



UNIVERSITAT<sub>DE</sub>  
BARCELONA

## Second Language Anxiety and Task Complexity

Kerry Anne Brennan



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- Compartigual 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - Compartigual 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-ShareAlike 3.0. Spain License.**



UNIVERSITAT DE  
BARCELONA

Second Language Anxiety and Task Complexity  
Tesi doctoral presentada per  
**Kerry Anne Brennan**

Director:  
Dr. Roger Gilabert Guerrero

Programa de doctorat:  
*Estudis Lingüístics, Literaris i Culturals*  
Departament d'Estudis Anglesos i de Llengües i  
Literatures Modernes

Universitat de Barcelona  
2016



## ACKNOWLEDGEMENTS

If I were to individually thank everyone who has played a part throughout this amazing journey, I could probably fill another one hundred pages. It has been an incredible process which has taught me so much. Also, I have experienced the concept of anxiety firsthand, both in a ‘facilitating’ and a ‘debilitating’ way.

I would like to thank all the experts in the field who have shared their wealth of knowledge with all of us. First and foremost thank you to my dear supervisor, Dr. Roger Gilabert, his encouragement and advice throughout this process has been invaluable; Dr. Andrea Révész, has been a tremendous support both through her work and her advice at different points throughout this project; Dr. Joan Carles Mora for his constant dedication to the doctoral program, despite being met with many roadblocks, and especially for making it possible for me to continue in this program at this time; Dr. Maria Luz Celaya for introducing me to the world of Applied Linguistics at UB; Dr. Elsa Tragant and Dr. Imma Miralpeix for giving me feedback on my initial ideas; Dr. Luke Plonksy for giving me ideas on different ways to look at my data; Dr. Kazuya Saito whose long and interesting discussions have helped me to see the ‘other side’ of academic life. And finally, to those who have had the biggest effect on this dissertation: Dr. Peter Robinson and Dr. Peter Skehan, without their interesting research and dedication to the world of SLA, this dissertation would not have been possible.

I would like to thank my colleagues from the University of Barcelona for all of their emotional support and academic discussions: Dr. Alexandra Malika; Dr. Gemma

*Second Language Anxiety and Task Complexity*

Artieda; (soon to be!) Dr. James Pownell; Dr. Jessica Mackay (for comments on my paper); Lena Vasylets (for assisting with data collection), and Dr. Mayya Levkina (for helping with participants). To my work colleagues and friends: Niamh, Chris and Martina, you have really been there for us in order to allow me to take ‘a break’ so that I could pursue this dream of mine. Thanks to Kathy de la Cruz for all of her technical support and help during the piloting stages of this project. Also, to everyone in London who made my ‘study abroad’ experience so special especially to: Liz, Mike, Jordi, Kevin, Michele, and Merima. I am also grateful to all the participants who helped make this study possible: the University of Barcelona students; those from Philadelphia, and all of the many others who helped me during the piloting stages of this project.

And to Jaime and Iñaki, who believed in me from the very start of this project even when I didn’t, and they have continued to do so spiritually, especially during the final stages.

I would like to thank my family for having always encouraged me and my undying need to learn more and explore the world. I am especially grateful to my mother for her constant support and tips on how to avoid negative anxiety, my father for all of his technical advice and the never-ending support he has provided in so many ways, Cynthia for comments made on sections of this paper, Wendy for reading me off all the names on the lists a million times, and finally, to Adam for all of his help with the piloting.

Finally, this dissertation is dedicated to my wonderful partner in life, Stephen, who has supported me and encouraged me throughout every step of this long and challenging process. I wouldn’t have been able to do it without him.

## SUMMARY

Throughout history there has been tremendous debate about what comprises the best methods and approaches to foreign or second language (L2) teaching and learning, without any decisive universal conclusions. Since the 1980s, however, task-based approaches to language teaching have offered what theoretically appears to be the most efficient option for adapting to the plethora of learner needs that may affect second language acquisition (SLA). Task-based syllabi have been based on the premise that task design should initiate from the learners' 'needs-analysis' based on which 'real-life' tasks are the most applicable to their situation (Long & Crookes, 1992; Long 2005). Once the needs are determined, pedagogic tasks and syllabi are subsequently designed and progressively sequenced in a more cognitively demanding manner which should lead to SLA.

However, the ongoing debate in task-based research and pedagogy has been the lack of a collective standard to effectively and systematically determine the best way to design, select, and implement pedagogic tasks in order to best facilitate L2 language learning and acquisition. SLA researchers have rigorously attempted to address this issue, and as a result, there have been promising findings. The task-based sequencing frameworks proposed by Robinson (2001, 2003, 2005, 2007; Robinson and Gilabert, 2007; Robinson 2011, 2015) and Skehan (1996, Skehan & Foster, 1999, 2001; Skehan, 2003, 2009) have both posited important theoretical and practical methods of selecting and implementing L2 tasks in order to promote SLA. As a result, their models have gained much attention in SLA research throughout the last decade.

*Second Language Anxiety and Task Complexity*

This dissertation describes some of the main tenets of these models, which includes giving importance to the effects that L2 task design features, contextual factors, and learner factors have on successful SLA. However, the main focus will be on how they define task design features and the effects these have on learner factors. Furthermore, it will be highlighted here that neither model has fully clarified how to address foreign language anxiety (FLA) as affected by task design features. FLA has been a widely researched variable in SLA, most frequently focusing on the impact that it has on L2 performance and acquisition, yet there are few studies conducted specifically in the task-based context. Therefore, the experimental study described in this dissertation set out to contribute to this gap in the research.

Thus, the aim of this study was twofold. First, the objective was to explore the effects of whether the L2 tasks in terms of task-type, task complexity, or sequencing of tasks, had any effect on the participants' state anxiety levels. The second objective was to explore whether state anxiety could be identified during task-based oral performance by using a scale designed for the current study. There were 108 EFL participants from the University of Barcelona who took part in this mixed-methods design study. First, the overall general FLA scores were determined through a widely used instrument: the Foreign Language Classroom Anxiety Scale (FLCAS; Horwitz, Horwitz, & Cope, 1986). Next, the state anxiety levels were rated through the use of two Likert scales that were adapted and designed for this study: The Anxometer and the Task Anxiety Reflection Scale (TARS). The TARS also included open-ended questions to add qualitative support to the findings. The results indicated that the state anxiety scales were successful at analyzing how the students felt during L2 oral task performance that had been

*Second Language Anxiety and Task Complexity*

manipulated in terms of task-type, task complexity, and sequencing. Additional results indicated that the FLCAS scale only mildly predicted the state anxiety ratings. These results have extended the research into how L2 task features, based on sequence and complexity, affect learners' state anxiety levels. Further studies are needed to examine what effects these findings have on task-based performance and acquisition.



## RESUM

Al llarg de la història ha existit un gran debat sobre quins són els millors mètodes i enfocaments per a l'aprenentatge i ensenyament de segones llengües o llengües estrangeres, un debat que no ha arribat encara a cap conclusió universal. Des dels anys 80, l'enfocament basat en tasques ha ofert el a nivell teòric és una opció eficient que s'adapta a tota una sèrie de necessitats per part dels aprenent susceptibles d'afectar de manera positiva els processos d'adquisició de segones llengües. Els programes basats en tasques han partit de la premissa de que el disseny de tasques s'hauria de començar a partir de l'anàlisi de les necessitats respecte de les tasques reals que millor s'adaptin a la situació dels estudiants (Long & Crookes, 1992; Long 2005). Un cop es determinen aquestes necessitats, es poden dissenyar programes en que les tasques es seqüencien de manera progressivament més complexa, el que pot portar potencialment a l'adquisició de segones llengües.

Tanmateix, un dels debats dins l'àmbit de la recerca i la pedagogia basades en tasques ha estat la manca d'uns estàndards col·lectius per determinar de manera efectiva i sistemàtica la millor manera de dissenyar, seleccionar i implementar les tasques de manera que facilitin l'aprenentatge i adquisició llengües estrangeres. La recerca en adquisició de llengües han intentar resoldre aquesta qüestió de manera rigorosa que ha portat a resultats prometedors. Els marcs per a la seqüenciació de tasques proposat per Robinson (2001, 2003, 2005, 2007; Robinson i Gilabert, 2007; Robinson 2011, 2015) i Skehan (1996, Skehan & Foster, 1999, 2001; Skehan, 2003, 2009) han postulat mètodes importants a nivell teòric i pràctic per a seleccionar i implementar tasques que portin a

l'adquisició de segones llengües. Aquests mètodes han estat centrals en la recerca en el camp de l'adquisició de la darrera dècada.

Aquesta tesi doctoral descriu els principals arguments d'aquests models que posen l'èmfasi sobre el efectes que la manipulació del disseny de les característiques de les tasques, el factors contextuals, i els factors dels aprenents poden tenir sobre una adquisició de segones llengües exitosa. Aquest treball es focalitza en com aquest models defineixen les característiques de disseny de les taques i el seus efectes sobre el factors del aprenents. De manera més concreta, el treball destaca com cap dels models ha aclarit com tractar l'ansietat respecte de les llengües estrangeres i com aquest queda afectada per les característiques del disseny de les taques. L'ansietat respecte de les llengües estrangeres ha esta estudiada àmpliament en el camp de l'adquisició de segones llengües en relació a l'impacte que té sobre l'ús i adquisició de segones llengües, sense que s'hagi estudiat en el context de l'aprenentatge basat en tasques. Aquest estudi experimental es planteja cobrir aquesta mancança en aquest camp de recerca.

L'objectiu d'aquest estudi es doble. D'una banda s'investiga els efectes que la seqüència de les tasques, en termes de tipus de tasca i complexitat cognitiva de la tasca, pugui tenir sobre el nivells d'ansietat dels participants. D'una altra banda, s'explora si els nivells d'ansietat poden ser detectats per una escala dissenyada específicament per al context de les tasques. 108 aprenents d'anglès com llengua estrangera de la Universitat de Barcelona participen en un estudi amb un disseny metodològic mixt. En primer lloc, es determinen els nivells d'ansietat general mitjançant un instrument àmpliament utilitzat a la recerca com és el FCLAS (Escala d'Ansietat a l'Aula de Llengües Estrangeres - Horwitz, Horwitz, & Cope, 1986). En segon lloc, es mesuren els nivells d'ansietat

*Second Language Anxiety and Task Complexity*

específica en la realització de la tasca mitjançant dues escales de Likert adaptades i dissenyades per a aquest estudi: l'Ansiometre i l'Escala de Reflexió sobre l'Ansietat de la Tasca (TARS). L'escala TARS també inclou preguntes obertes que aporten un suport qualitatiu als resultats. Els resultats indiquen que les escales d'ansietat específiques discriminen de manera efectiva els diferents nivells d'ansietat durant la realització de tasques el disseny de les quals ha estat manipulat a nivell del tipus de tasques, la seva complexitat cognitiva, i la seva seqüència. Anàlisis addicionals indiquen que l'escala d'ansietat general FLCAS només prediu de manera mínima a moderada el nivell d'ansietat específica durant la realització de les tasques. Aquests resultats aporten més llum sobre com les característiques de les tasques, basades en la seva seqüenciació i la seva complexitat, afecten el nivell d'ansietat dels aprenents de segones llengües. L'estudi conclou tot remarcant la necessitat de realitzar més estudis que investiguin la relació entre aquests resultats i els resultats lingüístics derivats de la realització de tasques i adquisició.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> .....	<b>ii</b>
<b>SUMMARY</b> .....	<b>iv</b>
<b>LIST OF FIGURES</b> .....	<b>xiii</b>
<b>LIST OF APPENDICES</b> .....	<b>xiv</b>
<b>KEY OF ACRONYMS USED IN THIS STUDY</b> .....	<b>xv</b>
<b>PART I: LITERATURE REVIEW</b> .....	<b>19</b>
<b>CHAPTER I. Task-based sequencing models.</b> .....	<b>20</b>
1.0. Introduction.....	20
1.1. An overview of models and approaches for foreign language teaching. ....	21
1.2. Operationalizing and selecting L2 tasks for task-based research. ....	26
1.3. L2 Task Classification, Sequencing and Syllabus Design Models.....	31
1.3.1. Peter Robinson’s models for L2 task design and sequencing.....	34
1.3.2. Peter Skehan’s L2 task-based framework. ....	38
1.4. Studies testing the theoretical task-based sequencing models.....	42
1.5. Individual differences and L2 task-based sequencing models. ....	47
<b>CHAPTER II. Individual Differences in Second Language Learning: How is Anxiety Identified?</b> .....	<b>53</b>
2.0. Introduction.....	53
2.1. Individual Differences (IDs) in SLA: Where does anxiety fit in?.....	54
2.2. Defining the construct of general anxiety: a psychological viewpoint. ....	60
2.3. Defining the construct of anxiety in the L2 context. ....	61
2.4. How has anxiety been measured in the foreign language context? .....	65
2.4.1. Foreign Language Classroom Anxiety Scale: (FLCAS; Horwitz et al. 1986). ....	67
2.4.2. Input Processing Output Anxiety Scale: (IPOAS; MacIntyre & Gardner, 1994). ....	70
2.4.3. Anxiety in different contexts within SLA.....	73
2.4.4. Task-based sequencing models and anxiety.....	74
<b>CHAPTER III. Purpose of study: Research Questions and Hypotheses</b> .....	<b>79</b>
3.1. The main focus of the current study. ....	79
3.2. Research questions and hypotheses. ....	83

<b>CHAPTER IV: Methodology and Design.....</b>	<b>86</b>
4.1. Introduction.....	86
4.2. Operational definitions of the variables.....	86
4.3. Creating the Task Anxiety Reflection Scale: TARS.....	90
4.4. The Piloting Stage.....	91
<b>CHAPTER V: The Experimental Study.....</b>	<b>97</b>
5.1. Study Design.....	97
5.2. Participants.....	97
5.3. The Instruments Used for the Current Study.....	100
5.4. Procedure.....	108
5.5. Scoring Procedures for Anxiety Scales.....	111
5.6. Statistical Analyses.....	111
<b>PART III: THE RESULTS.....</b>	<b>114</b>
<b>CHAPTER VI: The Analyses .....</b>	<b>115</b>
6.1. Introduction.....	115
6.2. Descriptive exploratory analysis.....	115
6.3. Results for research questions 1.....	117
6.4. Results for research question 2.....	128
6.5. Qualitative analysis.....	131
6.6. Additional analysis: Factor analysis of TARS.....	134
6.7. Additional analyses: Correlation of anxiety scales.....	136
<b>PART IV: DISCUSSION AND CONCLUSIONS.....</b>	<b>138</b>
<b>CHAPTER VII: Discussion .....</b>	<b>139</b>
7.1. Introduction.....	139
7.2. Addressing results to research question 1.....	140
7.3. Addressing results to research question 2.....	146
7.4. Additional results and justification for further analysis.....	147
<b>CHAPTER VIII: CONCLUSIONS.....</b>	<b>151</b>
8.1. Limitations.....	151
8.2. Implications for future research.....	152
8.3. Pedagogical implications.....	154
8.4. Summary of findings.....	156

8.5. Conclusions .....	157
<b>PART V: APPENDICES .....</b>	<b>168</b>

## LIST OF TABLES

Table 1. Classifications of task-types .....	32
Table 2. The Triadic Componential Framework .....	36
Table 3. The SSARC model for task-based sequencing. ....	38
Table 4. Skehan's Criteria for Task Difficulty. ....	39
Table 5. Task Complexity Tasks as Target (TAT) Studies. ....	45
Table 6. Complexity and Individual differences TAT studies. ....	46
Table 7. Summary of the main facets of Skehan's and Robinson's models .....	52
Table 8. Examples of Task Complexity studies using tools to assess IDs. ....	78
Table 9. Descriptive statistics of the participants. ....	100
Table 10. Task Sequence format used in study. ....	109
Table 11. Summary of Study design. ....	110
Table 12. Descriptive statistics of all TARS. ....	116
Table 13. Descriptive Statistics of all Anxometers. ....	117
Table 14. One-way ANOVA for TARS Scores. ....	119
Table 15. One-way ANOVA of the Anxometer Scores. ....	120
Table 16. Pairwise Comparison of Complexity: TARS Scores .....	121
Table 17. Pairwise Comparison of Complexity: Anxometer Scores .....	121
Table 18. Descriptive statistics of TARS scores according to sequence .....	123
Table 19. Descriptive statistics of all FLCAS. ....	129
Table 20. Examples of participants' comments from the TARS: SEQ B .....	132

Table 21. Pattern Matrix: TARS .....	135
Table 22. Correlation of all Anxiety scores.....	137
Table 23. Examples of participants' comments. ....	141
Table 24. Table Extraction Sums of Squared Loadings .....	147
Table 25. Quantitative and qualitative comparison for task 1 SEQ D .....	149

## LIST OF FIGURES

Figure 1. Options in Language Teaching. ....	22
Figure 2. Examples of percentage scores from FLCAS. ....	69
Figure 3. Relationship between anxiety, cognition, and behavior. ....	71
Figure 4. Yerkes and Dodson .....	77
Figure 5. Causal constructs of state anxiety included on TARS. ....	91
Figure 6. Examples of items from the FLCAS. ....	102
Figure 7. The Anxometer, adapted for this study.....	103
Figure 8. Example items from the TARS. ....	104
Figure 9. Causal constructs of state anxiety included in TARS scale. ....	105
Figure 10. Map tasks .....	106
Figure 11. Fire Chief tasks.....	107
Figure 12. Boxplots of the TARS scores for all tasks.....	119
Figure 13. Boxplots of the Anxometer scores. ....	120
Figure 14. Boxplots for Map Tasks: Anxometers and TARS .....	124
Figure 15. Boxplots for Fire Chief Tasks: Anxometers and TARS .....	125
Figure 16. Repeated Measures ANOVA for all four TARS.....	126

Figure 17. Repeated Measures ANOVA for all four Anxometers .....	126
Figure 18. Boxplots for FLCAS means scores by sequence. ....	129
Figure 19. Histograms of FLCAS by sequence.....	130
Figure 20. Scree Plot Factor Analysis: TARS .....	136
Figure 21. Comparison graphs of the Anxometer and TARS: Sequencing.....	143

## LIST OF APPENDICES

Appendix 1. Task as Target (TAT) Studies: Planning and Task Repetition. ....	169
Appendix 2. FLCAS-Castellano .....	171
Appendix 3. Anxometer and TARS from the pilot phase. ....	173
Appendix 4. Item analysis Correlation Matrix from Pilot of Spanish TARS .....	175
Appendix 5. The Task Anxiety Reflection Scale (TARS) .....	176
Appendix 6. Participants' Comments on the TARS: SEQ A .....	179
Appendix 7. Participants' Comments on the TARS: SEQ C .....	180
Appendix 8. Participants' Comments on the TARS: SEQ D .....	182



## KEY OF ACRONYMS USED IN THIS STUDY

CAF: Complexity, Accuracy, and Fluency; measurement of speech production.

