloting designed tasks, and selecting optimal measures. The above-mentioned are only examples of practices, which would benefit from more stringent approaches. In the domain of task-based language teaching, the aforementioned are the necessary prerequisites for a rigorous study design and constitute sound criteria for a study's internal validity. A systematic practical implementation of these fundamental, yet often neglected practices, is bound to lead to an overall standardization of procedures and greater study comparability, which, in turn, will render study findings more trustworthy and generalizable. Without incorporating the above mentioned as common practices, a rejection or acceptance of a hypothesis, which a large portion of TBLT explorations has focused on, is untimely.

Standardizing these procedures, along with carrying out more replication studies and carrying out systematic, extensive research syntheses, will render task-based language teaching a more solid ground for further conceptual and empirical investigations, and a mature domain within the field of second language acquisition, which I hope to actively participate in.

REFERENCES

Ahmadian, M. J. (2011). The effect of "massed" task repetitions on complexity, accuracy and fluency: Does it transfer to a new task? *The Language Learning Journal*, *39*(3), 269-280.

Ahmadian, M. (2012). The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research, 16*(1), 129–149.

Albert, A., & Kormos, J. (2004). Creativity and narrative task performance: An exploratory study. *Language Learning*, *54*(2), 277–310.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bannert, M. (2002). Managing cognitive load – recent trends in cognitive load theory. *Learning and Instruction*, *12*(1), 139-146.

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

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and instructed second language learning*. Bloomsbury Publishing.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*(2), 157–168.

Breen, M. P. (1987). Learner contributions to task design. In C.N. Candlin & D.F. Murphy (Eds.), *Lancaster papers in English language education (Vol. 7) Language learning tasks* (pp. 23-46). Englewood Cliffs, N.J.: Prentice Hall.

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.

Brown, G. K., Malmkjaer, K., & Williams, J. (Eds.). (1984). *Teaching talk: strategies for production and assessment*. Cambridge: Cambridge University Press.

Burke, S. (2001). Missing values, outliers, robust statistics and nonparametric methods. *LC-GC Europe Online Supplement*, *Statistics and Data Analysis, 2*, 19-24.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2000a). *Researching pedagogic tasks: second language learning, teaching and testing*. Harlow: Longman.

Candlin, C.N. (1987). Towards task-based language learning. In C.N. Candlin & D. Murphy (Eds.), *Lancaster papers in English language education* (Vol. 7). *Language learning tasks* (pp. 5-22). Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. (1993). *Human cognitive abilities*. New York: Cambridge University Press.

Cegarra, J., & Chevalier, A. (2008). The use of tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods*, *40*(4), 988–1000.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151–170.

Choong, K. P. (2011). Task complexity and linguistic complexity: an exploratory study. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *11*(1), 1-28.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.

Craik, F., & Lockhart, R. (1972). Level of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Cromer, R. (1974). The development of language and cognition: The cognition hypothesis. In B. Foss (Ed.*), New perspectives in child development* (pp. 184-252). Harmondsworth: Penguin.

Crookes, G. (1986). *Task classification. A cross-disciplinary review*. Center for Second Language Classroom Research, Social Science Research Institute, University of Hawai'i at Mānoa, USA.

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

Council of Europe (2011). *Common European Framework of Reference for Learning, Teaching, Assessment*. http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*(1), 5-43.

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

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25*(2), 220–242.

Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research, 4*(3), 193-220.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, New York: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59-84.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition*. *Monographies, 32*(3-4), 259-296.

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

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

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

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

Gilabert, R., & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In K. McDonough, & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (Vol. 34) (pp. 45-69). Amsterdam: John Benjamins Publishing Company.

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

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The cross-linguistic study of language acquisition* (Vol 1) (pp. 1008-1025). Hillsdale, NJ: Lawrence Erlbaum.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics, 12*(4), 345-364.

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

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Paris: Presses Universitaires de France.

Gwizdka, J. (2010). Distribution of cognitive load in web search. *Journal of the American Society for Information Science and Technology, 61*(11), 2167-2187.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*(6), 236-242.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(06), 803–831.

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

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

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

Hsu, H. (2012). Investigating the effects of planning on L2 text chat performance. *CALICO Journal, 29*(4), 619-638.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of rehearsal, elaboration and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 349-381). Cambridge: Cambridge University Press.

Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Blackwell.

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL, 3*(4), 193-225.

Ishikawa, T. (2007). The effect of manipulating 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, Buffalo, Toronto: Multilingual Matters.

Ishikawa, T. (2008). The effects of task demands of intentional reasoning on L2 speech performance. *The Journal of Asia TEFL*, *5*(1), 29-63.