CH: The Cognition Hypothesis

FLA: Foreign Language Anxiety

FLCAS: Foreign Language Classroom Anxiety Scale

IDs: Individual Differences

ISLA: Instructed Second Language Acquisition

LCM: Limited Capacity Model

L2: Throughout this study this will refer to any language other than the first language.

SLA: Second Language Acquisition

TARS: Task Anxiety Reflection Scale

TAT: Task as Target; research studies that give attention to features of L2 tasks within the research design.

TBLT: Task-based language teaching

TCF: Triadic Componential Framework

TOH: Trade of Hypothesis



## **PART I: LITERATURE REVIEW**

## **CHAPTER I: Task-based sequencing models.**

### **1.0. Introduction.**

This chapter begins with a brief overview of the principal teaching models and approaches that have been utilized in foreign and second language (L2) education throughout history. This summary works as a starting point for the reader to gain an understanding of the origins of the current research project. The discussion will lead to the strengths and weaknesses that arise when offering a task-based model as a solution for catering to L2 learners' individual needs. In the following sections, summaries of what has been proposed by Peter Robinson and Peter Skehan to resolve some of the shortcomings found in many task-based contexts will be described. Moreover, an outline will be given about the perspective of these researchers in terms of the interaction of task features and learner cognition. Finally, how each of their models considers learners' individual differences will be addressed.

The objective here is to highlight how the implications made from these theoretical models have intended to aide in the process of selecting and implementing tasks as pedagogic units, and then, sequencing them in a manner which optimizes L2 performance and acquisition. Finally, this chapter will conclude with how the tenets of these proposed models have been tested thus far, and consequently, what have been the main findings and gaps in the literature at this time. As such, aspects of these theoretical frameworks and empirical findings have guided the experimental study presented here. Therefore, this study has

intended to help contribute to this line of research by assessing the affective factor of anxiety in the context of L2 task-based complexity research.

The following main points will be included in this chapter:

1. An overview of models and approaches for foreign language teaching.
2. Suggestions for selecting, designing, and sequencing L2 tasks.
3. Models for L2 task sequencing and syllabus design.
4. A brief review of studies that have been influenced or guided by the suggested models and identifying gaps in the research.
5. Implications of how these models account for the learners' individual differences in SLA.

### **1.1. An overview of models and approaches for foreign language teaching.**

Throughout history, there have been a myriad of ways in which foreign language programs have been designed and executed around the world. To sum up the most common methods and approaches that have been developed throughout the years, included here is the pendulum of 'options in language teaching' (Figure 1; adapted from Long & Robinson, 1998). The concept of a pendulum helps illustrate the idea that foreign language teaching methods/approaches have had the tendency to fluctuate from one extreme to the other. Moreover, those that have intended to fall somewhere in the middle often shift back and forth, albeit slightly, incorporating aspects from either end of the spectrum. Often, these methods/approaches have stemmed from theoretical or practical findings and

some have been further supported (or discredited) by empirical or classroom-based research. Each option has been assumed to guide ‘best practices’ within the L2 classroom and throughout the L2 language learning process. However, it could be argued that many of these options have their strengths and weaknesses, some of which will be explored here.

Figure 1. Options in Language Teaching.

<u>Option 2</u>	<u>Option 3</u>	<u>Option 1</u>
<i>Analytic</i> Focus on Meaning (FoM)	<i>Analytic</i> Focus on Form (FonF)	<i>Synthetic</i> Focus on Forms (FoFs)
Natural approach Immersion	Task-based approaches (e.g., TBLT) Content-Based Teaching (e.g., CLIL)	Grammar Translation Audiolingual Method Silent Way Total Physical Response
(Procedural Syllabus, etc.)	(Process Syllabus, etc.)	(Structural/Notional- functional syllabuses, etc.)

*(Adapted from Long & Robinson, 1998, p. 16).*

The main problem with this ‘pendulum effect’ is that if the tendency is to fluctuate between extremes, it is difficult to confidently support choosing one method or approach over another. However, realistically, the way in which the L2 classroom syllabus or research design is structured is affected by many factors and influences such as personal preferences of those who are running the programs, financial budgets, the type of training that occurs, current trends, governmental or cultural support or beliefs, to name a few. Therefore, in order to select the most suitable language course structure or research design for the

population and context in which it is serving, those involved in the decision making process should be well-informed of the various options available.

Figure 1 (p.22) displays a three-way distinction which includes examples of some of the major teaching methods and approaches in foreign language education, found both historically as well as in modern day classrooms. On the right side of the chart (Figure 1, Option 1), we find what are considered to be the most traditional ones, or those which have become known as *focus on forms* (FoFs) approaches. These are predominantly language classrooms composed of *synthetic methods* (Wilkins, 1976) of language learning. In other words, the language learning syllabi have been based on some form of linear and accumulative progression of, for example, linguistic ‘structures’ or ‘notions’ (Long & Robinson, 1998).

Conversely, stemming from what some experts in the field have argued were weaknesses in those offered under Option 1, the teaching methods/approaches swung in the opposite direction (Figure 1, Option 2). The main concern was how ‘unnatural’ the Option 1 methods seemed to be in practice, and instead, they began to base language teaching models on what they considered to be a more ‘natural way’ of learning the target language. The claim was that with enough exposure to what was considered to be ‘comprehensible input’ of the target language (e.g., Krashen, 1985), in combination with more practice in language output, the L2 would be acquired in a similar way to that of first language development. Wilkins (1976) referred to the syllabi that adhered to this philosophy as *analytic* in structure (Long & Robinson, 1998).

Finally, arising from what were found to be inefficiencies with this format of language learning in the foreign language classroom setting, Option 3 was formed. For the most part, these new methods/approaches grew from the lack of clarity and structure in adequately supporting student development of certain L2 linguistic forms. Hence, the pendulum swung back towards Option 1, where once again, attention was paid to the specific structures that needed to be addressed in the L2. Therefore, Option 3 eventually became referred to as the *focus on form* (FonF) methods and approaches (Doughty, 2001).

Based upon SLA research findings at that time, it was deemed that a more successful language course structure was dependent on including components from both Option 1 and Option 2. Option 3 approaches, therefore, were to include ‘real-world’ communicative language situations, whilst at the same time utilizing pedagogic materials that focused on both L2 input and output. Concurrently, the curriculum carefully needed to highlight the target L2 linguistic structures, either implicitly or explicitly, at some point throughout the course. Despite these developments, the pendulum continued to swing. (See e.g., Doughty 2001; Doughty & Williams, 1998; Robinson & Long, 1998; for further discussion on this topic).

To sum up, when making the choice to implement any L2 teaching method or approach, it seems fundamental that the following three questions should be addressed:



1. What are the most efficient units of analyses and/or pedagogic units to meet the needs of the context or population being served?
2. How are these units effectively sequenced to form an L2 teaching syllabus?
3. How should the outcome or success be measured or analyzed?

Furthermore, apart from these three basic points, presumably there should be a theoretical framework guiding these choices. Additionally, theory must be eventually backed by both empirical and classroom-based research as well as qualitative feedback from all those involved (i.e., the researchers, teachers and learners).

Therefore, if a FonF (Option 3) approach is adopted, either in L2 research or pedagogy, those involved must first decide what the unit of analysis will be and/or how to structure the syllabus that will be followed. In contrast with FoFs (Option1), where some linguistic structure or ‘notion’ is deemed the focal point of any L2 program, task-based approaches propose using ‘pedagogic tasks’ as the units for creating a syllabus. Herein lies the first challenge of this type of approach: clearly defining what is meant by *task* in the context in which it is being used. Thus, an operational definition of task should be identified based on a theoretically and empirically proved framework. Accordingly, set guidelines should then be employed for L2 task selection, implementation and sequencing decisions (Van den Branden, 2006).

As applied linguists following a task-based approach, we need to clarify what is meant by ‘real-world’ tasks, both inside and outside the language classroom. As the current study has been guided by this approach, an analysis of how task has been operationalized will be addressed based on leading research in this area. Thus, we will consider some ways in which tasks have been defined, categorized, and utilized in the SLA literature thus far.

## **1.2. Operationalizing and selecting L2 tasks for task-based research.**

Throughout the past several decades, the use of L2 pedagogic tasks has become increasingly commonplace in the field of instructed SLA. As a result, there has been a call for a more systematically efficient way to design, use, and effectively implement them, both in the foreign language classroom and within SLA research. An inherent issue, however, is that with each appearance of the term *task*, there are often several definitions (or not) attached to it. Thus, the actual definition becomes somewhat arbitrary and difficult to interpret or generalize, especially across contexts (i.e., research or pedagogy in different situations). The variability arises depending on who (e.g., teacher, researcher or task designer) is using the tasks, for what purposes (e.g., to elicit language or to target/test a language structure), and in which situations (e.g., in a task-based classroom, in an experimental lab or to complete ‘real-life’ tasks). Due to this variable use of the term, as with other commonly used expressions in the field, it is pertinent to first clarify an operational definition in order to illustrate exactly what is being targeted. For this study, the focus will mainly be from the point of view of how

task is defined in the research context. However, it is through this type of research that pedagogical practice is best informed.

Many researchers rely largely on tasks as a unit of analysis within SLA research. Nevertheless, as illustrated earlier in the metaphor of the pendulum of language teaching options, their perspectives and ways of defining tasks may differ, which has often led to complicated findings. Let us consider how a leading researcher in this field has defined this term. Stemming from a theoretically psycholinguistic approach, as well as referencing ideas from several prominent SLA researchers (e.g., Prabhu, 1987; Nunan, 1989; Willis, 1996), Skehan emphasizes the components of a pedagogical task in which, “meaning is primary; there is some communication problem to solve; there is some sort of relationship comparable to real-world activities; task completion has some priority; the assessment of the task is in terms of outcome” (Skehan, 1998, p. 95).

At this point, Skehan’s definition does not reference linguistic structures as part of the focus of tasks, and furthermore, he warns that many types of common L2 classroom activities (i.e., gap-fills, drilling, etc.) should not be considered ‘pedagogic tasks’, although they often have been. He has also referred to these as ‘structure-trapping’ activities that may help with improving accuracy, but not necessarily to the natural acquisition of language learning. However, he later added that he is less concerned with the limitation of the idea that pedagogic tasks should replicate ‘real-world’ native language, and that it is sufficient for tasks to simply ‘resemble’ authentic samples (Skehan, 2003, p. 3). (For a review

of how tasks have been operationalized see e.g., Van den Branden, 2006; Samuda & Bygate, 2008).

Once the definition of task has been established by the researcher who is using it, within a specified context, then how it is applied or manipulated needs to also be clarified. Generally speaking, many SLA researchers use tasks to elicit some form of language outcome (e.g., telling a story about a past event to facilitate the use of past tense grammatical forms), regardless of the pedagogic approach that they follow. Thus, a two-way distinction has been created as a type of working definition of how tasks have been used for different purposes. However, the intention here is to focus on how tasks are applied, rather than defining the task itself; they have been distinguished through how the task has been defined within the research design structure (Brennan, 2009).

With the accumulation of more than thirty years of both empirical and classroom-based research using L2 tasks, researchers are finally able to gather some promising evidence to partially support or refute those projected theoretical models and hypotheses that have come to guide task-based decisions. In recent years, following the guidelines set forth by researchers such as Norris and Ortega (2003, 2006; Norris, 2010), SLA research meta-analyses and syntheses have increasingly strived to gain a better picture of the current situation in terms of systematically defining constructs in a manner that yields more generalizable findings. Thus, this type of research has helped demonstrate current practice about what has become accepted in the field, and moreover, which areas are still lacking

or not yet clearly understood. Findings from these types of studies will be included in the following sections.

Through one such synthesis, Brennan (2009) condensed the most common areas of SLA research that were utilizing L2 tasks, and furthermore, she identified which (or whose) theoretical constructs were guiding those studies. She first identified those studies in which researchers used L2 tasks for language elicitation purposes as *tasks as instruments*, where little or no attention was given to why the particular L2 tasks were being used. Conversely, those studies in which the use of L2 tasks (or features of the tasks) were specifically addressed, and in some aspect became one of the main variables, were termed *tasks as target* studies (from this point forward these will be referred to as TAT studies). The five main TAT research categories consisted of the following areas: *planning time, task-repetition, interaction, task complexity, and individual differences studies*.

At that time, the research focusing on task complexity and individual differences deemed to be the least systematic areas of TAT research in terms of methodological consistency and generalizable findings. Fortunately since then, however, there has been an influx of research examining task complexity and more focused attention has gradually been given to certain aspects of individual differences within task-based contexts (e.g., motivation, working memory, aptitude, anxiety, among others). See Tables 5 and 6 (pp. 45-46) for examples of complexity and individual differences TAT studies, respectively, and see Appendix 1 for examples of the remaining categories. Each study outlined, includes how the researcher has described the task-type they used, if there was a

theoretical or empirical reason for choosing or operationalizing the manipulation of the task, and if there were other variables involved in the research design.

As a result of the growing number of syntheses of task-based TAT research, a common concern found among them is that operationalizations of independent variables used (e.g., task complexity) have varied greatly, and sometimes researchers have not justified their methodological choices of task selection either theoretically or empirically. Consequently, the call has been made for a more systematic way of identifying constructs and subsequently obtaining and interpreting results more adequately from these studies (e.g., Long, 2005; Norris & Ortega, 2003; 2006; 2009; Norris 2010; Robinson 2001; Révész 2011; 2014; Skehan, 2009).

In addition, recent research has found the lack of consistent use of the various forms of ‘outcome measures’ of L2 performance, which is adverse to generalizing findings. As it has been acknowledged that a vital component of any task-based syllabus is being able to perform the target task, it is once again, pertinent that everyone involved is clear about what outcome measures are used. In turn, this has become problematic when comparing results that are not defined in the same way e.g., comparing end-of-term course or exam grades with the various ways of quantifying performance in terms of complexity, accuracy, and fluency (CAF) measures. However, this appears to be improving in recent years (e.g., Housen & Kuiken, 2009).

In sum, only when these variables both independent and dependent, are evidently and consistently operational across contexts, will we have a better understanding of how certain aspects of L2 tasks may affect performance and acquisition. Consequently, all those who use L2 tasks should be able to make more informed decisions about task selection and implementation processes. (e.g., Long, 2005, Robinson 2015, Skehan, 2009).

Therefore, one of the main intentions of the current study is to contribute to this line of TAT research, investigating aspects of both task complexity (i.e., task sequencing in terms of simple to complex) and individual differences (i.e., foreign language anxiety), as well as the interaction of the two. Furthermore, these variables will be explored, and therefore, supported by theoretical guidelines in order to justify the research design.

### **1.3. L2 Task Classification, Sequencing and Syllabus Design Models.**

Since the use of L2 tasks has become an increasingly popular choice of pedagogic unit in many student-centered or FonF language classrooms as well as in SLA research, there have been several proposals for how to create a more systematic protocol for selecting and sequencing tasks across contexts (See e.g., Baralt et al., 2014, p.10). However, upon investigating the various proposals, it appears that although some overlap can be found, there are still contradictory and ambiguous aspects, and therefore, there continues to be a call for a more rigorous exchange of ideas in order for L2 tasks to be used to promote SLA in a more reliable way.

In the mid-80s and 90s, researchers began to systematically classify and organize L2 tasks based on various frequently used ‘task-types’ (see Table 1 for some examples; Ellis 2003). Although many of them are still currently used in research and pedagogy, the way in which they are chosen or the order in which they are implemented, continues to vary tremendously. Additionally, the protocol for selecting and sequencing L2 tasks in order to drive the language learning process has been much less straightforward. Ellis (2003), for example, has offered a system that focuses on beginning with tasks that are ‘easy’ and then progressively adding more ‘difficulty’ to them in order to gradually ease the learner through the learning process. However, it often seems that those who apply this thought process, employ what Prabhu (1987) termed ‘commonsense judgment’ of sequencing tasks that facilitate a gradual and progressive ‘reasonable challenge’ for the learners (Skehan, 1996; Ellis 2003; Baralt, et al., 2014).

Table 1. Classifications of task-types

<i>A Pedagogic Classification (Willis, 1996)</i>	<i>A Discourse Domain/ Genre Classification (Swales, 1990)</i>	<i>A Cognitive Classification (Prabhu, 1987)</i>	<i>A Psycholinguistic Classification psycholinguistics (Pica, Kanagy &amp; Falodun, 1993)</i>
Listing	Narratives	Information-gap	Interactant relationship (One-way/two-way)
Comparing	Instructions	Reasoning-gap	Interaction requirement (required/optional)
Ordering and Sorting	Descriptions	Opinion-gap	Goal Orientation (convergent/divergent)
Problem-Solving	Reports		Outcome options (open/closed)
Sharing Personal Experiences Creative Tasks	Job Application Letters Recipes		

(From Ellis, 2003).



Long and Crookes (1992; Long, 2005), have claimed that the best starting point for any task-based syllabus is through first conducting a ‘needs analysis’ for task development. In other words, a task or syllabus designer should first identify which real-life tasks the learners would need to eventually carry out in the L2 and then base pedagogic target-tasks on those findings. Once those are formulated, the pedagogic tasks should then be designed and sequenced in a manner in which the tasks increase in complexity/difficulty; the outcome success is then measured by the learners’ ability to carry out, or approximate, real-life or target tasks (Long, 2005).