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second language task complexity*. *Researching the Cognition Hypothesis of language learning and performance* (pp. 307-347). Amsterdam: John Benjamins Publishing Company

Jackson, D. O., & Seuthanapornkul, S. (2013). The Cognition Hypothesis: A Synthesis and meta-analysis of research on second language task complexity. *Language Learning*, *63*(2), 330-367.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, *63*(1), 87-106.

Jarvis, S., & Daller, M. (Eds.). (2013). *Vocabulary knowledge: Human rating and automated measures*. Amsterdam: John Benjamins Publishing Company.

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 143-164). Amsterdam: John Benjamins Publishing Company.

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System: An International Journal of Educational Technology and Applied Linguistics*, *37*(2), 254-268.

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition, 34*(4), 627– 658.

Kim, Y., & Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity*. *Researching the Cognition Hypothesis of language learning and performance* (pp. 287-306). Amsterdam: John Benjamins Publishing Company.

Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research, 13*(4), 301-347.

Klein, W., & Perdue, C. (Eds.). (1992). *Utterance structure: Developing grammars again*. Amsterdam: John Benjamins Publishing Company.

Kormos, J. (2000). The role of attention in motoring second language speech production. *Language Learning, 50*(2), 343-384.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning, 62*(2), 439-472.

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

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

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*(1), 48-60.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, London: Routledge.

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

Lee, Y. (2002). *Effects of task complexity on the complexity and accuracy of oral production in L2 Korean*. Unpublished PhD dissertation. University of Hawai'i at Mānoa, USA.

Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency*. *Complexity, accuracy and fluency in SLA* (pp. 171-197). Amsterdam: John Benjamins Publishing Company.

Littlewood, W. (2004). The task-based approach: some questions and suggestions. *ELT Journal, 58*(4), 319-326.

Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching, 40*(3), 243-249.

Logan, G. D. (1988). Toward an instance theory of automatisation. *Psychological Review, 95*(4), 492–527.

Long, M. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hylstenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77-99). Clevedon: Multilingual Matters.

Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly, 24*(4), 649-666.

Long, M. H. (2005). *Second language needs analysis*. Cambridge, New York: Cambridge University Press.

Long, M. H. (2006). *Problems in SLA*. Mahwah, New Jersey: Lawrence Erlbaum.

Long, M. H., & Norris, J. (2000). Task-based language teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597-603). London: Routledge.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*(01), 37–66.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malicka, A., & Levkina, M. (2012). *Measuring task complexity: Does L2 proficiency matter?* In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43-66). Amsterdam: John Benjamins Publishing Company.

Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon: Multilingual Matters

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.

Matthiessen, C., & Thompson, S. A. (1988). The structure of discourse and 'subordination'. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (Vol. 18) (pp. 275– 329). Amsterdam: John Benjamins Publishing Company.

Meara, P. & Miralpeix, I. (2004). D_Tools, Swansea: Lognostics (Centre for Applied Language Studies), University of Wales Swansea.

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

Meraji, S. R. (2011). Planning time, strategy use, and written task production in a pedagogic vs. a testing context. *Journal of Language Teaching and Research, 2*(2), 338-352.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259.

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

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

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, *12*(1), 11-37.

Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing, 19*(2), 61-81.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*(3), 214–255.

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

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 716-761). Blackwell.

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

Norris, J. M., & Ortega, L. (2012). Assessing learner knowledge. In S. M. Gass & A. Mackey (Eds.), *The handbook of second language acquisition* (pp. 573-589). New York: Routledge.

Norris, J. M., & Ortega, L. (Eds.). (2006). *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins Publishing Company.

Notash, M. Y., & Yousefi, M. H. (2011). Uptake in task-elicited L2 performance: Can task complexity matter? *Journal of Language Teaching and Research, 2*(3), 508-516.

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

Nuevo, A. M., Adams, R., & Ross-Feldman, L. (2011). Task complexity, modified output, and L2 development in learner–learner interaction. In P. Robinson (Ed.), *Second language task complexity*. *Researching the Cognition Hypothesis of language learning and performance* (pp. 175-202). Amsterdam: John Benjamins Publishing Company.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

Nunan, D. (2004). *Task-based language teaching*. Cambridge*:* Cambridge University Press.

Orr, J. M., Sackett, P. R., & DuBois, C. L. Z. (1991). Outlier detection and treatment in I/O Psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology, 44*(3), 473-486.

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

Ortega, L. (2004). Aproximaciones cognitivo-interaccionistas al aprendizaje de segundas lenguas mediante tareas. *Estudios de Lingüística Inglesa Aplicada*, *5*, 15-38.

Pallotti, G. (2009). CAF: Defining, refining, and differentiating constructs. *Applied Linguistics, 30*(4), 590-601.

Paas, F., Touvinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63-71.

Paas, F., & Van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and instruction. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning* (pp. 9–34). London: Multilingual Matters.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61–86.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Quesada, J., Kintsch, W., & Gómez, E. (2005). Complex problem-solving: a field in search of a definition? *Theoretical Issues in Ergonomics Science, 6*(1), 5-33.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reigeluth, C. M. (1999b). The elaboration theory: Guidance for scope and sequence decisions. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2) (pp. 425–453). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Révész, A., Michel, M. C., & Gilabert, R. (2012). *Measuring cognitive task demands: A step forward in task complexity research*. Paper presented at the 22nd European Association of Applied Linguistics (EuroSLA) Conference, Poznan, Poland.

Richards, J., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. Harlow, Essex, England: Longman.

Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning 45*(2), 283–331.

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57.

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

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Blackwell.

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

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

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). Amsterdam: John Benjamins Publishing Company.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics*, *30*(4), 533-554.

Robinson, P., & Ellis, N. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York, London: Routledge.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics, 45*(3), 161-176.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111-141). Amsterdam: John Benjamins Publishing Company.

Sasayama, S. (2011). Cognition Hypothesis and second language performance: comparison of written and oral task performance. *Second Language Studies, 29*(2), 107-129.

Sasayama, S., & Izumi, S. (2012). Effects of task complexity and pre-task planning on EFL learners' oral production. In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 23-42). Amsterdam: John Benjamins Publishing Company.

Schank, R. (1999). *Dynamic memory revisited*. New York: Cambridge University Press.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Segalowitz, N. (2010). Fluency, second language acquisition, and cognitive science. In N. Segalowitz (Ed.), *Cognitive bases of second language fluency* (pp. 1-28). New York, London: Routledge.

Shiau, Y. S., & Adams, R. (2011). The effects of increasing reasoning demands on accuracy and complexity in L2 oral production. *University of Sydney Papers in TESOL*, 6, 121-146.

Skehan, P. (1996a). A framework for the implementation of task based instruction. *Applied Linguistics 17*(1), 38-63.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (1999). Modeling second language task performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532.

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

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

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language learning* (pp. 183–205). New York: Cambridge University Press.

Slobin, D. I. (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.

Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. In M. P. García-Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 157-177). Clevedon, Buffalo, Toronto: Multilingual Matters.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). New York: Cambridge University Press

Swales, J. (2009). The concept of task. In K. Van de Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching: A reader* (pp. 41–55). Amsterdam: John Benjamins Publishing Company.

Sweller, J. (1999). *Instructional designs in technical areas*. Melbourne: ACER.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Allyn and Bacon.

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics, 28*(1), 113-135.

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System: An International Journal of Educational Technology and Applied Linguistics, 37*(3), 482-495.

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

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

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Van de Branden, K. (2006). *Task-based language education: from theory to practice*. Cambridge: Cambridge University Press.

Van Merriënboer, J. J., Krischner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educational Psychologist*, *38*(1), 5-13.

Van Patten, B. (1985). Communicative value and information processing in second language acquisition. In P. Larson, E. Judd & D. Messerschmitt (Eds.), *On TESOL '84: A brave new world for TESOL* (pp. 89-100). Washington, DC: TESOL.

Van Patten, B. (1999). Attending to form and content in the input. *Studies in Second Language Acquisition, 12*(3), 287–301.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3*(2), 159–177.

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

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(2), 263-281.

Wigglesworth, G. (1997a). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85–106.

Wigglesworth, G., & Elder, C. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly, 7*(1), 1-24.

Willis, J. (1996). *A framework for task-based learning*. Essex: Longman.

Woods, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Process*, *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.

Appendix 1. Pilot study 1.

Simple check-in task. Receptionist's prompt sheet

CHECK-IN 1: PAREJA

Recepcionista

Trabajas de recepcionista en un hotel en Barcelona. Acaba de llegar una pareja y quieren hacer el check-in. Dales la bienvenida y haz las siguientes cosas en el orden indicado.

Eres tú el que empieza la conversación.

Tus tareas

1. *Pregúntale al cliente su nombre y apellidos.*
2. *Pregúntale al cliente su número de reserva.*
3. *Pídele al cliente su carné de identidad.*
4. *El número de habitación del cliente: 409.*
5. *Ofrécele al cliente estos servicios adicionales:*

pensión completa: 30€ al día
desayuno: 10€
horario: 08.00-22.00

acceso 24/7: 15€
una hora: 2€

1 día: 15€
1 semana: 100€

(Pregunta por cada cosa por separado y espera la reacción del cliente.)

La reserva del cliente:

- *una suite en la última planta con vistas al mar.*

Problemas

- Todas están ocupadas. Discúlpate por la situación.