Whereas Skehan (2003) has agreed that this is ideal, he and Ellis (2003) have claimed that it is not always practically feasible to carry out such a needs analysis to create all L2 course syllabi. Thus, this is where we find the next challenge facing those who use L2 tasks as pedagogic units. First, there is the decision about which L2 tasks to begin with, and then, how to methodologically sequence them in an order which best promotes the L2 language learning process. Additionally, there is the ongoing question of how to systematically define what makes a task ‘simple’/ ‘easy’ or ‘complex’/ ‘difficult’ for the language learner.

Nowadays, the models and hypotheses initiated by Robinson and Skehan, along with their colleagues and predecessors, are among the most influential in guiding this line of research thus far. However, their methods and positions about how to approach the construct of task complexity and exactly how the factors included in each of their frameworks should lead to greater L2 performance and acquisition greatly differ. As a result, task complexity research has been applied

and interpreted in a number of ways. In addition, although both models initiate from a cognitive-interactive perspective, they diverge in their hypotheses on task outcomes as well as methods of how to define task features. Consequently, aspects of both of these frameworks have been challenged, and in some cases at least partially supported (or criticized), by a growing number of research studies. (See Table 5 and 6, pp. 45-46, for examples of these studies).

### **1.3.1. Peter Robinson's models for L2 task design and sequencing.**

#### *1.3.1.1. The Cognition Hypothesis for promoting L2 development.*

Upon review of the various seemingly insufficient L2 task sequencing proposals set forth in the 80s and 90s (see for example Candlin 1984, 1987), Robinson began to formulate his own theoretical constructs. To begin, Robinson stated that task-based classification and syllabus design systems should be: (a) “motivated by a theory”; (b) “empirically researchable”; and (c) “operationally feasible” (Robinson, 2007, p. 13). Thus, he began with his theoretical interpretation of the Cognition Hypothesis (CH), which has proposed a method of designing and sequencing tasks according to the gradual increase of *cognitive complexity*; starting from what are deemed to be the simplest versions of L2 pedagogic tasks and gradually increasing demands necessary to complete the target-tasks, such that they eventually emulate real-world tasks in the L2. Robinson has posited that this process leads to the restructuring of L2 linguistic forms and concepts, and therefore, promoting greater accuracy and lexical complexity, and eventually,

language development. He has based this claim on empirical-research findings of L2 performance measures, SLA theory, as well as research in the field of cognitive psychology (Robinson, 2001, 2003, 2005, 2007, 2011, 2015).

#### *1.3.1.2. The Triadic Componential Framework for L2 task selection.*

In keeping with the tenet of having an ‘operationally feasible’ method of identifying which factors should be considered when choosing and carrying out L2 tasks in a logical manner, Robinson has offered a classification system that comprises different aspects of task design, complexity sequencing, situational conditions, and learner factors. More specifically, the Triadic Componential Framework (TCF; see Table 2, p.36) focuses on L2 factors that involve task design dimensions and procedural processes, while being cognizant of learner differences, which he encompasses in the following three sections: ‘*task-complexity*’ (cognitive factors), ‘*task-condition*’ (interactive factors) and ‘*task-difficulty*’ (learner factors). The TCF offers a solution for the lack of a universal systematic taxonomy of tasks/task features, and therefore, to work as an initial guide for teachers, researchers and task/syllabus designers alike; in order to methodically test and investigate how certain task related factors contribute to L2 performance and acquisition (Robinson, 2001a; 2001b; 2003; 2005; 2007; Robinson & Gilabert, 2007; Robinson, 2010, 2011).

Robinson has operationalized *task complexity* in terms of the task features that could be manipulated in order to impose cognitive demands on the L2 learner. As can be seen in Table 2, this includes ‘resource-directing variables’

which are those task features that are claimed to place cognitive/conceptual demands on the learner e.g., aspects such as time reference (e.g., +/- here and now). These variables are intended to manipulate the learners' focus on the accurate use of linguistic or grammatical structures. Additionally, there are 'resource-dispersing variables' which are meant to induce performative/procedural demands on the learner. These variables are postulated to facilitate the learner to utilize whatever cognitive resources needed in order to execute the tasks e.g., attentional resources. Together, with careful application of both resource-directing and resource-dispersing task variables, Robinson claims that this will lead to greater accuracy and complexity of the L2 (Baralt, et al., 2014, p. 13).

Table 2. The Triadic Componential Framework

<b><i>Task Complexity</i></b> (Cognitive factors)	<b><i>Task Condition</i></b> (Interactive factors)	<b><i>Task Difficulty</i></b> (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)
<b><i>(a) Resource-directing variables</i></b> making cognitive/conceptual demands	<b><i>(a) Participation variables</i></b> making interactional demands	<b><i>(a) Ability variables</i></b> and task-relevant resource differentials
+/- here and now +/- few elements -/+ spatial reasoning -/+ causal reasoning -/+ intentional reasoning -/+ perspective-taking	+/- open solution +/- one-way flow +/- convergent solution +/- few participants +/- few contributions needed +/- negotiation not needed	h/l working memory h/l reasoning h/l task-switching h/l aptitude h/l field independence h/l mind/intention-reading

<b>(b) Resource-dispersing variables</b> making performative/procedural demands	<b>(b) Participant variables</b> making interactive demands	<b>(b) Affective variables</b> and task-relevant state-trait differentials
+/- planning time +/- single task +/- task structure +/- few steps +/- independency of steps +/- prior knowledge	+/- same proficiency +/- same gender +/- familiar +/- shared content knowledge +/- equal status and role +/- shared cultural knowledge	h/l openness to experience h/l control of emotion h/l task motivation h/l processing anxiety h/l willingness to communicate h/l self-efficacy

“The Triadic Componential Framework for task classification – categories, criteria, analytic procedures, and design characteristics” (From Robinson & Gilbert, 2007). \* h is ‘high’ and l is ‘low’

### 1.3.1.2. The SSARC model for task sequencing.

Using the theory base of the CH and the task features outlined in Table 2 (i.e., the TCF), Robinson advances his theoretical research agenda. In attempt to make his ideas more ‘empirically researchable’, he has designed the SSARC model (See Table 3, p. 38). This model aims to support task sequencing decisions based on increasing complexity, by adhering to the following two principles:

1. “Only the cognitive demands of tasks contributing to their intrinsic conceptual and cognitive processing complexity are sequenced”.
2. “Increase resource-dispersing dimensions of complexity first (e.g., from + to – planning time), and then increase resource-directing dimensions (e.g., from – to + intentional reasoning)”.

(See Baralt, *et al* 2014, pp. 16-18; Robinson 2010, 2015).

To sum up, Robinson has proposed a framework for using L2 tasks in a structured and universal manner that is intended to drive SLA. Based initially on the theoretical tenets of the CH, and guided by the task features outlined in the TCF, he then offers the systematic formula of task-based sequencing presented by the SSARC model. Therefore, he includes a recommendation for how to select and manipulate L2 tasks whilst considering different contextual factors, and at the same time minimizing obstacles caused by individual differences. As a result, these ideas are intended to promote L2 performance and acquisition.

Table 3. The SSARC model for task-based sequencing.

Step 1. SS (stabilize, simplify) = $i \times e [(s'rdisp) + (s'rdir)]^n$
Step 2. A (automatize) = $i \times e [(c'rdisp) + (s'rdir)]^n$
Step 3. RC (restructure, complexify) = $i \times e [(c'rdisp) + (c'rdir)]^n$

\*  $i$  = current interlanguage state;  $e$  = mental effort;  $s'$  = simple task demands;  $c'$  = complex task demands;  $rdisp$  = resource dispersing,  $rdir$  = resource directing;  $n$  is the potential number of practice opportunities. (From Baralt et al., 2014, p.17)

### **1.3.2. Peter Skehan's L2 task-based framework.**

#### *1.3.2.1. Skehan's task-selection framework.*

“The purpose of having a system such as this is that it allows tasks to be analysed, compared and, best of all, sequenced according to some principled basis” (Skehan 1996, p.53). Broadly speaking, Skehan set out to consolidate what had already been demonstrated in SLA research in the 1980s and 90s as well as to help contribute to resolving some of the shortcomings of foreign language teaching

methodologies at that time (e.g., too much ‘structure trapping’ in models such as the ‘PPP’ and not enough focus on structures in ‘TBLT’), while at the same time considering the effect of learner differences on L2 learning and acquisition.

Similarly to Robinson, he has supported a theoretically psycholinguistic FonF approach to language teaching and research. He has proposed a framework (see Table 4) which focuses on both language cognition and performance conditions surrounding L2 tasks, further broken down into: ‘*code complexity*’ (i.e., linguistic complexity), ‘*cognitive complexity*’ (i.e., processing and familiarity of content) and ‘*communicative stress*’ (e.g., time pressure). (Skehan 1996, p. 52; 1998; Robinson 2001, p. 194-195).

Table 4. Skehan’s Criteria for Task Difficulty.

Code Complexity	(syntactic and lexical difficulty)
Cognitive Complexity	(content of what is being said, i.e., Levelt’s conceptualization stage)
Cognitive processing	(amount of on-line computation required while doing task)
Cognitive familiarity	(does task rely on ready-made or pre-packaged solutions?)
Communicative Stress	(how quickly task has to be done)
Time pressure	(speaking vs. writing, listening vs. reading)
Modality	(# of participants, # of relationships involved)
Scale	(how important is it to do the task? Any real consequence?)
Stakes	(can participants exert an influence on how task is done?)
Control	(can they negotiate tasks, request a clarification?)

(From Baralt et al. 2014, p.9).

In contrast to Robinson's TCF for task selection, Skehan has focused more on the manipulation of commonly used L2 task-types in SLA research such as 'narrative tasks', 'personal tasks', and 'decision-making tasks'. Furthermore, those who use Skehan's model generally operationalize task features using the distinction of 'form-focused' and 'meaning-focused' tasks. Based on the work of Van Patten (1990) and Skehan (1992), he has stressed the importance of sequencing L2 tasks in a manner, "so that there is a balance between focus on form and focus on communication", (Skehan 1996, pp.47-50). Furthermore, he has proposed that a task-based syllabus, or sequence of tasks, should begin with the least demanding tasks and incrementally increase these demands as the learner becomes 'more comfortable' with the forms. However, it seems that there is no strict guideline in which to adhere when implementing or sequencing tasks within this framework.

#### *1.3.2.1. Performance outcomes and the Limited Attention Capacity Model/Approach*

Stemming from practical findings in this area, Skehan and Foster have postulated that L2 performance (i.e., CAF measures) may be affected by the combination of the task processing conditions based on the factors outlined in Table 4.

Furthermore, through the empirical work conducted by Skehan, Foster, and colleagues, they have contended that the more complex the task is, the less attention the learner gives to form, which has led to the theoretical proposal of what has become known as the Limited Attention Capacity Model (LCM; Skehan 1998; 2009; Skehan & Foster, 1999, 2001).



Skehan has further proposed that the performance on speech dimensions such as general CAF measures can be predicted through focusing on different aspects of L2 tasks as well as the level of ‘difficulty’ attributed to them, once task structures have been manipulated. One way this has been empirically analyzed is through studies which use, for example, independent variables such as different amounts and types of planning-time strategies. (See e.g., Foster & Skehan 1997; 1999; Tavakoli & Skehan, 2005; Tavakoli & Foster, 2008).

#### *1.3.2.2. The Trade-Off Hypothesis.*

Predominantly building upon the empirical findings from what have been termed here as task-based TAT ‘planning-studies’ (see Appendix 1), Skehan and Foster have formulated what is now known as The Trade-Off Hypothesis (TOH) for predicting outcome measures in L2 performance. Based on clear operationalizations of L2 performance outcome measures in terms of CAF, the TOH claims that as fluency measures increase, complexity and accuracy measures are in competition with one another. This is in opposition of the claims made through Robinson’s CH. Skehan has further emphasized the comparison between these two task-based research models (i.e., the CH and the TOH), by stating that his own hypothesis, or the TOH, is more of a “research-then- theory” way of looking at task performance and sequencing effects whereas Robinson’s CH is more of a “theory-then-research” account of the different relationships that occur between during task performance (Skehan, 2007).

Drawing from empirical studies Skehan and his colleagues have conducted, as well as findings from cognitive psychology (e.g., Levelt, 1989), he sets a research agenda to challenge the claims made through the TOH, through more systematic studies. As previously mentioned, these are predominantly what would be considered TAT planning and repetition studies (i.e., Ellis 2009; Skehan 1996, 2009; Skehan & Foster, 1999, 2001).

#### **1.4. Studies testing the theoretical task-based sequencing models.**

Fortunately, there have been a growing number of studies attempting to test various aspects of the aforementioned frameworks. Since the research agenda set forth by Robinson and Gilabert (2007), where they strongly recommended the need for more systematic research in this area, there has been an increase of studies focusing on the construct of task complexity. For example, Jackson and Suethanapornkul (2013) have recently conducted a research synthesis and meta-analysis based on an original pool of 170 studies that had focused either theoretically or empirically on the idea of task complexity. The research synthesis/meta-analysis concluded that although research in this area is fortunately taking shape, there continues to be a need for a more consistent system on how studies are designed and interpreted, in order to gain a true understanding of what has now been accepted in research and consequently what can be confidently applied to current instructed SLA practices.

Those studies which have been guided by the research agenda set forth by the CH, and the components of the TCF, have predominantly operationalized task

complexity through the use of manipulating *resource-directing* variables of +/- here-and-now (e.g., Gilabert 2005), +/- few elements (e.g., Gilabert, Barón, & Llanes, 2009), +/- intentional reasoning (e.g., Kuiken & Vedder, 2007); *resource-dispersing variables* of +/- planning time (e.g., Guará-Travares, 2008). However, there is an ongoing controversy related to the lack of clearly justifying how tasks have been defined in terms of ‘resource-dispersing’ or ‘resource-directing’ or, moreover, how tasks have been manipulated in terms of complexity. Often it has appeared that researchers have made tasks more complex through ‘instinct’. (See e.g., Jackson & Suethanapornkul, 2013, for the research synthesis and meta-analysis of the major studies examining the CH/TOH between 1996 and 2010; Baralt, et al. 2014; Sasayama, Malicka, & Norris, forthcoming).

In a recent review article about the two aforementioned frameworks, Andrea Révész (2014) has advocated for the importance of focusing on the independent variables, or in this case, the factors that may or may not have a contributing effect on specific learning outcomes. Some examples of studies working towards these goals are conducted by those researchers, that, rather than simply making intuitive decisions, they defend their strategies of task selection or task sequencing by citing the empirical support for grading tasks or by including external measurements that may empirically support such decisions such as: time-estimation scales, self-rating difficulty scales, as well as through eye-tracking data (e.g., Révész, Michel, & Gilabert, 2015; See Tables 5 & 6, pp. 45-46, for some examples).

Through this perspective of task complexity, studies using the TCF as a guide for task selection, design, and sequencing there has been promising research conducted. For example, there have been more studies with the progression of three levels of complexity rather than the most common dichotomy of simple and complex tasks (e.g., Kim, 2009; Malicka, 2014). There have also been studies examining the combination of variables (e.g., Kim & Tracy-Ventura, 2011). Additionally, there have been positive findings supporting more careful and transparent choices made about which CAF measures to use when analyzing task performance outcomes (e.g., Housen, 2009). Therefore, although it appears that research has begun to progress it seems researchers need to continue to include reasons why such decisions have been made supporting their research design and operationalization of variables, both dependent and independent, in order for studies to be replicated, and in turn, sufficiently contribute to the application of findings (Norris, 2010; Révész, 2011, 2014).

Table 5 and Table 6 (pp. 45- 46) summarize some examples of TAT studies that have either empirically tested or were theoretically guided by the aforementioned models. Included are those studies that focus on task complexity as well as task complexity in combination with individual differences, respectively. The way in which the tasks were operationalized in each study, what or who was the theoretical driving force motivating the research design or questions, and then, which aspects of the tasks were targeted is summarized. Clearly this is not an exhaustive list, as it is not within the scope of this paper to do a complete meta-analysis or synthesis (see e.g., Jackson & Suethanapornkul,

2013; Sasayama, Malicka, & Norris, forthcoming). The tables present an overview for the reader about how this type of research has taken shape over the past decade, and furthermore, which aspects are gaining momentum and which are still in need of further research.

Table 5. Task Complexity Tasks as Target (TAT) Studies.

<i>Researcher(s)</i>	<i>Description of task-type used</i>	<i>Theory/Research supporting task selection</i>	<i>Constructs targeted</i>
Gilbert (2005)	Narratives	Foster & Skehan (1996); Mehnert (1998); Ortega (1999); Skehan & Foster (1997), Robinson (2001)	+/- planning time, Complexity here-&-now vs. there-&-then
Robinson (2005)	Narratives	Schmidt (1990;2001); Long (1996)	Complexity of here-&-now vs. there-&-then
Nuevo (2006)	Narratives	Robinson (2001) and Skehan (1998)	High and Low complexity
Gilbert (2005)	Dialogic narratives, Instruction-giving, & Decision-making	Pica <i>et al.</i> (1993); Gilbert (2005; 2007); Robinson, (1995; & others)	Comparing task-type; interaction ; Increasing task complexity: +/- here-&-now; +/-elements; +/-reasoning demands;
Gilbert (2007)	Monologic narratives, Instruction-giving, & Decision-making	Levelt (1989, 1993); Kormos (2000, 2006) Skehan 1998; Robinson, (2002). Cornell et al. (1994) and Chown, Kaplan and Kortenkamp (1995).	Increasing task complexity; task-type; Increasing task complexity: +/-here-&-now; +/-elements; +/-reasoning demands; self-repairs
Robinson (2007)	Dialogic narratives: resource-directing, one-way; closed	Referenced many researchers & the CH	Complexity and “uptake”; anxiety, +intentional reasoning
Michel <i>et al.</i> (2007; Michel, 2013)	Leave a msg.; one-way, decision making phone conversation; two-way	CH (Robinson)	Complexity +/- one-way flow, +/- intentional reasoning
Kuiken & Vedder (2007)	Letter writing about reasons for choosing a holiday destination	Limited Attentional Capacity Model (Skehan & Foster), CH (Robinson)	Complexity+/- few elements, +/- intentional reasoning
Gilbert, Barón, & Llanes (2009)	Instruction-giving task same as in Gilbert 2007b.	Instruction-giving task same as in Gilbert 2007b.	Complexity +/- few elements ;+/- interaction; recasts, self-repairs
Kim (2009)	12 Decision-making tasks related to course themes: work, vacation, and university life.	CH (Robinson)	Complexity of ++/+/- few elements, +/- reasoning demands
Michel (2011)	2 Decision-making tasks; Leaving a phone message/phone call discussing and describing which electronic gadgets to buy	Limited Attentional Capacity Model (Skehan & Foster), CH (Robinson)	Complexity +/- few elements ;+/- interaction
Malicka (2014)	Dialogic problem solving scenarios in a hotel reception; 3 levels of complexity.	Tasks based on needs analysis of learners in a tourism school. -Limited Attentional Capacity Model (Skehan & Foster), CH (Robinson)	Complexity of ++/+/- few elements, +/- reasoning demands (Task Difficulty as supporting task complexity)
Levkina & Gilbert (2014)	Spatial reasoning, describing a flat; writing task; information-gap	CH (Robinson)	Complexity of ++/+/- few elements, +/- spatial reasoning

(Adapted from Brennan, 2009)

Table 6. Complexity and Individual differences TAT studies.

<i>Researcher(s)</i>	<i>Description of task-type used</i>	<i>Theory/Research supporting task selection</i>	<i>Constructs targeted</i>
Mehnert (1998)	Instruction task, Exposition task	Clark & Clark (1977); Faerch & Kasper (1983;1986); Crookes (1989)	Easy and difficult, different planning times
Ortega (1999)	Monologic Narrative re-tell	Ellis (1987); Crookes (1989); Foster & Skehan (1996); Doughty and Williams; (1998); Long (1991)	FonF, Learner strategies
Robinson (2001)	Dialogic map task; +/- familiar area	Sachs (1983); Meisel (1987)	+/- few elements ;Task difficulty
Cohen (2003)	Hypothetical examples	Doughty and Williams, (1998); Skehan, (1998).	Strategies & Learning Styles, Task-types
Ortega (2005)	Monologic Narrative	Ellis (1987); Crookes (1989); Foster & Skehan (1996); Ortega (1999)	Learner driven strategies, FoM vs. FonF
Kawauchi (2005)	Picture-based Narratives	Wigglesworth (1997); Foster & Skehan (several studies by these)	Different proficiency levels, type of learner strategies
Robinson (2007)	Picture sequencing narratives, + dual task	Referenced many researchers; the Cognition Hypothesis and Table 3	Anxiety affecting increasingly complex tasks
Trebits & Kormos (2008)	Narratives	Baddeley, (2003)	WM, Complexity
Kormos & Dörnyei (2007)	Oral Argumentative dyad	Kyösti Julkunen (1989); Skehan's (1996); Skehan & Foster's (1997)	Motivation; +planning;
Guará-Taveras (2009)	There -& -then, picture narratives	Mehnert (1998); Robinson (1995)	+/- planning time, WM and speech production
Révész (2009)	2 Interactive decision-making tasks;	CH; Robinson & Gilabert (2007)	+/- few elements, +/- intentional reasoning; self-confidence, anxiety and communicative confidence; recasts
Tavakoli (2009)	Dialogic Narratives (picture/video retell) & Reflective Interviews	Candlin (1987); Skehan (1998); Robinson (2001);Tavakoli & Skehan (2005)	Learner & Teacher perceived task difficulties
Kim & Ventura (2011)	Reasoning demands-oral, mono and dyads;	(Kuiken & Vedder, 2002, 2007)	+/- few elements; Language Anxiety
Révész (2011)	Argumentative task in a simple and a complex version	Task adapted from Watson, DeSanctis, and Poole (1988); CH (Robinson)	Causal reasoning WM; Dual Task; Developmental readiness, Attention control
Révész (2012)	Reenacting/ witnessing a crime.	CH (Robinson)	WM
Baralt (2013)	Reflecting & commenting on people's intentions in different situations	CH (Robinson)	+/- intentional reasoning ; Online vs. face-to-face interaction; Feedback; Learner anxiety and difficulty
Révész, Michel, & Gilabert (2015)	Same as previous studies by these three researchers	CH (Robinson)	Validating Task complexity with external measures

(Adapted from Brennan, 2009) \*WM is working memory and CH is the Cognition Hypothesis.

### **1.5. Individual differences and L2 task-based sequencing models.**

Long has confirmed that, “A one-size-fits-all approach has long been discredited by research findings...” (2005; p.1). This statement is in reference to the rationale made for conducting a needs analysis or a more tailor-made approach to task-based syllabus design in order to focus on the numerous learner factors that contribute to the language learning process. It is clear that both Robinson and Skehan are in agreement with the fact that learner factors should be addressed as they have incorporated this aspect into their projected task-based sequencing frameworks. Nevertheless, upon review of recent task-based TAT literature, those focusing on individual differences continue to be the least consistent, systematic empirical studies conducted in this area. Perhaps that is the point; there is no ‘one-size-fits all approach’. Thus, the question is whether we, applied linguists who use L2 tasks, can strive to be more empirically efficient and well-informed when making pedagogical or research-based choices in this area.

Broadly speaking, the various learner factors (i.e., individual differences) that may affect L2 learning have often been distinguished into two main categories: *cognitive factors and affective factors*. More specifically, these can be further divided into a variety of sub-categories which may affect the way in which L2 learners’ progress in the language learning process. Aspects of both cognitive and affective factors have been extensively researched within instructed SLA as well as in other areas of general education. However, the controversy remains of how to effectively manage or support these factors within the language learning environment. The two frameworks described in the previous sections, each have

offered ways in which to incorporate these variables in their projected task-based models for syllabus design. Nevertheless, once again, the testing of learner differences within a task-based environment has been the least systematic thus far.

As can be seen in Table 2 (p. 36), the TCF includes a list of various learner factors under the branch of *task difficulty* which distinguishes between learner 'abilities' (e.g., aptitude, intelligence, cognitive style, etc.) and 'affective variables' (e.g., confidence, anxiety, motivation, etc.). Robinson has claimed that the former group of variables is important for task or syllabus designers, whereas the latter group is deemed to be the methodological judgment of whoever is administering the tasks (e.g., teachers or researchers). This emphasizes the need for those involved in both the preliminary stages of task design as well as the implementation of L2 tasks, to be aware of how particular features of the tasks may affect L2 performance (and eventually acquisition). This also may vary depending on the context in which the tasks are carried out.

Skehan and Foster have also incorporated many of these aspects of learner factors into their work. In fact, it appears to be at the core of much of the work they have carried out. Theoretically, Skehan (1998) has incorporated both the affective (i.e., communicative stress) and cognitive (i.e., cognitive complexity) aspects into his framework as seen in Table 4 (p.39). Moreover, the effects of learner differences are further postulated through the LCM, and subsequently, the TOH. These are seen through the examination of both the processing abilities as well as affective factors, such as, motivation or learner strategies used when carrying out L2 tasks. To illustrate this, Skehan emphasizes that based on



learners' specific language learning goals in combination with other individual differences, the learners may consciously decide whether to focus on one of the components of language (i.e., CAF measures), or not.

According to the examples of the TAT studies presented in Table 6, p.46 (task complexity and individual differences), performance measures in conjunction with learner variables have begun to receive attention in research. Some examples include an examination of *learner strategies* that have been used to approach tasks in different contexts (e.g., Cohen, 2004; Ortega, 1999, 2005); *working memory* in combination with other factors such as complexity and planning (e.g., Guara-Tavares, 2008; Revesz, 2012); *motivation* or the effects of different interlocutors in interaction studies (e.g., Kormos & Dornyei, 2007); *learner perceptions* of task complexity or task difficulty (e.g., Robinson, 2007), as well as learners' *anxiety* levels and how they affect task performance (e.g., Robinson, 2001; Baralt, 2010; Kim & Tracy-Ventura, 2011).

Robinson has further stressed that as we gain a clearer understanding of how the first two aspects of the TCF (i.e., cognition and context) function, the third dimension of 'learner factors' also needs to be considered, as they can also play a major role in L2 performance and acquisition. He has noted that "research into interactions between L2 learners' cognitive processing abilities, their motivational and affective dispositions, and the demands of tasks has begun to be systematically approached in recent years [e.g., Baralt, 2010; Dornyei, 2002; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Robinson, 2002a, 2005b]" (Robinson, 2011, p.319). However, although progress has been made, it is still

unclear exactly how to sufficiently operationalize each of these ‘learner factor’ variables presented in these frameworks.

Robinson has suggested that researchers cannot always account for individual differences when designing and implementing L2 tasks. However, the researcher or teacher may be able to make *on-line* adjustments or adaptations based on the feedback they receive from the learners, that is, if learners are able to express how they are feeling at the moment of carrying out the tasks. Thus, as research in these areas develops, the empirical findings should work towards informing those *in situ* choices made by researchers or educators. Accordingly, more systematic research should be conducted in order to quantitatively and qualitatively understand what happens to learners during L2 task performance (Robinson, 2002, 2007, 2015).

In sum, the TAT studies as illustrated in Table 6 that have set out to pinpoint exactly how certain features of L2 task complexity (and how they are sequenced) in relation to various learner factors have affected L2 performance, has thus far been the most difficult to consolidate. Understandably, this line of research has the least cohesive and synthesized information available in the literature. Therefore, and similarly to the challenge of defining what is meant by constructs such as ‘task complexity’, there has been a number of ways in which ‘learner factors’ have been operationalized and measured in the TAT literature. Consequently, even though certain areas have increasingly sparked interest among researchers, conclusions and recommendations remain inconsistent.

Thus, as it is pertinent that constructs are well-defined as well as how they are viewed and assessed within SLA is understood, the variables included in the subset of 'learner factors' need closer attention. Within the individual differences TAT line of research, this is perhaps the most challenging feat, as it is often problematic to identify and then isolate certain variables due to the many subcategories found within this broad domain. While there have been a growing number of studies focusing on L2 task sequencing based on increasing complexity in relation to certain learner factors, to my knowledge, there has been little research investigating the effects of task sequencing and how it affects the learner factor of anxiety (Robinson 2001, Baralt, 2010; Baralt & Gurzynski-Weiss, 2011; Kim & Tracy-Ventura, 2011).

Therefore, Chapter II will address the operationalization of the construct of anxiety within SLA. Following the previously suggested guidelines for defining constructs (Norris and Ortega, 2003, 2006; Norris, 2010; Révész, 2014, among others), first, the concept of 'general anxiety' will be explored through a theoretical psychological and cognitive perspective, and then, how this factor came to further develop within the context of foreign language learning. Next, how this factor has been researched and assessed within the foreign language context will be addressed. Finally, the focus will be on the lack of consistent and validated methods for assessing the 'on-line' anxiety that may occur during L2 performance within the task-based context, and specifically within task complexity research.

Table 7. Summary of the main facets of Skehan's and Robinson's models

<b><i>Researcher:</i></b>	<b>P. Robinson</b>	<b>P. Skehan</b>
<b><i>Hypothesis:</i></b>	CH: As <i>accuracy</i> and <i>complexity</i> increase, <i>fluency</i> decreases.	TOH: As <i>fluency</i> increases, <i>accuracy</i> and <i>complexity</i> are in competition.
<b><i>Task properties or Cognitive demands:</i></b>	Task Complexity (TCF; See Table 2, pp. 36 & 37, column 1)	Task Demands; Cognitive Complexity (See Table 4, p.39)
<b><i>Sequencing Tasks:</i></b>	SSARC Model (See Table 3, p.38)	Balance of 'focus on form' and 'focus on meaning', progressively increasing demands of tasks.
<b><i>How they account for IDs:</i></b>	Task Difficulty Learner Factors; Affective and Cognitive (See Table 2, p.37, column 3)	Task Difficulty; Communicative Stress (See Table 4, p.39)

## **CHAPTER II: Individual Differences in Second Language Learning: How is Anxiety Identified?**

### **2.0. Introduction.**

This chapter sets out to give the reader an overview of how the unique cognitive and affective behaviors and characteristics of humans have been labeled, categorized, and assessed in the field of SLA. First, an overview is included outlining how some of the leading researchers have categorized these learner differences, which will from this point forward be referred to as individual differences (IDs). The main categories will be presented focusing on how anxiety has been perceived within the broad spectrum of IDs. Then, how anxiety has come to be defined and assessed specifically within the foreign language learning context will be addressed. Moreover, how research on this factor has been applied to the task-based context. Finally, there will be an outline of how some of the empirical studies mentioned in Chapter I have used assessment tools to identify IDs in the task-based setting.

The following main points will be covered in this chapter:

1. A broad overview of how IDs have been labeled and categorized.
2. How anxiety is recognized within these categories of IDs.
3. Anxiety through a psychological context; a working definition.
4. How anxiety has become a construct in foreign language learning.
5. A summary of how anxiety and other IDs have been assessed.

## **2.1. Individual Differences (IDs) in SLA: Where does anxiety fit in?**

Volumes of literature have been and continue to be written about how learner characteristics, or IDs, may affect performance and development within disciplines such as educational psychology and SLA, among others. Although it is not within the confines of this paper to review every aspect of this wide-ranging topic, it is important to first gain a brief insight into how these factors have been consolidated into what is now established within this line of SLA research, and furthermore, how the construct of anxiety has been considered within an overall umbrella of IDs. (For a more in depth review on this line of research see e.g., Dörnyei Z. & Ryan, 2005; 2015; Ellis 2008; Robinson, 2002; Skehan, 1989).

Gardner and MacIntyre (1992) have contemplated that, "...there are probably as many factors that might account for individuals in achievement in a second language as there are individuals." (p. 212). According to this view, it seems unimaginable to attempt to organize these innumerable differences into distinct, testable, and generalizable categories. Consequently, it is understandable why the empirical work conducted in these areas has included the most diverse consolidating findings, operationalizations of variables and methodological choices. Despite the fact that the topic of IDs is a vast one, researchers throughout history have attempted to identify some commonalities in order to gain a better understanding of what may affect the language learning process and L2 achievement as well as inform pedagogical practice.

In their book, *The Psychology of the Language Learner Revisited*, Zoltán Dörnyei and Stephen Ryan (2015) give a straightforward and critical overview of the current development of the prevalent IDs affecting SLA throughout the last decade. They begin by stating that they, "...find the concept of learner characteristics genuinely intriguing- and sometimes also genuinely annoying!" (p. xiv), emphasizing the fact that although there has been plentiful fascinating research conducted on the subject, there is still information yet to be discovered and/or resolved.

First, Dörnyei and Ryan challenge the rigorousness of the original paradigm of IDs (2005) in which it was declared that by identifying particular learner variables namely motivation, aptitude, and learning strategies, they would gain insight into the, "...*why, how long, how hard, how well, how proactively, and in what way* the learner engaged in the learning process". They then expanded upon this idea by proclaiming that:

Individual learner characteristics are not stable but show salient temporal and situational variation, and neither are they distinct and monolithic but involve, instead, complex constellations made up of different parts that interact with each other and the environment synchronically and diachronically (2015, p.6).

However, through the consolidated research of IDs conducted thus far, it is apparent that one of the ongoing goals has been to more precisely define the individual factors that these constellations encompass. For this reason, researchers and educators should be clear about how to best identify and interpret how and which IDs may be affecting the L2 processes (e.g., what exactly does a researcher mean by *anxiety* and what information is included in the studies to support such labels). Furthermore, as previously stated, it is through this knowledge that researchers and educators are able to make more informed decisions. However, as will be seen throughout the following sections, it is anything but clear-cut.

Despite the plethora of IDs in SLA, many have been labeled and conceptualized in various manners. Nonetheless, it continues to be problematic to efficiently identify, isolate, and then investigate all of the emotions and characteristics that make up individuals in order to determine what *exactly* may influence skill development, and furthermore, pinpointing at *exactly* the points along the learning process the effects of these factors occur. Although IDs are assumed to play a role in SLA, Dewaele (2009, 2012) referred to the attempt at identifying these learner differences as the ‘holy grail’ search for an underlying source that would be the key (or lack thereof) to L2 development (Dörnyei and Ryan, 2015, p.5). Nevertheless, we will now consider some of the ways in which researchers have attempted to define and categorize IDs, specifically focusing on how anxiety has been viewed. Then finally, concluding with the most influential theories and definitions that have steered the current study.



Skehan (1989) created a ‘canonical’ model of IDs which came to be an influential reference point in SLA. Through an extensive theoretical review of the research of what was happening at the time, he formulated an outline of those learner characteristics which seemed to most affect L2 attainment. Stemming from this, it was found that the learner variables including *age of onset*, *language aptitude*, and *motivation* had been the most consistently and widely researched IDs in SLA thus far and are still currently considered to be the best predictors of L2 attainment (Skehan & Dörnyei, 2003). Conversely, *anxiety*, although there has been a lot of research conducted in the area, has continued to be the most ‘curious’ one often ‘puzzling’ SLA researchers and educators to date (Dörnyei, 2009). Despite the fact that researchers recognize the predicament of defining and assessing the construct of anxiety, many caution against ignoring it completely. However, exactly how it should be assessed and handled in the language learning context has been less straightforward than some of the other IDs.

According to Ellis (2008), the main IDs being investigated in the field of SLA have been: *language aptitude*, *learning style*, *motivation*, *anxiety*, *personality*, *learner beliefs*, and *learner strategies*. He has further categorized the above IDs into four main areas (p.529):

1. “Abilities” (i.e., cognitive capabilities for language learning; e.g., *aptitude*).
2. “Propensities” (i.e., cognitive and affective qualities involving preparedness or orientation to language learning; e.g., *anxiety*).