<u>Tus opciones:</u>
- ofrécele al cliente otra habitación (1ª planta); 20% descuento. Estándar parecido)
- ofrécele al cliente una habitación de peor calidad y añade desayunos
- ofrécele al cliente un reembolso de 50%
Tu opción preferida es la primera.

Simple check-in task. Client's prompt sheet

CHECK-IN 1: PAREJA

CLIENTE

Tu y tu pareja acabáis de llegar a un hotel en Barcelona y q queréis hacer el check-in.

El recepcionista empezará la conversación.

1. *Tu número de reserva: BA498605*
2. *Tu carné de identidad:: 65749384D*
3. *Tu decides si quieres aceptar o rechazar los servicios adicionales.*

Tu reserva:

- *una suite en la última planta con vistas al mar. (Todas están ocupadas!)*

Tus opciones:

- *Insiste en que te den la habitación que habías reservado;*
- *Pide un reembolso del 100%.*

No te interesa que te den otra habitación ni un reembolso de 50%.

Complex check-in task. Receptionist's prompt sheet

CHECK-IN 2: FAMILIA

RECEPCIONISTA

Trabajas de recepcionista en un hotel en Barcelona. Acaban de llegar dos familias con hijos (4 adultos y 3 niños) y quieren hacer el check-in. Dales la bienvenida a los clientes y haz las siguientes cosas en el orden indicado.

Eres tu el que empieza la conversación. Hablas con uno de los adultos.

Tus tareas

1 *Pregúntale al cliente su nombre y apellido.*
2 *Pídele al cliente algún documento de identidad.*
3 *Pídele al cliente su número de reserva.*
4 *Habitaciones de los clientes: 206 y 207.*
5 *Ofrécele al cliente los siguientes servicios adicionales:*


(Pregunta por cada cosa por separado y espera la reacción del cliente.)

La reserva del cliente:

-dos habitaciones que dan a la calle principal
- en cada habitación hay dos camas desmontables
- según el cliente, los niños tienen 15% de descuento.

Problemas:

- *Todas las habitaciones que dan a la calle principal están ocupadas.*
- *No hay camas desmontables disponibles en este momento.*
- *El descuento para niños es solamente en temporada baja.*
- *No hay posibilidad de reembolso.*

Complex check-in task. Client's prompt sheet

CHECK-IN 2: FAMILIA

CLIENTE

Sois una familia con un hijo y otra familia con dos hijos. Acabáis de llegar
a un hotel en Barcelona y queréiss hacer el check-in.

El recepcionista empezará la conversación.

1 *Dale al recepcionista tu nombre y apellidos.*
2 *Tu número de reserva: CA439839453*
3 *Tu carné de identidad: 34573967B*
4 *Decide tu mismo/a si quieres aceptar o rechazar los servicios adicionales.*

Tu reserva

- dos habitaciones con dos camas desmontables que dan a la calle principal.

- niños: 15% de descuento.

+Complex check-in task. Receptionist's prompt sheet

CHECK-IN 3
UN GRUPO DE INVESTIGADORES
RECEPCIONISTA

Un grupo de diez quiere hacer el check-in. Se alojan para 4 días y participarán en una conferencia que se celebra al lado del hotel. Dales la bienvenida y haz las siguientes cosas en el orden indicado.

Eres tu el que empieza la conversación. Hablas con una de las personas del grupo.

Tus tareas:

1. *Pregúntale al cliente su nombre y apellidos.*
2. *Necesitas su número de carné de identidad.*
3. *Necesitas el número de reserve del cliente.*
4. *Ofrécele al cliente los siguientes servicios adicionales.*

(Pregunta por cada cosa por separado y espera la reacción del cliente.)

La reserva del cliente:

- *habitación individual (x10)*
- *prepagado: la primera noche*

Problema:

Overbooking. Discúlpate por la situación.

Tus opciones

- ✓ *distribuye el grupo en 2 otros hoteles de la misma cadena (ubicación lejos del lugar de la conferencia)*
- ✓ *ofrece alojamiento en un hotel diferente cerca del lugar de la conferencia (10% más caro).*

No hay posibilidad de reembolso.

+Complex check-in task. Client's prompt sheet

CHECK-IN 3:
UN GRUPO DE INVESTIGADORES

CLIENTE

Sois un grupo de 10 investigadores y acabáis de llegar a Barcelona para participar en una conferencia. Queréis hacer el check-in en vuestro hotel.

El recepcionista empezará la conversación.

1. *Tu carné de identidad: 453475389R*
2. *Tu número de reserva: VA342783*
3. *Decide tu si quieres los servicios adicionales.*

Tu reserva:

- *Habitaciones individuales (x10; segunda planta)*
- *Prepagada la primera noche.*

No te interesan los otros hoteles que sugiere el recepcionista. Insiste en un reembolso para todo el grupo.