3. “Learner cognitions about L2 learning” (i.e., conceptions and beliefs about L2 learning).
4. “Learner actions” (e.g., learning strategies).

As can be seen with the various ways in which IDs have been viewed in the field of SLA, it is not surprising that investigating them has become a central area of research throughout the last few decades. As a result, a variety of quantitative tools, usually in the form of Likert Scale questionnaires, have been developed and validated (For a list of the most commonly used scales of this type in current SLA research, see Table 21.1.; Ellis, 2008).

In the 70s, Scovel (1978) conducted a review of the IDs literature, which has often been regarded as a seminal work that initiated the SLA research agenda for investigating anxiety as it affects L2 achievement. However, more than two decades later, Scovel (2001) still has claimed that “anxiety is the most misunderstood affective variable of all” (Dörnyei, & Ryan, 2015, p.177). Additionally, MacIntyre (1999) and Skehan (1998) cautioned that although the most commonly referenced literature reviews at the time (e.g., Scovel 1978; Young, 1992) were important reference points to consider, they should also be interpreted with caution as they were not necessarily focusing on what was considered to be representative of the current SLA research.

Subsequently, various researchers have put forth theoretical frameworks in order to systematically guide research that addresses both cognitive and affective IDs that are thought to interact with SLA in specific ways. Stemming from these models, much empirical work began which has led researchers to conceptualize constructs even further. In the case of conceptualizing anxiety for example, this can be seen in the following five distinctions made: *Anxiety as a symptom of cognitive deficit; anxiety and multilingualism; anxiety in personality; anxiety and idiodynamic variations; positive aspects of anxiety* (Dörnyei, & Ryan, 2015, pp. 177-179).

Therefore, in the following sections we will take a closer look at how this complex variable has been further explored, measured, and treated through cognitive perspectives and models. First, the construct of anxiety will be reviewed in the context of cognitive psychology. Next, an overview of how research conducted in the area of IDs in SLA has led to the operationalization of this factor within the field of foreign language learning will be addressed. Then, the assessment of anxiety in the L2 context will be explored. Finally, we will look at how this factor could be analyzed in light of the cognitive-interactive task-based syllabus design models set forth by Robinson and Skehan.

## **2.2. Defining the construct of general anxiety: a psychological viewpoint.**

Before we explore how anxiety has been assessed in the foreign language setting, it is important to see how this construct has been defined in general, as this is where the definition that has been used in many cognitive-based SLA studies originates. As the task-based sequencing models of Skehan and Robinson that were described in Chapter I both stem from a cognitive approach to SLA, we will begin by examining how this construct has been defined from this perspective. One of the most commonly referenced studies since the 70s, and still active today in the field of cognitive psychology, is the work conducted by Eysenck on anxiety and how emotional factors affect task performance.

Initially based on research findings of that time (e.g., see Skovel, 1979), Eysenck followed a two-way distinction of the different types of anxiety that humans may experience. Firstly, Eysenck defined *trait anxiety* as, “a semi-permanent predisposition to experience anxiety having an important hereditary component” which he distinguished from *state anxiety* as, “a transient emotional mood or condition” which could manifest at any point in time and in various situations (1979, p. 363). Eysenck has subsequently proposed theoretical models, one of which has postulated what cognitively happens to learners who experience these anxious reactions during task performance. His perspectives have important implications for the current study, and therefore, will be explored further.

One example of Eysenck’s influence in the development of IDs was in the ‘The Big Five Model’. The model was built upon Eysenck’s (1984) ‘three-factor

model' which became well-known in the 80s as a paradigm of *personality trait* research which was composed of what were considered to be five 'broad' areas interestingly known as OCEAN: *Openness; Conscientiousness; Extraversion-Introversion; Agreeableness* and *Neuroticism-Emotional Stability*. These traits were all meant to affect SLA in different ways, depending on additional factors (e.g., aptitude). Anxiety was considered to fall under the latter category, and therefore, it has often been referred to in SLA research in this way; as a personality trait. This has had implications for how it has been evaluated in terms of quantitative scales as will be seen in Section 2.6 (Dörnyei & Ryan, 2015).

### **2.3. Defining the construct of anxiety in the L2 context.**

“Language Anxiety is a complex psychological construct requiring investigation from a variety of perspectives and approaches” (Young, 1992, p.157). This statement still rings true more than two decades later and, as a result, has undergone vast exploration since that time. Moreover, researchers within the context of SLA and applied linguistics have operationalized anxiety in a number of ways and along different stages of the language learning process. Although the proposed definitions have overlapped in some ways, they have often been interpreted and applied in different manners. Therefore, a review is included here of the most commonly referenced definitions used by some of the most prominent researchers in this area.

Derived predominantly from research conducted in the area of general anxiety and IDs (e.g., Eysenck, 1979; Scovel, 1979; Spielberger, 1983), Horwitz et al. (1986) and MacIntyre and Gardner (1991) initially made the distinction that anxiety can be broken down into three main areas: *trait*, *state* and *situational-specific*. First, following the definition given by Eysenck, they have conceded *trait anxiety* to be a generally stable emotion; it could be either genetic or acquired at a young age, which usually manifests in a variety of situations. Research in this area has long established that this type of anxiety can have a major influence on “cognitive, affective and behavioral functioning” in general (MacIntyre & Gardner, 1991, p. 88). Therefore, it could be posited that one who has this type of anxiety would most likely be more susceptible to feeling anxious in the language learning context. What the research is not yet clear about though is whether this leads to positive or negative results in terms of language learning success. Secondly, the aforementioned researchers also have agreed with Eysenck that *state anxiety* is the feeling of uneasiness that may be “experienced at a particular moment in time”. In contrast to those who have trait anxiety, the learners who display this emotion at some point in time are not deemed to be typically ‘anxious’ people *per se*, however, they may display momentary and/or unpredictable bursts or symptoms of anxiety at different points in time. Therefore, this type of anxiety could also potentially occur in the language learning setting; however, it less predictable.

Finally, Horwitz et al. (1986) and MacIntyre and Gardner (1991) have argued that *situation-specific anxiety*, is the anxiety that has developed from a

series of repeated events in a particular environment or situation, and therefore, is only exhibited in those contexts e.g., test taking (Spielberger, 1983). Accordingly, the anxious emotions which manifest in the L2 learning context are considered to be the situation-specific anxiety. Furthermore, it could be considered a ‘stable’ or a ‘trait-like’ disposition which is always associated with the foreign language context (Dewaele, Patrides, & Furnham, 2008, p. 914).

Nowadays, it has been generally accepted that this situation-specific anxiety is a prominent factor that affects L2 performance, acquisition, and achievement. It has been demonstrated predominantly through the analysis of anxiety during speaking and listening activities (for a comprehensive research review on the topic see e.g., Horwitz, 2001; 2010). Despite this knowledge, it is still unclear just how researchers and educators should effectively proceed. In the following sections we will examine what has been discovered in this area and how this construct has evolved in the field of SLA throughout the last several decades.

Throughout the 80s, the aforementioned researchers embarked upon a rigorous research agenda to more adequately and universally be able to identify the situation-specific anxiety that came to be known as Foreign Language Anxiety/Language Anxiety (FLA/LA), Communication Anxiety (CA), or the anxiety specific to the foreign language classroom or language learning setting. The work that was initially conducted by Horwitz et al., specifically defined FLA as the, “negative affective response to some experience in language class. With

repeated occurrences, anxiety becomes reliably associated with the language class and differentiated from other contexts” (Horwitz et al., 1986, p. 128).

Similarly, but with a slightly different angle, MacIntyre and Gardner (1989) defined CA as the “...anxiety in interpersonal settings, especially French speaking situations, and was significantly related to both the learning and recall of vocabulary items” (p.296). MacIntyre (1999) further distinguished CA into four specific areas within the language learning environment:

1. *Academic* - (e.g., displayed through any number of performance measures in the L2 setting; Final grades; CAF measures, etc.).
2. *Cognitive* – ‘disruption of thinking’, (e.g., TAT studies testing the CH or the TOH).
3. *Social Communicating*- using L2 in different social contexts with native speakers or peers and the emotion that it may bring to the language efficiency (e.g., Willingness to Communicate (WTC) studies).
4. *Personal*- self-image reports when speaking in the L2 or projected self-view. There have been a variety of qualitative interviews or journal analyses about this topic (e.g., Mackay, 2015 or other studies on motivation).

The pioneers of this line of research and the most frequently referenced surrounding FLA have done a systematic job of identifying this ever-present, albeit elusive, affective factor in the L2 classroom. However, questions still



remain about how the findings generated from these studies should apply to what currently happens in the field of L2 education and research.

As previously mentioned, there has been a plethora of studies throughout the past thirty years looking at this affective factor through all types of lenses including: psychology, psycholinguistics, testing, and education. However, the most influential studies surrounding the various aspects of FLA and its effect on SLA were initiated during the 80s and 90s. It is the findings generated from those studies that have driven the research conducted in this area, still today. However, it is evident that the theories, models, and constructs that were presented at that time are still being tested and challenged, and possibly with skepticism over the continued unresolved issues about how to appropriately analyze, and furthermore, cope with this factor.

#### **2.4. How has anxiety been measured in the foreign language context?**

As it has already been noted, there have been a myriad of perspectives and interpretations about how anxiety can be defined and assessed within the context of SLA. In addition, the point has been made about the importance of all those involved in the language learning environment to be well-informed about how to identify and appropriately manage and communicate across fields about this ‘curious’ yet common factor among language learners, which may manifest in different learning contexts and situations. Furthermore, many have warned about

using the various tools that assess FLA with caution (e.g., MacIntyre, 1999; MacIntyre & Gregerson 2015; Skehan 1998; Horwitz 2001; 2010).

Researchers and educators alike have often pondered some of the following frequently posed questions about FLA: whether anxiety could be a *trait* or a *state* emotion; whether it is fleeting or temporal; if it can be considered genetic or is it an emotional reaction that is somehow conditioned through environmental factors, and therefore, if there could be some cognitive influence. There has also been speculation about whether it is the cause or the effect of what manifests (e.g., poor performance) or whether it is a facilitating or debilitating emotion. Moreover, it remains controversial how exactly to measure when and why these anxious feelings occur. Finally, probably the most important question that has been raised by far is: whether educators can and should do anything to support those who experience anxiety within the classroom, or are there any preventative measures that could be applied to those who experience negative forms of anxiety? In sum, these are just a few of the numerous questions that have been posed throughout the years about this complex learner factor. Therefore, researchers have been making a concerted effort to find the answers to some of these questions (Horwitz, 2010).

Since anxiety has been considered problematic to assess there have been both quantitative and qualitative assessment tools used to investigate various aspects of this factor, and within various contexts. Some areas that have been investigated in relation to this factor within SLA are: L2 language skills (e.g., reading, writing, speaking, and listening); the effects of proficiency or outcome

measures (e.g., whether it is the cause or effect); assessing anxiety in the context of different L2s (e.g., Korean, Japanese, Spanish, etc); the anxiety of multilinguals, among others. However, despite over thirty years of research conducted on the topic of FLA, there continues to be a call for more ‘consistency and precision in the measurement of anxiety’. (See Horwitz, 2010, for a list of milestones in FLA research).

The following section is broken down into four parts. The first one describes the work carried out by Horwitz and her colleagues. The second section describes the scales and analyses that have been developed by MacIntyre and Gardner. The third section includes a brief summary of additional scales and methods that have been used in order to assess FLA in different contexts or in combination with different variables. Finally, there is a brief overview of the different types of tools specifically used in the task-based context, and then, a brief discussion of how the decision was made about which ones have been used for this dissertation.

#### **2.4.1. Foreign Language Classroom Anxiety Scale: (FLCAS; Horwitz et al. 1986).**

The work initiated by Elaine Horwitz and colleagues has become the most widely referenced in the area of FLA, specifically because of the tool that they developed: the Foreign Language Classroom Anxiety Scale (FLCAS; Horowitz, Horwitz & Cope, 1986). In fact, it has become the most standardized quantitative tool for measuring this factor. Moreover, the theory behind it has become a

framework and reference point for much research conducted in the area of FLA, still today. Their work has supported the idea that FLA should be considered a situation-specific anxiety that manifests within the foreign language learning environment (most specifically the L2 classroom).

The FLCAS originated from a qualitative approach, where the researchers set out to resolve what had apparently been troubling foreign language students at an American university. The items created for the scale were largely based on feedback from volunteers who had participated in a student discussion group which took place during what was called a “Support Group for Foreign Language Learning”. The group had been created as a pilot to see what type of feedback the researchers would find about how students felt about the language learning experience. The researchers noted that of a possible 225 total students, 75 volunteered to participate in the group. Interestingly, many had similar negative feelings in relation to the foreign language classroom environment (Horwitz et al., p.128). Therefore, with the feedback generated from these participants as well as through an analysis of other instruments that were being utilized at the time, they created a 33-item 5-point Likert rating scale (see Horwitz et al., 1986 for the full scale in English). A Spanish translated version is included in Appendix 2. Figure 2 presents two items from the FLCAS and examples of scores yielded from the original study.

**Figure 2. Examples of percentage scores from FLCAS.**

Items with Percentages of Students' Scores.

Strongly Agree	Agree	Neither agree or Disagree	Disagree	Strongly Disagree
1. I never feel quite sure of myself when I am speaking in my foreign language class.				
11	51	17	20	1
2. I <i>don't</i> worry about making mistakes in language class.				
11	23	1	53	12

*(Horwitz et al., 1986, p.29)*

As we can see in Figure 2, there was a high percentage (51%) of participants that agreed with the first question claiming that they did not feel sure about themselves in the language class. Additionally, more than half of the participants (65%) admittedly worried about making mistakes in the class. A further factorial analysis revealed that the constructs within this scale were deemed to be most closely related to the following: “1) communicative apprehension; 2) test anxiety; and 3) fear of negative evaluation”, within the context of the L2 classroom (p. 127). Horwitz et al. further claimed that “the results demonstrate that students with debilitating anxiety in the foreign language classroom setting can be identified and they share a number of characteristics.” (p. 129). As translated version of this scale was used in the current study (see Appendix 2) and a description is given in Chapter V, Section 5.3, where the methodology and design are described.

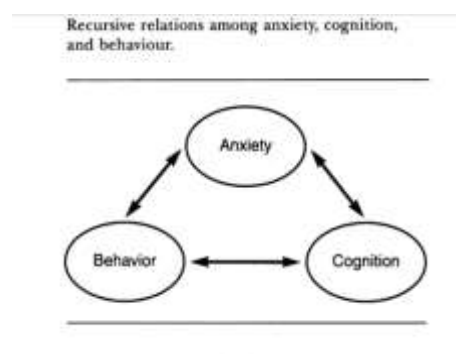
Since the introduction of the FLCAS rating scale thirty years ago, there have been a large number of studies that have utilized the FLCAS as a quantitative measurement to identify FLA. It has been used in a variety of settings as well as been translated into many languages. Thus, Horwitz has conducted two literature reviews during that time (2001, 2010). Firstly, she has claimed that the results in SLA research pertaining to this factor have been relatively 'uniform'. Secondly, she has found that much of the research conducted on this factor is consistent with their original findings; those deemed to be high-anxiety learners as demonstrated by quantitative measures of FLA (i.e., the FLCAS) tend to have low L2 achievement scores. However, she further stressed the importance of clearly defining how anxiety is being defined and in specifically what context(s) it is being investigated (Horwitz, 2010).

#### **2.4.2. Input Processing Output Anxiety Scale: (IPOAS; MacIntyre & Gardner, 1994).**

Concurrently to the work that was being conducted by Horwitz and colleagues, MacIntyre and Gardner were also rigorously experimenting with ways to assess anxiety in the L2 context. They took a slightly more psycholinguistic approach on how the effects of anxiety could hinder the language learning process. In Figure 3 (p.71) MacIntyre (1995) illustrates the concept that anxiety, cognition, and behavior are all interconnected but that it is difficult to ascertain which one is the cause and which one is the effect. Therefore, basing their work on an information-processing framework (e.g., Tobias, 1979, 1986), in conjunction with a thorough

analysis of both quantitative and qualitative instruments that measured different facets of anxiety, MacIntyre and Gardner proposed a new tool which was referred to as the Input Processing Output Anxiety Scale (IPOAS; 1994). As they were somewhat skeptical of strictly adhering to the available quantitative methods, they set out to investigate this tool as well as alternative modes for analyzing this factor.

**Figure 3. Relationship between anxiety, cognition, and behavior.**



*(From MacIntyre, 1995)*

To begin, MacIntyre and Gardner conducted a thorough review of a variety of tools which were being used at the time. They analyzed over 20 different types of affective scales that were designed for testing various aspects of anxiety including 4-point and 7-point Likert rating scales including ‘the Anxometer’, which assessed state anxiety, as well as several qualitative scales such as ‘can do’ and reflective essays. Upon conducting a series of factor analyses of the variety of anxiety scales that had been found, they yielded two orthogonal dimensions of anxiety: General Anxiety and Communicative Anxiety (MacIntyre, 1989).

Therefore, also basing their work on an information-processing framework (e.g., Tobias, 1979, 1986) in conjunction with a thorough analysis of both quantitative and qualitative instruments that measured different facets of anxiety, MacIntyre and Gardner offered the IPOAS. It could be argued that this tool is more holistic than the FLCAS, as it is designed to tap into the anxiety that may occur during the different stages of language learning. They proposed that there were three stages (i.e., input, processing, and output) during the language learning process in which the learner's anxiety could manifest, and consequently, could affect L2 performance and acquisition.

Ellis (2008) has mentioned that, “relating anxiety to a processing model, as proposed by MacIntyre and Gardner, may ultimately help teachers to fine-tune their interventions by focusing on specific stages of the learning processes”. Sparks and Ganschow (1995) however, argue that this ‘provided support for the idea that anxiety creates disruption in cognitive activity at each of the stages’ of SLA, maintaining the argument that perhaps the tools were not testing anxiety, but rather proficiency or linguistic abilities. Ganschow et al. (1994), which was later contested by MacIntyre (1995), challenged the work that had been conducted in this area and the attempt to label FLA, stating that this phenomenon actually stemmed from deficits in L1 rather than the foreign language learning context. They also advised caution about tools becoming the standard in assessing SLA (Ellis, 2008, p.541).