Simple recommendation-giving task. Receptionist's prompt sheet

LAS ESPECIALIDADES LOCALES

RECEPCIONISTA

Un cliente de viaje de negocios que se ha alojado esta mañana te pide que le recomiendes un sitio para cenar esta noche. **El cliente empezará la conversación.**

Tus tareas:

1. Descríbele las diferentes opciones al cliente.
2. Sugiérele una opción teniendo en cuenta su perfil y las características de los restaurantes.

Tus opciones:

Cocina:	*mediterránea*	**Cocina:**	*española*
Menú de noche:	*35 - 40 €*	**Menú:**	*22€*
Menú de mediodía:	25€	**Menú para niños:**	*15€*
Situación:	*Gràcia*	**Situación:**	*Port Olímpic*
Acceso:	*a pie*	**Acceso:**	*2 líneas de metro*

Críticas...

"Un poco caro pero merece la pena"

Críticas...

"Excelente relación calidad-precio"

Simple recommendation-giving task. Client's prompt sheet

LAS ESPECIALIDADES LOCALES

CLIENTE

Estás en Barcelona de viaje de negocios durante dos días. Viajas solo/a. Te apetece cenar fuera esta noche y le pides al recepcionista una recomendación.

Eres tú el que empieza la conversación.

Tus tareas:

1. Pídele al recepcionista una recomendación.
2. Dile al recepcionista tus preferencias.
3. Cuando el recepcionista haya descrito las opciones, dale la información sobre tu presupuesto.

Tu perfil:

1. Preferencias: una especialidad local.

2. Presupuesto: 20-30€.

Complex recommendation-giving task. Receptionist's prompt sheet

MARISCO

RECEPCIONISTA

Una familia de cuatro personas ha llegado a tu hotel esta mañana. Les gusta que les recomiendes un sitio para comer.

Hablas con uno de los adultos.

El cliente empieza la conversación.

Tus tareas:

1. Descríbele las cuatro opciones al cliente.
2. Sugiérele la opción que más se ajuste al perfil del cliente.
3. Justifica tu decisión.

Tus opciones:

Cocina: *marisco*
Precio: *30€*
Menú para niños: *15€*
Ubicación: *Port Olímpic*
Acceso: *4 paradas de metro de metro*

¡Vistas impresionantes desde el segundo piso!

Cocina: *española*
Precio: *18€*
Menú para niños: *12€*
Ubicación: *Eixample*
Acceso: *2 buses/5 paradas*

Música en vivo a partir de las 9 PM.

Cocina: *marisco*
Precio: *25€*
Ubicación: *Gràcia*
Acceso: *a pie*
No hay menú para niños.

Críticas…
"Los camareros se desviven para tenerte contento"

Cocina: *mediterránea*
Precio: *15€*
Menú para niños: *12€*
Ubicación: *Plaça Catalunya*
Acceso: *5 paradas de metro*

Críticas…
"Comida estupenda y unas raciones muy generosas…
el servicio un poco lento pero en general merece la pena"

Complex recommendation-giving task. Client's prompt sheet

MARISCO

CLIENTE

Sois una familia de cuatro personas y estáis de vacaciones en Barcelona. Habéis hecho el check-in y os gustaría que el recepcionista os recomendara un sitio para comer.

Tu empiezas la conversación.

Tus tareas:

1. Pídele al recepcionista una recomendación.
2. Dale al recepcionista tu preferencia.
3. Cuando el recepcionista haya descrito las opciones, dale el resto de la información.

Tu perfil:

1. Tu preferencia: marisco
2. Los niños no comerán marisco así que es necesario que el restaurante tenga un menú para niños.
3. Tienes un presupuesto reducido y prefieres no gastar demasiado dinero.
4. Acceso: a pie

+Complex recommendation-giving task. Receptionist's prompt sheet

ALGO EXÓTICO

RECEPCIONISTA

Un grupo de 10 compañeros de trabajo de viaje de familiarización se han alojado en tu hotel. Les gustaría que les recomendaras un sitio para cenar esta noche.

El cliente empieza la conversación.

Tus tareas:

1. Descríbele las diferentes opciones al cliente.
2. Recomiéndale al cliente una opción teniendo en cuenta su perfil y las características de los restaurantes.
3. Convence al cliente de que escoja el restaurante ARGENTINO.

Cocina: *argentina*
Precios: *30-35€*
Ubicación: *El Born*
Acceso: *a pie*

Jazz en vivo los fines de semana
Menús para grupos a precios muy accesibles

Cocina: *vietnamita*
Menú: *25€*
Ubicación: *Gràcia*
Acceso: *a pie*

Piano en vivo a partir de las 21.00
¡El servicio es muy ineficiente!

Cocina: *escandinava*
Menú de noche: *28€*
Menú de mediodía: **22€**
Ubicación: *Sarrià*
Acceso: *5 paradas de metro + 10 min. a pie*

No hay actividad cultural pero hay un disco en la planta baja.

Cocina: *mejicana*
Menú: *25€*
Menú para niños: ***18€***
Ubicación: *L'Eixample*
Acceso: *bus y metro*

Descuento para grupos (mínimo diez personas)

Cocina: *griega*
Menú de mediodía: *18€*
Menú de noche: *25€*
Ubicación: *Eixample*
Acceso: *bus o metro*

Proyecciones de películas a partir de las 20.00

Cocina: *marisco*
Menú laborables: *20€*
Menú: *30€*
Ubicación: *Port Olímpic*
Acceso: *2 líneas de metro*

Karaoke a partir de las 22.00

+Complex recommendation-giving task. Client's prompt sheet

ALGO EXÓTICO

CLIENTE

Sois un grupo de 10 compañeros de trabajo. Estáis de viaje de familiarización en Barcelona. Os gustaría cenar fuera esta noche y estáis buscando "algo exótico".

Tus tareas:

1. Pídele al recepcionista una recomendación.
2. Dale al recepcionista tu preferencia.
3. Cuando el recepcionista haya descrito las opciones, dale el resto de la información.

Tu perfil:

1 Preferencia: algo exótico.

2 Presupuesto: 20€ por persona aproximadamente.

3 Os gustaría un sitio que no sólo ofrezca comida sino también alguna actividad cultural (un espectáculo, música en vivo, etc.)

4 A dos personas no les gusta la comida mejicana.

5 Una persona es alérgica al marisco.

6 Una persona utiliza silla de ruedas así que es importante el acceso fácil al sitio.

Appendix 2. Affective variables questionnaire

NAME ______________________
LAST NAME ______________________
TASK ______________________

Please tick the appropriate box on the scale below.

1. This task required ...

no mental effort at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	extremely great mental effort

2. This task was ...

… not difficult at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	extremely difficult

3. During this task I felt ...

... extremely relaxed	☐	☐	☐	☐	☐	☐	☐	☐	☐	not relaxed at all

4. On this task I ...

… did extremely well	☐	☐	☐	☐	☐	☐	☐	☐	☐	did not do well at all

Appendix 3. Simple task

Resolver problemas en la recepción

¿Qué opciones hay?

La situación

Trabajas de recepcionista en el Hotel Verdi en Barcelona. Acaban de llegar tres clientes y les gustaría saber que habitaciones hay en el hotel para escoger la que les guste.

Tu tarea

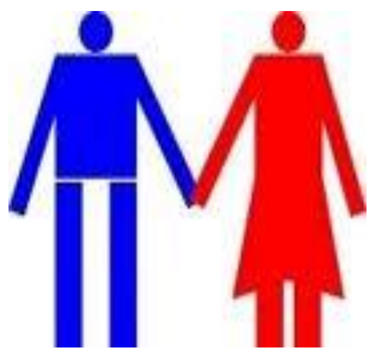
- Explícales a los clientes las diferentes opciones.

Tus recursos

- Un listado de perfiles de clientes que nos suelen visitar
- Un listado de habitaciones disponibles

Algunos de los clientes que nos visitan…

Parejas jóvenes



Las parejas jóvenes son unos de nuestros clientes más frecuentes. Normalmente vienen de vacaciones para algunos días. Suelen tener un presupuesto bajo y por eso les gustan las opciones baratas, pero también

Universitarios

Como nuestro hotel se encuentras a unos pasos de la Universidad, recibimos muchos profesores de fuera. Normalmente están dispuestos a pagar un poco más pero disfrutar de mejor calidad.

Grupos de amigos

Dependiendo de la edad, buscan opciones realmente baratas o prefieren pagar un poco más por buena calidad. Les suelen gustar estudios y suites.

LAS OPCIONES DISPONIBLES

Lujosa

Precio	50€/ noche
Ubicación	3ª planta
Vista	vista del mar
Desayuno	15€
Descuento	desde 5 noches

Instalaciones

- mini-bar
- aparcamiento gratis
- caja de seguridad
- teléfono
- cuarto de baño
- secador

Estándar

Precio	15€/ noche
Ubicación	planta baja
Vista	a la calle
Desayuno	no incluido
Oferta	-

Instalaciones

- TV
- internet
- nevera
- secador
- cuarto de baño
- teléfono

Guest room

Precio	25€/ noche
Ubicación	2ª planta
Vista	al jardín
Desayuno	incluido
Descuento	desde 5 noches