### **2.4.3. Anxiety in different contexts within SLA.**

Although research has cited speaking in the foreign language to be the most anxiety provoking activity (Horwitz et al., 1986; Horwitz, 2001, 2010; Young, 1990, 1992), we must take these findings with caution. Evidently there has been ample research since the mid-80s which has demonstrated this fact, and furthermore, those who have displayed high FLA/CA have often displayed poor foreign language performance or achievement, especially in the context of speaking in a foreign language. Nevertheless, we must recognize that this factor can manifest itself in various ways and contexts and that there are many factors that comprise an L2 course which may provoke anxiety (e.g., the method or approach used, how many students are in the class, among many others).

For example, consider the learner who has received a high score on an anxiety scale, e.g., the FLCAS, he/she may feel more anxious in a communicative group class rather than a traditional reading-writing approach context, or vice versa. Horwitz (2001) warns that, as a result, there may be confounds that exist in these types of studies. Therefore, it is important to view results from these studies with caution, and moreover, take a close look at how the study is designed. It is vital to understand how anxiety is being operationalized, in which context and finally, whether there are any other variables concurrently being examined.

Although L2 speaking has been the most commonly assessed skill in terms of FLA, there have also been a number of studies focusing on other language skills. A variety of tools have been generated to assess anxiety in these other skill

areas including: reading (FLRAS; Saito et al., 1999), listening (FLLAS; Elkhafaifi, 2005), and writing (SLWAT; Cheng, 2002; Cheng, Horwitz, & Schallert, 1999). Horwitz (2001), however, has further cautioned that although there are increasingly more studies using these tools to assess anxiety in a variety of ways, it is important not to confuse it with other variables. Thus, once again, it is important when using tools such as these that the construct is clear both in the idea of what type of anxiety is being addressed, the context where it is occurring, as well as the outcome measures that are being used (Horwitz, 2010; MacIntyre & Gregerson, 2012).

SLA researchers have provided a picture of this ‘constellation’ of IDs and have focused on how fleeting yet ‘curious’ the affective factor of anxiety can be within this plethora of learner factors. However, it continues to perplex all those involved. Therefore, researchers in this field urge educators to be aware that anxiety may manifest in the language learning environment, and moreover, where possible, adapt the teaching method in order to support learners who are experiencing this emotion or other similar affective emotions (Dewaele et al., 2008; Horwitz 2001, 2010; MacIntyre & Gregersen, 2012; Révész et al., 2015; Tavakoli, 2009; Young, 1992).

#### **2.4.4. Task-based sequencing models and anxiety.**

Revisiting Long’s (2005) claim that the “one-size-fits-all” approach has long been disproved, seems especially relevant considering the vast array of IDs that exist, especially in the case of anxiety. Therefore, the task-based sequencing models

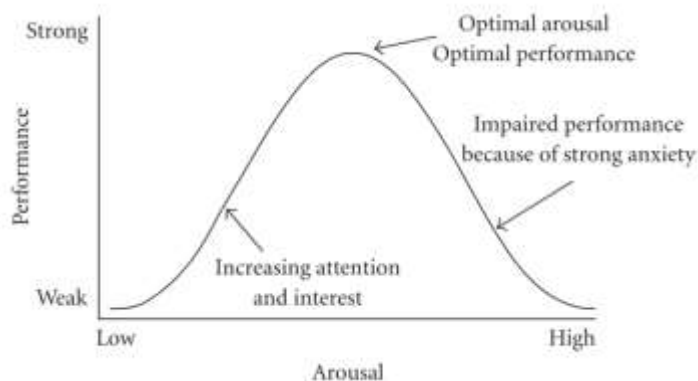
proposed by Robinson and Skehan in Chapter I, have offered a plausible match for these elusive learner factors. However, it has been argued that neither of the task-based models has overtly operationalized exactly how the factor of anxiety may be affected by the task sequencing processes they propose. Thus, there is evidently a gap in the research that needs to be filled in this area. If the definitions of anxiety that have been examined here in Chapter II are appropriately applied to the task-based sequencing models postulated in Chapter I, together researchers could gain a better understanding of how to best support the language learning process for these learners. Moreover, although anxiety has been deemed an affective factor in both task-based models it is pertinent that researchers in this area further explore the cognitive implications that may interact during task-based performance stages.

Derakshan and Eysenck, for example, have postulated that anxiety affects overall task performance, claiming the idea that the mental ‘effort’ taken to carry out tasks is affected by the cognitive effort taken up by the feeling of ‘worry’ or ‘distracting thoughts’ that generally occur when in this emotional state (2009). Although Derakshan and Eysenck’s work does not specifically focus on the context of language learning *per se*, it is relevant to the experimental study conducted here due to the importance given to the interaction between the emotional manifestations and cognitive processing which, consequently, may interfere with task-based performance. Furthermore, they have suggested that this process can lead to positive results by sequencing tasks in order from ‘easy’ to

‘difficult’. Hence, this line of work has been referenced in support of methodological choices in the area of sequencing L2 tasks (e.g., Eysenck, 1997).

Eysenck has made reference to the concept illustrated in Figure 4 (p.77) as an explanation of the interaction of anxiety with task performance (1979, 1997). In accordance with the research conducted using task-based sequencing models and those that are focusing on how to adapt to the learners’ anxiety that occurs in the language learning environment, the Yerkes-Dodson ‘law’ offers interesting insight into the effects of various levels of anxiety on task performance. Therefore, in adhering to this ‘law’ when following a task-based sequencing model, one must determine what level of anxiety the learner has while carrying out L2 tasks. In addition, and to gain further understanding of this interaction, careful attention should be given to how the level of anxiety corresponds with the features of the task (e.g., task type, complexity, and the sequence in which the tasks are carried out). Therefore, the L2 researcher or teacher can confidently decide which features of the tasks are manipulated in order to facilitate optimal performance. However, a system of accurately assessing the learners’ anxiety needs to occur in order to better understand what is happening during the process.

Figure 4. Yerkes and Dodson



(‘Yerkes and Dodson’, Hebbian - Diamond D.M. et al., 2007).

\*‘Arousal’ is considered here to be the emotion that leads to ‘anxiety’.

To conclude, Table 8 summarizes some of the task complexity studies (i.e., using a task-sequencing research design) that have included an instrument for assessing IDs in their research. Only the ones conducted by Kim and Tracy-Ventura (2011) and Baralt and Gurzynski-Weiss (2011) focused specifically on anxiety. Kim and Tracy-Ventura (2011) found that there were no significant effects of task complexity on anxiety; however, they also found that their operationalization of task complexity may have been weak. Baralt and Gurzynski-Weiss (2011) included a clear operationalization of both state anxiety and task complexity. Although they set out to compare state anxiety in computer-mediated interaction and face-to-face interaction and they did not find significant differences in the anxiety levels of the learners when comparing modalities. By using a mixed-method design they did, however, find that state anxiety was affected by task complexity. The findings from these two studies have both

contributed to the research on task complexity, but the latter study has given evidence that the computer-mediated L2 context did not pose more or less stress on the participants. These two studies have extended the research of the assessment of anxiety in light of task-based complexity and sequencing models and are, therefore, relevant to the study carried out for this dissertation.

Table 8. Examples of Task Complexity Studies using tools to assess IDs.

<i>Task Complexity Studies</i>	<i>Scales/Questionnaires</i>	<i>Construct</i>
(Gilabert, 2007; Michel, 2011)	5 questions; 5-point Likert Scale	Task difficulty, stress, confidence, motivation and frustration.
(Robinson, 2001, 2007)	1.5 questions; 9-point Likert Scale 2. 18 questions; IPOAS (MacIntyre & Gardner, 1994)	1. Task difficulty, stress, confidence, motivation and frustration. (post task) 2. Language Anxiety. (pre-test)
(Kim & Tracy-Ventura, 2011)	6 items; 6-point Likert scale; Adapted from Sheen (2008)	Language Anxiety.
(Révész, 2011)	Adapted from Kormos and Dornyei (2004), and McCroskey and McCroskey (1988)	Linguistic self-confidence, language anxiety, learners' self-perceived communicative competence; Perception of task difficulty by teacher and student.
(Baralt & Gurzynski-Weiss, 2011)	15 items; 5-point Likert scale; 'loosely based on the IPOAS MacIntyre & Gardner, (1994)'	State Anxiety.
(Malicka, 2014)	4 items; 9-point Likert scale	Difficult; Mental effort; Anxiety; Performance
(Révész, Michel, & Gilabert, 2015)	Subjective ratings and expert judgements.	Mental Effort; Difficulty

## **CHAPTER III: Purpose of Study; Research Questions and Hypotheses**

### **3.1. The main focus of the current study.**

Robinson has urged researchers to “thoroughly address, quantitatively, the twin issues for educational research raised, about how to sequence pedagogic task characteristics, and also how to map components of task complexity to individual differences between learners, so as to optimize learning and performance for groups and individuals” (Robinson, 2001, p. 255). Following this statement, the empirical study described here set out to contribute to the field of applied linguistics by contributing to the findings of these types of studies. First, the study examines the effects of task-based sequencing and complexity models postulated by Robinson and Skehan, and second, how these models interact with the learners’ anxiety during L2 task performance.

The current study, therefore, has taken a cognitive approach to the research design which focuses on whether task-type, task complexity, or task sequencing, based on varying complexity levels or task-types, has any effect on the learners’ state anxiety while carrying out L2 oral tasks. The goal was also to examine the relationship between those who are already deemed to have situation-specific anxiety, or in this case FLA, prior to participating in L2 tasks and how that compares to a more task-specific measurement of their state anxiety upon performing L2 oral tasks. The following section details how the frameworks that

have been described thus far have influenced the design, methodology, and operationalization of terms used in the current study.

As discussed in Chapter II, anxiety has proven to be one of the most prevalent affective learner factors influencing the L2 learning process, yet still continues to baffle many. This has been demonstrated through the extensive research and literature reviews conducted by a number of the most prominent researchers in this area (e.g., Dörnyei & Ryan, 2015; Horwitz 2001, 2010; MacIntyre 1999; MacIntyre and Gregerson, 2012).

Consequently, we have seen that a number of instruments have been designed as a method of identifying those L2 learners who are affected by this phenomenon (e.g., the FLCAS; IPOAS). However, as previously noted, those measurements have not always been conveyed as an adequate means for anticipating *exactly* when, if, or why learners will become anxious in particular L2 situations. Moreover, those tools have sometimes been criticized for the lack of clarity or validity of what construct is specifically being tested (see Ganschow et al., 1994 for more on this argument). Rather, it appears that they work more as a baseline to categorize those L2 learners that *may* have the tendency to exhibit some form of anxiety about aspects of the foreign language classroom setting (e.g., fear of making mistakes), although, it is not always clear as to what the source is or what should be (or is) done in these situations.

As a result, there have been numerous studies that have included either tailor-made scales or various types of available scales assessing the anxiety that could occur when engaging in particular L2 activities (i.e., listening, reading,



speaking, or writing) or within specific L2 classroom contexts. However, there is very little literature about vigorously investigating the construct of anxiety in a task-based context (see Chapter II). More specifically, there is a lack of research examining whether learners who may be deemed to have high FLA display any difference in their state anxiety levels, as influenced by specific features of the L2 tasks or the way in which they are sequenced. Therefore, one of the main objectives of the current study is the proposal of a tool which could be applied to situations that utilize L2 tasks either individually or within task-based sequencing models, for example, those aforementioned models set forth by Skehan or Robinson. The findings from this study set out to contribute to the research in both these areas, namely the research on FLA/LA in relation to SLA, and in combination with the research conducted on task-based sequencing models, where currently there is a gap in the literature.

Through an extensive piloting stage, which will be further discussed in the following chapter, an instrument was developed and then applied to the current study. The tools used were designed or adapted to examine whether the type of L2 task given, or the sequencing of the L2 tasks in terms of complexity levels, had any effect on the learners' state anxiety. Furthermore, a baseline was established of the learners' overall FLA levels, which was measured by the FLCAS scores, prior to task performance.

Using the results generated from this experimental study, an analysis was then conducted comparing the general FLA scores (FLCAS) to those responses generated from the more task-specific state anxiety scales (the Anxometer and the

Task Anxiety Reflection Scale, which will be further explained in the following chapter). As previously mentioned, the FLCAS has already been deemed a reliable method to evaluate whether students have FLA, whereas the latter two scales have only been used in the current study (the Anxometer has been used in previous studies, however, the one used here was an adapted version). In sum, the idea was to investigate whether these more timely tools could provide further insight into how L2 tasks features may interact with the learners' state anxiety. In which case, these scales could potentially be useful for foreign language teachers, researchers, and task designers alike, specifically when implementing and designing L2 speaking tasks.

Therefore, the aim of this study was twofold: first, to explore the effects of whether the sequence of L2 tasks in terms of task-type and task complexity had any effect on the participants' state anxiety levels. The state anxiety levels were rated through two Likert scales designed/adapted for this study which will be explained in the following section. Second, to analyze whether the participants' FLA levels, as indicated by the FLCAS scores, were a good predictor of the state anxiety levels that occurred during a sequence of four L2 oral tasks that had been manipulated in terms of task-type and task complexity.

### 3.2. Research questions and hypotheses.

The following research questions have been addressed in the experimental study:

*Research question 1: How do the participants' state anxiety scores generated by the Anxometer and the Task Anxiety Reflection Scale (TARS) compare, according to: (A) the complexity of the L2 tasks (i.e., simple vs. complex); (B) the sequence in which the L2 tasks were conducted (i.e., simple-complex or complex-simple); (C) the L2 task-type (i.e., the Map tasks or the Fire Chief tasks)?*

- *Hypotheses and predictions.*

The null hypothesis applies here. Although there are numerous studies that have proven FLA affects language performance, especially when carrying out speaking tasks (Daly, 1991; Horwitz, et al., 1986; MacIntyre & Gardner, 1991, 1994; Young, 1990, 1992), to my knowledge, there are few studies looking at exactly what specific aspects of the L2 tasks may influence state anxiety as affected by the type of task, the complexity of the task, and the sequence in which the tasks are carried out. Although there have been studies that have utilized affective ratings scales during task-based performance, they have usually been in combination with other affective factors such as motivation or task difficulty. Furthermore, often those scales have been used as a method of validating the construct of task complexity within a task sequencing design. In terms of the TOH and CH, it could be assumed that the participants would become anxious

when carrying out a more complex task. However, currently these models are intended to predict performance in terms of the CAF outcome measures and not affective measures.

*Research Question 2: To what extent do the situation-specific anxiety scores of the Foreign Language Classroom Anxiety Scale (FLCAS) predict state anxiety during oral task-based performance?*

- *Hypothesis and predictions.*

Again, the null hypothesis applies here. Although the FLCAS has been a widely accepted and utilized instrument to assess the construct of FLA in general, it is assumed that it would also be a good indicator of how certain learners would react during performance of L2 tasks. Therefore, it could be predicted that there will be a positive correlation between the scores on the FLCAS and the TARS and the Anxometer. However, it is argued here that although there may be a positive correlation of results at some point, the FLCAS does not provide *enough* information about the causes; that is, what specifically about the L2 tasks causes momentary or *state anxiety*. Therefore, it is predicted that the TARS will offer further information about the specific task factors that may contribute to the anxiety that manifests during L2 oral task performance.

## **PART II: THE EXPERIMENTAL STUDY**

## **CHAPTER IV: Methodology and Design**

### **4.1. Introduction.**

This chapter presents the methodological processes, the research design, and the procedures that took place during the piloting stages, and concludes with the details of the experimental study carried out for this dissertation. First, a description of the methodological decisions made about the research design, the participants, and the operationalization of constructs, as well as which L2 tasks and instruments were used for the final study is presented. To clarify why several of these decisions were made, a summary of the piloting stages is also provided. However, only those points that impacted the current study are included. Additional details from these stages can be found in the Appendices. To conclude, the methods used for coding and analyzing the generated data are presented.

### **4.2. Operational definitions of the variables.**

In keeping with the standards set forth by Norris and Ortega (2003), to clearly identify and theoretically and/or empirically justify variables and constructs being examined, the following section details those used in this study. The variables included here have been theoretically motivated by the proposed models described in Part I of this dissertation.

There were a number of considerations taken into account about how to define terminology and variables during the initial stages of this dissertation.

First, the operationalization of the constructs was determined, the L2 tasks

designed or selected, and the necessary instruments were adapted or created. In order to make informed choices about the definitions used here, a literature review was conducted and previous experimental studies related to these areas were explored (as has been discussed in Part I). Additionally, several pilot studies (which will be described in the following section), were conducted in order to create a theoretically and empirically supported research design.

#### *4.2.1. Operationalizing Task Complexity and Task-type.*

Task complexity has come under much scrutiny throughout the past decade, and as a result, researchers have begun to set guidelines for those investigating this construct to ensure their research is relevant to others in the field (e.g., Norris & Ortega 2009; Révész, 2011, 2014; Jackson & Suethanapornkul, 2013). Bearing in mind the theoretical frameworks set forth by Skehan and Robinson, task complexity has been defined through a cognitive approach to the demands placed on the learner while carrying out L2 tasks and how these can affect performance. More specifically, task complexity has been operationalized here in light of the task design features outlined in the TCF (Chapter I, Table 2, p.36). This study included two task-types (i.e., a map task and a problem-solving task) that were each manipulated to have two levels of task complexity. In other words, each task had a ‘simple’ and a ‘complex’ version that were manipulated through the resource-directing aspect of +/- elements (e.g., more choices to make or different people involved). Additionally, these tasks have been used in studies which

employed methods to empirically validate that the tasks were deemed to be more complex by participants as well as teachers (see e.g., Gilabert 2007; Gilabert et al., 2009; Révész et al., 2015).