Instalaciones

- internet
- nevera
- TV
- cuarto de baño
- teléfono
- secador

Suite

Ubicación última planta
Precio 35€/ noche
Vista al jardín
Desayuno no incluido
Oferta a partir de 7 noches, 10%
Instalaciones

- internet
- nevera
- TV
- cuarto de baño
- teléfono
- secador

Estudio

Ubicación última planta con terraza
Precio 80€/noche
Vista horizonte urbano
Desayuno incluido
Oferta descuento en la temporada baja
Instalaciones

- cocina
- wi-fi
- TV
- cuarto de baño
- aire acondicionado
- aparcamiento gratuito

Appendix 4. Complex task

Resolver problemas en la recepción

¿Qué ha pasado con mi habitación?

La situación

Trabajas de recepcionista en el Hotel Verdi en Barcelona. Ha habido un problema y algunos clientes tendrán que ser trasladados a otras habitaciones. Haz lo siguiente:

Tu tarea

- Discúlpate por la situación. Sé muy amable, ¡no quieres perder los clientes!
- Explica las diferentes habitaciones donde se pueden trasladar los clientes.
- Recomiéndale una opción a cada cliente. Puedes ofrecer la misma solución a más de un cliente.

Tus recursos

- Un listado de tus clientes
- Un listado de habitaciones disponibles

L O S   P E R F I L E S   D E   T U S   C L I E N T E S

¡Recuerda que las habitaciones que habían reservado no están disponibles!

Compañeros de trabajo

Reservaron *3 suites en la última planta*
Presupuesto *100€/noche*
Vista *al jardín*

Comidas *pensión completa*

Un grupo de estudiantes

Reservaron *3 habitaciones económicas*
Presupuesto *30€/noche*
Vista *piscina*

Comidas -

Un profesor universitario

Reservó *guest room*
Presupuesto *65€/noche*
Vista *vista al mar con terraza*

Comidas -

LAS OPCIONES DISPONIBLES

SUITE

Ubicación	1ª planta
Precio	100€/noche
Vista	piscina
Comidas	desayuno

ESTÁNDAR

Ubicación	todas las plantas
Precio	40€/noche
Vista	a la calle
Comidas	desayuno 10€

DOBLE

Ubicación	3ª planta
Precio	50€/noche
Vista	a la calle
Comidas:	-

DELUXE

Ubicación	última planta
Precio	120€/noche
Vista	vista al mar con terraza
Comidas	desayuno

JUNIOR SUITE

Ubicación	2ª planta
Precio	80€/noche
Vista	al jardín
Comidas	-

Appendix 5. +Complex task

Resolver problemas en la recepción

¿Sabes afrontar una crisis?

La situación

Trabajas de recepcionista en el Hotel Verdi en Barcelona. Ha habido overbooking y varios clientes tienen que ser trasladados a otros hoteles. Eres responsable de trasladar a 3 clientes.

Tus tareas

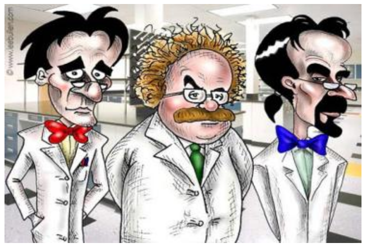
- Discúlpate por la situación. Sé muy amable, ¡no quieres perder los clientes!
- Explica las diferentes opciones disponibles. ¡Asegúrate de informar bien a los clientes!
- Recomienda la solución que te parece mejor según los perfiles de los clientes y la disponibilidad de los hoteles.
- Justifica tus soluciones. ¿Por qué la opción que acabas de recomendar es la mejor?

Tus recursos

- Un listado los perfiles de tus clientes
- Un listado de hoteles disponibles

LOS PERFILES DE TUS CLIENTES

Un grupo de investigadores



Reservaron *6 habitaciones individuales*
5 de octubre

Precio *35€/ habitación individual*
Duración de estancia *3 noches*
Petición *habitaciones en la misma planta*
De este hotel les gusta...
Ubicación (el centro)
Internet gratis las 24 horas
Desayuno incluido

Una familia de cuatro

Reservaron *1 habitación doble con*
2 camas desmontables
1 de octubre
Precio *70€/ 4 personas*
Duración de estancia *1 semana*
Petición *habitación en la última planta con vistas al mar*
De este hotel les gusta...
Ubicación (el centro)
piscina
Paseo de 10 min hasta
el transporte público

Un hombre de negocios

Reservó *suite en la última planta*
1 de octubre

Precio *140€/ noche*
Duración de estancia *2 noches; posiblemente más*
Petición *terraza y vista al mar*
De este hotel le gusta...
Ubicación (el centro)
wi-fi gratis
aparcamiento gratis
Información adicional *cliente habitual*
(5 visitas previas)

T U S   O P C I O N E S

PLAZA HOTEL

Ubicación	*suburbios*
Transporte público	*paseo de 20 min*
Precios	*45€/ hab. Indiv.* *60€/hab. doble* *70€/estudio de 3 camas* *no hay suites*
Wi-fi	*10€/ todo el día*
Vista al mar	*+10€ todas las habitaciones*
Terraza	*solamente hab. en la última planta*
Aparcamiento	*no hay*
Piscina	*no*
Comidas	*no incluidas*
Disponibilidad	*5 hab. ind. : a partir del 7 de octubre* *doble: disponible* *estudio: disponible ahora*

AVENIDA DE ESPAÑA

Ubicación	*el centro*
Transporte	*paseo de 5 min*
Precios	*50€/hab. indiv.* *100€/ estudio familiar* *75€/hab. doble* *170€/suite*
Wi-fi	*gratis*
Vista al mar	*no hay*
Terraza	*en algunas habitaciones*
Aparcamiento	*gratis*
Piscina	*no hay*
Comidas	*no incluidas*
Disponibilidad	*todas disponibles ahora*

CATALONIA HOTEL

Ubicación	*cerca del centro*
Transporte	*paseo de 10*
Precios	*50€/ hab. Indiv.* *69€/ hab. doble* *no hay estudio* *130€/suite*
Wi-fi	*2€/ hora*
Vista al mar	*no disponible ahora*
Terraza	*cargo adicional*
Parquing	*15€/día*
Piscina	*sí*
Comidas	*Desayuno incluido*
Disponibilidad	*todas disponibles ahora* *estudio disponible para 5 días*

HOTEL W

Ubicación	*al lado del mar*
Transporte público	*2 líneas de metro cerca*
Precios	*285€/noche/hab. doble* *335€/noche/King-size bed* *2240€/estudio para 4 personas*
Wi-fi	*gratis*
Vista al mar	*la mayoría de las habitaciones*
Terraza	*algunas habitaciones*
Aparcamiento	*sí*
Piscina	*sí*
Comidas	*desayuno*
Disponibilidad	*todas disponibles ahora*

EMPERADOR HOTEL

Ubicación	*suburbios*
Transporte público	*autobús: 15 minutos hasta el centro*
Precios	*40€/habitación individual* *70€/hab. doble* *140€/ estudio*
Wi-fi	*10€/24h*
Vista al mar	*no hay*
Terraza	*algunas habitaciones*
Aparcamiento	*20€/24h*
Piscina	*ahora no disponible*
Comidas	*no incluido*
Disponibilidad	***e**studio disponible durante 6 días* *Hab. individual y doble disponible*

Camas desmontables disponibles en todas las habitaciones por un cargo adicional de 10€.

Appendix 6.

Second pilot study
Time judgment task and affective variables questionnaire

Time judgement task

How long did it take you to complete the task?

Estimated time

__________ in minutes and _________ seconds

Real time

____

Affective variables questionnaire

Please tick the appropriate box on the scale below.

5. This task required ...

… no mental effort at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	extremely great mental effort

6. This task was ...

… not difficult at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	extremely difficult

7. During this task I felt ...

… extremely relaxed	☐	☐	☐	☐	☐	☐	☐	☐	☐	not relaxed at all

8. On this task I ...

… did extremely well	☐	☐	☐	☐	☐	☐	☐	☐	☐	did not do well at all

Appendix 7. Context and general instructions to all tasks

Bienvenidos a la sesión de entrenamiento "Resolver problemas en la recepción"

- ***¿De qué se trata?***

Imagínate que acabas de comenzar a trabajar como recepcionista en un hotel. Hoy es tu segundo día de trabajo.

El hotel donde trabajas, el hotel Verdi de Barcelona, es famoso por su trato personal con el cliente. Por este motivo, según la nueva política del hotel, los trabajadores tienen que pasar una o más sesiones de entrenamiento para practicar sus habilidades y estar más seguros de si mismos a la hora de tratar con los clientes. La sesión de hoy trata sobre resolver problemas en la recepción y ¡es tu turno!

- ***¿Qué es lo que tienes que hacer?***

Harás una tarea o tres tareas en las cuales tendrás que tratar con los clientes. Cada tarea tiene la misma estructura:

(1) las instrucciones
(2) la información sobre los clientes
(3) la información sobre las habitaciones y/o los hoteles.

Para cada tarea tendrás 1 minuto para estudiar los puntos 1-3. Pasado un minuto, empezarás a hacer la tarea. Te diré cuando tienes que empezar.

Si tienes cualquier duda, pregúntamelo a mi. Cuando estés preparado, empezaremos la primera tarea.