Therefore, two different task-types in this study were cognitively differentiated on two levels. The first task was a map task, which has frequently been used in research due to the many linguistic and conceptual elements that can be manipulated (e.g., prepositions, landmarks). In this case, the participant was prompted to leave a voice message asking a friend to do some errands for him. The errands involved going to several locations, requiring the participant to follow the visuals on a map in order to give the precise directions. Here, as mentioned previously, the complexity was manipulated by the number of elements included on the map which guided the intended route. The simple version included a clear and visually supportive path, whereas the complex version was less straightforward, and included elements of ambiguity as well as detours to prevent the person from taking a simple or direct route (e.g., the elevator was broken so they needed to find the stairs or they had to choose from various symbols representing different choices).

The second task was a ‘problem-solving’ situation which has been referred to as the ‘Fire Chief’ task (e.g., Gilabert 2005, 2007; Gilabert et al., 2009; Révész et al., 2015). The participants were instructed to imagine that they were part of a volunteer rescue situation and had to indicate which course of action they would take to rescue the people from a burning building. The simple version had obvious choices offering clear exit routes and no clear obstacles. The complex version had



several elements that made the choice of action much more complicated to decide (e.g., someone trapped in the elevator, a pregnant woman and an old woman both needed some assistance; See p. 106 for a visual of the tasks).

#### *4.2.2. The operationalizations of anxiety for the current study.*

Two aspects of anxiety have been operationalized for this study. First, Horwitz et al.'s FLCAS was employed as a baseline measurement to assess if the L2 learners of English were considered to have FLA prior to carrying out the L2 tasks.

Therefore, this aspect of the construct was considered to be a 'stable' situation-specific anxiety, or FLA, which is deemed to occur in the context of learning or using a foreign language. Second, as the intention of this study has been to delve deeper into this factor by comparing the more stable FLA with the more transient emotion that could manifest during task-based performance, this has been considered to represent state anxiety. The state anxiety has been assessed through the administration of two Likert scales, which were designed and adapted for this study: the Anxometer and the Task Anxiety Reflection Scale (TARS: See Appendix 3 for the full TARS and Figure 7 for the Anxometer).

Furthermore, as different aspects of anxiety have been measured they have, therefore, represented different variables in the research design. The state anxiety scores generated from the two scales represent the dependent variable in this study. Conversely, the situation-specific anxiety ratings have represented the independent variable; however, only when addressing the second research question.

### **4.3. Creating the Task Anxiety Reflection Scale: TARS.**

At the start of the study, the intention had not necessarily been to create such a scale. However, it became inevitable that one was needed due to the fact that, to my knowledge, one such tool did not exist for the purpose of gaining insight into what *specifically* may have an effect on the state anxiety during L2 task-based performance, regardless of whether the learner had FLA. Consequently, this became one of the main objectives of the study. Thus, there were various stages that needed to take place in order to create the state anxiety scales used, prior to conducting the final study.

To begin, using Dörnyei's *Guidelines for Conducting Surveys and Questionnaires* (2003), the researcher set out to create a scale with the intention of capturing how certain features of L2 speaking tasks could potentially affect levels of state anxiety. The initial scale comprised five constructs which can be seen in Figure 5. The constructs were designed to tap into the different features of the learner's process, through a perspective similar to the IPOAS designed by MacIntyre & Gardner (1991). The anxiety that occurs during the input, output and processing stages of task L2 performance. (These were originally decided based upon email correspondence and discussions with Dr. Robinson and Dr. Gilabert). These constructs were originally represented by 20 items (i.e., ten positively and negatively worded statements); however, two items were initially removed as they were considered redundant and later two more were removed as they were related to the input which was not relevant to this study. The responses were made up of

a 5-point Likert scale ranging from 1 being, “I completely agree”, to 5 meaning, “I completely disagree with the statement”, similarly to the FLCAS scale. The scale was created in English for the first and second pilots, and then finally, translated into Spanish. (See Appendix 3 for the English version of the TARS used for the second pilot).

**Figure 5. Causal constructs of state anxiety included on TARS (Pilot).**

<b>Constructs</b>	<b>Examples</b>
<b>Anxiety due to Performance Demands (PD)</b>	Reaction to time pressure or being recorded.
<b>Anxiety due to Input Demands (ID)</b>	The emotions related to processing input necessary to carry out L2 tasks.
<b>Anxiety due to Legibility/Clarity of materials (MD)</b>	Tolerance to uncertainty/ambiguity in the materials (e.g., instructions, visuals).
<b>Anxiety due to Outcome Demands (OD)</b>	Feelings manifested with the realization of lack of knowledge/proficiency needed to carry out L2 tasks.
<b>Anxiety caused by Conceptual Demands (CD)</b>	Decision making ability needed to carry out tasks (regardless of the language).

#### **4.4. The Piloting Stage.**

The piloting stage had several goals. Principally, the objectives were to assess the tools and the format of the research design that would be used for the final study.

There were three pilot studies conducted and several smaller pre/post piloting

phases. The first pilot study will not be included here as it aimed at creating L2 speaking tasks which were not relevant instruments for the current study.

The second pilot occurred in the spring of 2012. The study aimed to test simple and complex versions of L2 speaking tasks as well as the newly designed version of the anxiety reflection scale (eventually referred to as TARS). The participants (N=46) were high school students in Philadelphia, U.S., who were taking Spanish as a foreign language. Parents had to sign a consent form to provide parental permission for those participants aged 16-18<sup>1</sup>. The researcher collaborated with the Spanish teacher to design L2 speaking tasks based on a needs-analysis of the topics taken from the class curriculum for their Spanish class. The materials were related to finding work in Spain. The L2 tasks were manipulated in terms of complexity according to +/- number elements in terms of Robinson's TCF (Chapter I, Table 2, p.36). The tasks were put on a PowerPoint format so the participants were able to view them and carry out the tasks individually in a computer lab. These tasks were not used in the final study due to lack of support for complexity levels as well as the participants were from a different demographic, and therefore, the tasks were no longer suitable.

The first version of the TARS was created in English as it was the mother tongue of the participants (see Appendix 3). Additionally, to establish an initial baseline of their FLA the participants completed an initial self-assessment to report whether they believed that they had 'high', 'medium' or 'low' foreign

---

language anxiety. Their Spanish teacher also labeled them according to how each of them behaved and reacted to various situations in the foreign language classroom.

Next, the participants were randomly assigned to one of two groups. The first group did the tasks in a sequence of simple to complex and the second group completed the same tasks but the sequence was counterbalanced for complexity. The participants recorded themselves doing the two tasks in a language lab,<sup>2</sup> and then, immediately completed the anxiety reflection scales (on paper) for each of the two tasks. The recordings were all uploaded to a sound file in Dropbox and the anxiety reflection scales were handed in to the teacher upon completion. All data were mailed or electronically sent to the researcher. An item analysis and an exploratory factor analysis were conducted to see if any of the items on the scale should be eliminated and furthermore, if they loaded onto the intended constructs. The results from the item analysis can be found in Appendix 4.

#### *4.4.1. Results, limitations, and what changed for the subsequent pilot.*

After analyzing and reflecting on the data and initial research design, as well as feedback received from experts about the original scale, some changes needed to be made to the TARS. Subsequently, the scale was amended based on the feedback as well as results drawn from the exploratory factor analysis conducted

---

<sup>1</sup> The researcher was not present during the recordings but the participants often recorded themselves as a requirement for their Spanish course. Additionally, a consent form of confidentiality of personal information was signed by all of the participants' parents, which was kept on file at the high school.

at that time. Additionally, as the decision was made to use only English as-a-foreign language (EFL) students who resided in Barcelona, the scale was translated into Spanish. Therefore, in keeping with the recommendations made by Dörnyei Z. (2007), the scale was then translated into Spanish and then back-translated into English by a bilingual translator in order to validate the wording. Additionally, prior to the following pilot, there were several pre-pilot discussions and evaluations about the exact wording of the items by several bilingual English-Spanish speakers; two of these individuals worked in the area of translation or bilingual education and one had no background in translation or foreign language education but was able to provide critical feedback about some of the wording. All of this feedback was taken into consideration when creating the final version.

The final version included 16 items that comprised four constructs (Figure 5). The construct of anxiety in terms of the input or receptive understanding was removed. This was due to the fact that the study was aimed to focus on oral skills and the concept of understanding the materials given was incorporated in another construct. The following is a summary of the main changes that occurred due to findings from this pilot:

- The FLCAS was used as a baseline for assessing the participants' general FLA instead of the self-rating/teacher-ratings.
- The study location would take place only in Barcelona for two reasons; first, the high school where the study had taken place no longer allowed for this type of research to be conducted there, and

second, as the researcher was located in Barcelona it was more convenient to obtain a larger sample size there.

- The L2 tasks used would be based on previously conducted task complexity studies that were already proven reliable in terms of complexity levels.

#### *4.4.2. Results and limitations or changes made from the third pilot.*

The final pilot was conducted in the autumn of 2014 with 60 participants at the University of Barcelona, who were deemed to have a similar make up to the type of subjects that were to be included in the study presented here. The results from the final pilot were analyzed using a mixed-method approach in order to triangulate the results generated. For the quantitative analysis, an exploratory factor analysis was once again conducted to determine whether the items all loaded onto the projected constructs. With the results generated, minor adjustments were made and the scale was then re-piloted with ten additional participants who completed a ‘stimulated recall’ and post-study interview in order to gain a more qualitative and descriptive insight into what was exactly happening while carrying out the L2 speaking tasks.

Additionally, all the participants were given the option to comment or ask questions at any point of the study; either by raising their hands or by writing comments after any of the items. This was added to triangulate the quantitative data collected, which has been suggested when obtaining construct reliability.

This contributed to the modifications made including the addition of the open-ended question following each item, where participants were invited to make further comments about the statements. An example of one comment made after completing the Simple Map task was:

*“I do not feel comfortable to give directions in my first language and much less in English”.*

This participant also displayed high anxiety ratings on both Map tasks during the performance stage of the corresponding task, especially on the more complex one.

The main adjustments made from this pilot were the following:

- The biggest limitations from this pilot were in terms of logistics i.e., students did not complete all parts. More specifically; the Anxometer was skipped by many participants. Thus, the decision was made to change the Anxometer, to make it more attractive and simplistic (as recommended by Dörnyei, 2003), and to transfer all paper materials to an electronic format on SurveyMonkey, except for the FLCAS.
- Due to the responses from the stimulated recall, the decision was made to include open-ended comments question as a way in which the students could provide further information about their feelings.



## **CHAPTER V: The Experimental Study.**

### **5.1. Study Design.**

In sum, a repeated-measures mixed experimental design was conducted in which participants performed two oral tasks with two levels of complexity followed by the completion of the final versions of state anxiety scales: the Anxometer and the TARS. All data were collected digitally, which was then followed by quantitative and qualitative analyses.

### **5.2. Participants.**

In total, there were 300 volunteers who participated in some part of this dissertation. However, those who participated in the piloting stages have already been described. The total number of volunteers who took part in the final study was initially 160 (i.e., they completed the initial language background survey, signed the consent form, and/or did the FLCAS or the main study). The goal was to have a sample size of at least 150 in order to have the results from the questionnaire be statistically reliable (Dörnyei, 2003). However, not all of the participants' results were included in the final analyses. Several of the participants were excluded due to the fact that English was considered to be one of their first languages. Furthermore, those who did not complete all parts of the study or could not be identified (i.e., he or she did not label one part of the project), were not included. Therefore, there were a total of 108 participants in the final study, a further breakdown can be found in Table 9 (p.100).

There was a similar demographic to those participants from the third pilot study; English as-a-foreign-language (EFL) university students who were based in Barcelona. They comprised University of Barcelona students, the majority of which were studying English Pedagogy at the degree level, and the remainder of them were studying various other subjects and were taking English as an elective course. The University students consisted of five different intact classes that were either in their first, second, or third year of their undergraduate studies. The majority of the participants claimed to have learned English through a traditional approach to language learning (i.e., Figure 1, Option 1, p.22).

In order to gain a more global understanding of the participants, a language background questionnaire was collected either by paper or online, at the beginning of their course or at some point throughout the study. The questionnaire is included in the external CD-ROM Appendix. However, the following are some examples of the topics covered:

- First language and other known foreign languages.
- Age they began learning English.
- How often they use English and their other language(s) on a daily basis.
- Reasons for currently learning English.

Those who did not have either Spanish or Catalan as their primary language were also not included in the final data analysis due to the fact that there was not a translated version of the TARS readily available, for all of the first

languages. However, those participants were able to complete the FLCAS in English and a few opted to complete the study in Spanish as a foreign language. Once again however, their results were not taken into account at this point.

Although a proficiency test was not given at the time of the study, all of the courses in which the students were enrolled required proof of an entry level of at least B2 (High Intermediate level), according to the Common European Framework (CEFR). Furthermore, as part of the Language Background questionnaire the participants were asked to estimate their levels, such as, whether they had taken an official exam (e.g., IELTS), and to include a description of the types of schooling they have had in English (however, this was not done in great detail which will be further addressed in the Discussion section). According to the responses, all of the participants claimed to have had at least twelve years of formal instruction in English and to have at least a B2 (High Intermediate) level. 70% of the participants from all four groups claimed to have at least a C1 (Advanced) level of English. It is possible that this may have had some effect both on the performance and/or the anxiety they exhibited during the study and will be further addressed in the Discussion chapter.

The participants included in this study were selected as they were deemed to have a high proficiency level even though a level test was not conducted at the time the study was carried out. Additionally, since many of the participants were studying either English Pedagogy, it would be assumed that they were motivated or they were good language learners. However, this was just an assumption, and therefore, in the future it is recommended that along with the FLCAS and the

language background questionnaire there should be other learner factor scales as well (i.e., language aptitude, working memory, motivation, etc.).

**Table 9. Descriptive statistics of the participants.**

Sequence Group	Sex	Age <i>M (SD)</i>
Sequence A <i>n=24</i>	4 Males; 20 Females	21.17 (1.43)
Sequence B <i>n=27</i>	4 Males; 22 Females	21.07 (3.04)
Sequence C <i>n=29</i>	6 Males; 22 Females	20.96 (1.69)
Sequence D <i>n=24</i>	5 Males; 19 Females	21.17 (2.57)
N=108	19 Males; Females 83	21.09 (2.25)

Table 9 presents the number of participants which were included in the final study, in accordance with the sequence in which they participated.

### **5.3. The Instruments Used for the Current Study.**

The data collected were predominantly quantitative in nature. But several participants responded to more open-ended questions or a ‘stimulated-recall’ interview (during the third pilot/pre-pilot study), in efforts to triangulate the findings. The following questionnaires/scales were used in the current study. Steps 1, 3, and 4 were all formatted and entered into the survey and questionnaire software program (SurveyMonkey Inc., 1999-2016):

1. The Language Background Questionnaire (see CD-ROM Appendix)
2. The Foreign Language Classroom Anxiety Scale; FLCAS (Appendix 2, p.170)
3. The Anxometer (Figure 7, p.103)
4. The Task-based Anxiety Reflection Scale; TARS (Appendix 5, p.175 & 176)

### *5.3.1. Foreign Language Classroom Anxiety Scale; FLCAS*

Following the second pilot study, it was decided to change from the self/teacher ratings of FLA to a more comprehensive and reliable baseline of the participants' situation-specific anxiety. As the FLCAS has been one of the most widely used scales for assessing whether a foreign language student is considered to have general FLA, it was therefore chosen here. Moreover, as most of the participants had either Spanish or Spanish/Catalan as their first languages, a Spanish translated version of the FLCAS was administered (taken from: Pérez-Paredes & Martínez, 2000-01). This was conducted according to the recommendations that the administration of a scale of this nature (emotional) should be translated into the participants' first language (Dörnyei, 2003).

The FLCAS is a 33-item scale which uses a 5-point Likert Scale rating system. For example, a rating of 1 is, "I totally agree" ("Estoy totalmente de acuerdo) with the given statement (or item) where a score of 5 means, "I totally

disagree” (Estoy totalmente en desacuerdo) with the given statement. Figure 6 includes examples of items from the FLCAS.

**Figure 6. Examples of items from the FLCAS.**

Item in English	Spanish translation
<i>2. I don't worry about making mistakes in language class.</i>	<i>2. No me preocupa cometer errores en clase.</i>
<i>4. It frightens me when I don't understand what the teacher is saying in the foreign language.</i>	<i>4. Me asusta no entender lo que el profesor está diciendo en idioma extranjero.</i>

Scores generated were calculated by coding the items, where all negative statements were reversed and then recoded, following the methods that were suggested in previous studies that used this scale (Aida, 1994; Horwitz, 1986). Once the final score was tallied from a possible score of 33 (low anxiety) to 165 (high anxiety), the subjects were labeled with either: high, medium, or low anxiety levels. A Cronbach's alpha of .93 (N=108) was obtained (Horwitz, 1986).

### 5.3.2. The Anxometer

An adapted version of the Anxometer was used, which allowed the subjects to give an overall initial state anxiety rating of how they felt, or global gut reaction, upon completion of each of the four oral tasks (MacIntyre & Gardner, 1991; Robinson 2003, 2007, among others). As previously mentioned, this scale was

originally designed to assess the construct of state anxiety of the subject either before, during or after completing an activity in a foreign language. This was used to complement the more detailed reflection scale, the TARS, which will be discussed in further detail in the following section. During the piloting stages of the study, the Anxometer was administered, on paper, which was placed above the TARS scale, based on a 0-100% rating system (MacIntyre & Gardner, 1991). However, this became problematic due to the fact that either the participants skipped it, or misinterpreted what they had to do, which greatly affected the analysis of the results. Therefore, the decision was made for the final study to make this scale more attractive and user-friendly (Dörnyei, 2003). Additionally, it was digitalized and programmed so that the participants were unable to move on to the next stage/question until they had selected an anxiety rating. The scale was included within the overall computer-based format and was based on a 1-10 rating as well as emoticons representing each level of anxiety. (See Appendix 3 for the original version of the Anxometer).

**Figure 7. The Anxometer, adapted for this study.**



### 5.3.3. The Task-based Anxiety Reflection Scale (TARS; Spanish Version):

A Cronbach's alpha for the 16 item scale was found to be .85 which was, therefore, found to be within normal limits of internal construct reliability.

Furthermore, Figure 8 provides an example of several items of the 5-point Likert-Scale. In addition, Figure 9 includes the 4 constructs targeted in the final version of the reflection scale. (See Appendix 5 for full TARS).

**Figure 8. Example items from the TARS.**

2. El hecho de que me estuvieran grabando me ha puesto nervioso/a.

1 Muy de acuerdo   
 2 De acuerdo   
 3 Indiferente   
 4 En desacuerdo   
 5 Totalmente en desacuerdo

Comentarios (opcional):

3. Me he sentido bajo presión para completar la tarea a tiempo.

1 Muy de acuerdo   
 2 De acuerdo   
 3 Indiferente   
 4 En desacuerdo   
 5 Totalmente en desacuerdo

Comentarios (opcional):

4. No estaba seguro/a de lo que tenía que hacer.

1 Muy de acuerdo   
 2 De acuerdo   
 3 Indiferente   
 4 En desacuerdo   
 5 Totalmente en desacuerdo

Comentarios (opcional):



**Figure 9. Causal constructs of state anxiety included in TARS scale.**

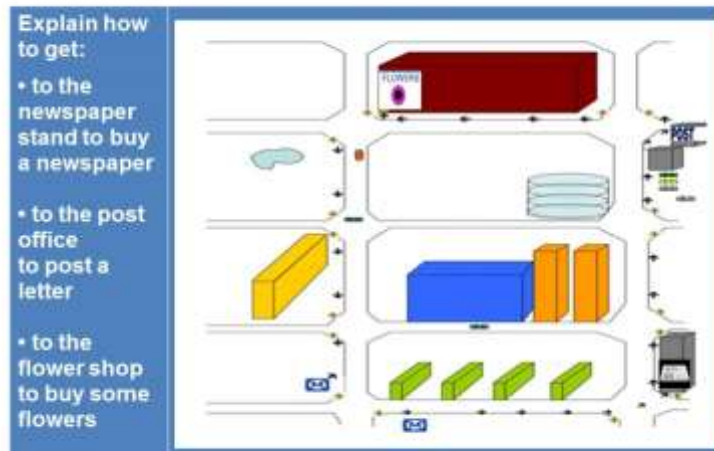
Constructs	Examples
<b>Anxiety due to Performance Demands (PD)</b>	Reaction to time pressure or being recorded.
<b>Anxiety due to Legibility/Clarity of materials (MD)</b>	Tolerance to uncertainty/ambiguity in the materials (e.g., instructions, visuals).
<b>Anxiety due to Outcome Demands (OD)</b>	Feelings manifested with the realization of lack of knowledge/proficiency needed to carry out L2 tasks.
<b>Anxiety caused by Conceptual Demands (CD)</b>	Decision making ability needed to carry out tasks (regardless of the language).

#### 5.3.4. *Tasks Used for Current Study*

For ease of administration, the tasks used to facilitate the language samples were also all digitally formatted and incorporated directly into the questionnaire software program (SurveyMonkey Inc., 1999-2016). The subjects carried out two different oral tasks each consisting of simple and complex versions. As previously mentioned, these specific tasks were chosen as they have been previously validated as tools that are considered to be simple and complex, defined by features of the TCF (Michel, Révész & Gilabert, 2012, 2013; Révész et al. 2015). All tasks are included in Figures 10 and 11, however, see the CD-ROM Appendix for the full materials used for Sequence A.

Figure 10. Map tasks

## SIMPLE MAP (SM) TASK



## COMPLEX MAP (CM) TASK

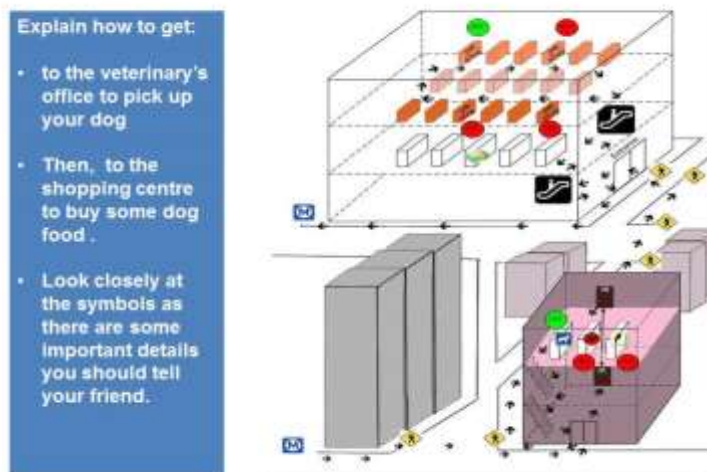
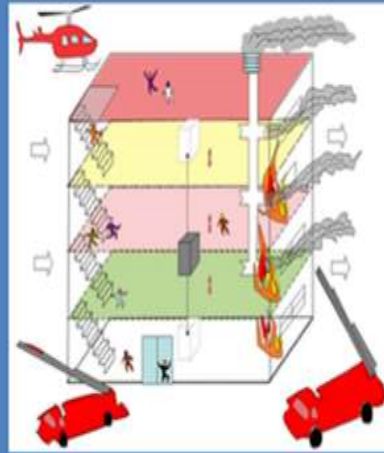


Figure 11. Fire Chief tasks

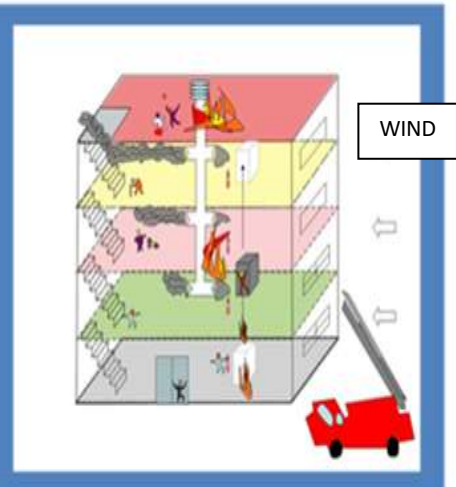
## SIMPLE FIRE CHIEF (SFC) TASK

- Explain what actions you would take to save as many people as possible.
- Explain the sequence in which you would take those actions.
- Justify your choice of actions and their sequence.



## COMPLEX FIRE CHIEF (CFC) TASK

- Explain what actions you would take to save as many people as possible.
- Explain the sequence in which you would take those actions.
- Justify your choice of actions and their sequence.



*Additional instruments*

- Praat recording software was used to collect all of the language samples.
- Headphones with attached microphones were used to minimize the sounds of the voices of the other participants who were simultaneously recording their language samples.

**5.4. Procedure.**

In accordance with the ethical standards of the University of Barcelona, as well as the *APA Manual sixth edition (2010)*, all participants signed a consent form stating that they were voluntarily taking part in the research study and that they were able to ask questions or could withdraw from the project at any point. As there were three parts to complete throughout the duration of one term, some subjects did not complete all of the components. Therefore, those subjects were only included in the analysis according to that in which they participated. The exact number of participants for each section will be stated in the analysis of the Results section.

As previously stated, the participants all completed a language background questionnaire either at the onset of the course or at some point throughout the course term. It was collected either on paper or digitally through *SurveyMonkey*, (the digital language background was deemed much more effective). In addition, the Spanish version of the FLCAS was administered on paper in order to establish the general FLA level of each of the participants. This also acted as a way to

allow for the participants to decide if they would be willing to continue to participate in the research project (i.e., they needed to complete a consent form at the same time as well as began to understand the nature of the project).

The main part of the study took place in a computer lab at the University of Barcelona. The L2 task materials/prompts and anxiety scales were prepared by the researcher prior to the participants arriving. Additionally, the Praat software program to record their language samples was open and ready for them to use. The participants were randomly assigned to one of the four sequence groups (A, B, C, & D; See Table 10); each participant completed the same four oral tasks, although again, they were counterbalanced for sequence of task-type and complexity level. A brief introduction and general verbal instructions were given to the entire group by the researcher (in English), and additionally, the instructions were visually provided on the screens of each computer. (Again, see the CD-ROM for the complete Sequence A).

Table 10. Task Sequence format used in study.

<b>SEQUENCE</b>	<b>TASK 1</b>	<b>TASK 2</b>	<b>TASK 3</b>	<b>TASK 4</b>
<b>A</b>	SM	CM	SFC	CFC
<b>B</b>	CM	SM	CFC	SFC
<b>C</b>	SFC	CFC	SM	CM
<b>D</b>	CFC	SFC	CM	SM

To begin, the participants were instructed to record themselves completing one practice speaking task, in order to assure that everything was working properly and that they understood how to proceed with the materials. At that point, they were prompted to ask the researcher if they had any questions or problems with any of the equipment or if they were not clear about something, but that they should try their best to complete the tasks alone. By the end of the session (which lasted between 20 and 45 minutes, depending on the participant), they each had five language samples recorded, including the practice task. In addition to the language samples, there were a total of four completed Anxometers and TARS scales where the participants responded to how they felt about each individual task.

Table 11. Summary of Study design.

<i>Instruments used</i>	<i>Brief description of procedure</i>
1. Language Background Questionnaire	Conducted either during class or for homework on paper or through SurveyMonkey.
2. FLCAS	Conducted during class time, on paper. The professor explained a general idea about the questionnaire and students needed to sign for consent to participate in the study.
3. 4 L2 speaking tasks and the Anxometer and the TARS (including open-ended questions).	Conducted during class time, in a computer lab. Researcher and researcher's assistant were present to assist participants with any technological problems.

### **5.5. Scoring Procedures for Anxiety Scales.**

First, the scores for each of the anxiety rating scales (i.e., the FLCAS, the Anxometer, and the TARS) were tallied. All responses were transferred to *excel* 2010 and the sums were calculated. Upon calculating the scores of the FLCAS, the participants were then grouped into: High, Medium and Low foreign language anxiety levels. The total possible score for the FLCAS was 33 to 165. The TARS scores were calculated for each individual task as well for the total of all four tasks. The possible totals for the TARS scores ranged from 16 to 80 for an individual task, and 64 to 320 for all four tasks. The total Anxometer scores could range from 1 to 10 for an individual task, and 4 to 40 for all four tasks. Once all of the TARS and Anxometer scores were coded and calculated for each participant, the results were further analyzed.

### **5.6. Statistical Analyses.**

In sum, a repeated-measures mixed design was used to analyze the data generated from this study. For the analyses discussed in the Results chapter, the data were grouped according to the sequence in which they were completed, and then adapted in accordance with each of the research questions. The software programs used were IBM SPSS version 19/23, excel 2010, and a calculator. The scores displayed are in adherence with the guidelines from the *APA manual* (2010). Additionally, the researcher followed suggestions made by Larsen-Hall, (2012) and Plonksy (2015) about reporting results.

First, the data were all checked for normality through descriptive analyses.

Second, inferential statistics were carried out in order to address the two research questions. Additionally, examples of the qualitative responses were included to triangulate the findings. Finally, additional quantitative analyses were completed in order to support the reliability of the instruments that were used in the study.

### *5.6.1. Quantitative Analysis*

- *Within-subject analysis*

A within-subject analysis was conducted using an ANOVA comparison of the FLCAS scores (the general FLA) with the results from the Anxometer and then the TARS (the state anxiety scores), corresponding to each of the four tasks. The results were used to answer research question number 2.

- *Between-subject analysis*

Next, the scores generated from the three anxiety scales were compared using a one-way between-groups ANOVA design. The groups were organized by sequence in order to assess if there were any differences in levels of anxiety based on task-type, complexity level, and/or sequence. The analysis was intended to support or refute the hypothesis posed in research question 1.

Additionally, a correlation analysis was conducted comparing the results from: a) the FLCAS b) the Anxometer, and then c) the TARS. This was conducted to compare the situation-specific anxiety scores with the state anxiety



scores of the Anxometer and the TARS, upon completion of each of the four tasks.

#### *5.6.2. Additional Analyses of the instruments used*

A factor analysis of the TARS was conducted including all the participants that completed this portion of the study for each of the four tasks. This was intended to support construct validity of the scale designed for this study.

#### *5.6.3. Qualitative analysis*

An analysis was conducted of the responses to the open-ended questions from the TARS, as well as the answers given during a stimulated recall interview, which was conducted with ten participants during the piloting stages. This information aimed to lend further support when responding to the research questions as well as to the reliability of using a tool such as the TARS for this study.

## **PART III: THE RESULTS**

## **CHAPTER VI: The Analyses**

### **6.1. Introduction.**

This chapter is divided into three parts. First, the results generated from the quantitative analyses regarding the two posed research questions and hypotheses are presented. This includes the results of the descriptive analyses followed by the inferential statistics, according to each independent variable (i.e., task complexity, sequencing, and task-type). Next, an analysis of the FLCAS scores is presented; first through a descriptive exploration, and then by conducting a correlation analysis of the FLCAS, TARS, and Anxometer ratings. The second section includes examples of the responses drawn from the qualitative analysis used as a qualitative resource to further support the quantitative findings. Finally, there is a supplementary section which includes results of the additional analysis conducted.

### **6.2. Descriptive exploratory analysis.**

The descriptive analysis was conducted in order to establish a normal distribution of data as well as to test for outliers. There were a few minor outliers found, however, as they did not reach beyond 3 standard deviations above or below the mean, it was decided to maintain all cases for the calculation. Therefore, the scores were adjusted by calculating  $M-2.5(SD)$ , and the new scores were added to the data. The descriptive analysis was conducted once again with the new scores generated, and no outliers were found.

As described in the Methodology section, the participants were randomly assigned to one of four different sequence groups based on a Latin square design. The groups were labeled by sequence (i.e., A, B, C, or D) as can be seen in Table 10, which will be referenced in all tables that refer to the task sequence. All participants completed the same four L2 oral tasks, but the tasks were counterbalanced for complexity, sequence, and task-type. Additionally, the tasks were given the following descriptive titles that will be used in most tables in this chapter unless otherwise indicated: Simple Map (SM), Complex Map (CM), Simple Fire Chief (SFC), and Complex Fire Chief (CFC).

By eyeballing both the tables and figures (histograms and QQ plots) the data appeared to be relatively normally distributed. Table 12 and Table 13 present the descriptive analysis of the TARS and Anxometer ratings for all of the tasks in each sequence. The mean scores found already appear to capture the fact that there is a difference when comparing the anxiety ratings displayed during simple and complex versions of each task. Additionally, the Skewness and Kurtosis values all fell within the acceptable range of -1.96 to +1.96.

Table 12. Descriptive statistics for all TARS.

Task (N=108)	<i>M</i>	<i>SD</i>	<i>S</i>	<i>K</i>
Simple Map	42.24	8.19	.122	-.260
Complex Map	47.71	8.63	-.078	.134
Simple Fire Chief	43.78	8.40	-.190	-.135
Complex Fire Chief	47.60	7.49	-.129	.345

Table 13. Descriptive Statistics of all Anxometers.

Task (N= 108)	<i>M</i>	<i>SD</i>	<i>S</i>	<i>K</i>
Simple Map	5.24	1.71	.091	-.147
Complex Map	6.00	1.84	.351	-.467
Simple Fire Chief	5.60	1.83	.182	-.204
Complex Fire Chief	6.02	1.76	.109	-.166

### 6.3. Results for research questions 1.

*How do the state anxiety scores generated by the Anxometer and the Task Anxiety Reflection Scale (TARS) compare, according to: (A) the complexity of the tasks (i.e., simple vs. complex); (B) the sequence in which the tasks were conducted (i.e., simple-complex or complex-simple); or (C) the task-type (i.e., the Map tasks or the Fire Chief tasks)?*

#### 6.3.1. Results for inferential analyses of part (A): Task complexity.

A one-way between subjects ANOVA was conducted to compare the effects of complexity, sequencing, and task-type on the participants' state anxiety ratings upon completion of the four oral L2 tasks. First, the results of the effects of task complexity are presented. Box plots have been included here as this has been strongly recommended as a way to clearly visualize the range of variance in the mean scores (Larson-Hall, 2012). As can be seen in Figures 12 and 13, there were only one or two low-level anxiety outliers found for all tasks except for the SM task. However, once again, the decision was made to include these participants in the data, therefore, the following calculations were done for the TARS outliers:

CM:  $47.7 - (2.5(8.8) = 22) = 25.7$ ; SFC:  $43.7 - (2.5(8.6) = 21.5) = 22.2$ ; CFC:  $47.5 - (2.5(7.9) = 19.75) = 27.75$ . Then, as well for the Anxometer outliers: CM:  $5.97 - (2.5(1.86) = 4.65) = 1.32$ ; SFC:  $5.59 - (2.5(1.83) = 4.57) = 1.02$ ; CFC:  $6.02 - (2.5(1.76) = 4.4) = 1.62$ .

The results are presented in Table 14 and 15 as well as the boxplots in Figure 12 for the TARS. The TARS means yielded an  $F(3, 107)$ ,  $p < .001$ . A Bonferroni post hoc pairwise comparison revealed a significant difference seen between the Simple and Complex tasks, only a slight difference between the two Simple tasks, and almost no difference was found between the two Complex tasks. The results are displayed in more detail, of both the TARS and the Anxometer scores on Tables 16 and 17, p. 121. Confidence Intervals were also included, as recommended by Larsen-Hall and Plonksy (2015), to establish an accurate range of scores.

Table 14. One-way ANOVA for TARS Scores.

TASK	<i>M</i>	<i>SD</i>	<i>95% CI</i>	
			<i>LB</i>	<i>UB</i>
SM	42.24	8.19	40.67	43.80
CM	47.65	8.79	45.98	49.33
SFC	43.72	8.57	42.08	45.35
CFC	47.48	7.87	45.97	48.98

N= 108

\*Reminder that SM is Simple Map is SM, Complex Map is CM, Simple Fire Chief is SFC, and Complex Map is CM.

Figure 12. Boxplots of the TARS scores for all tasks.

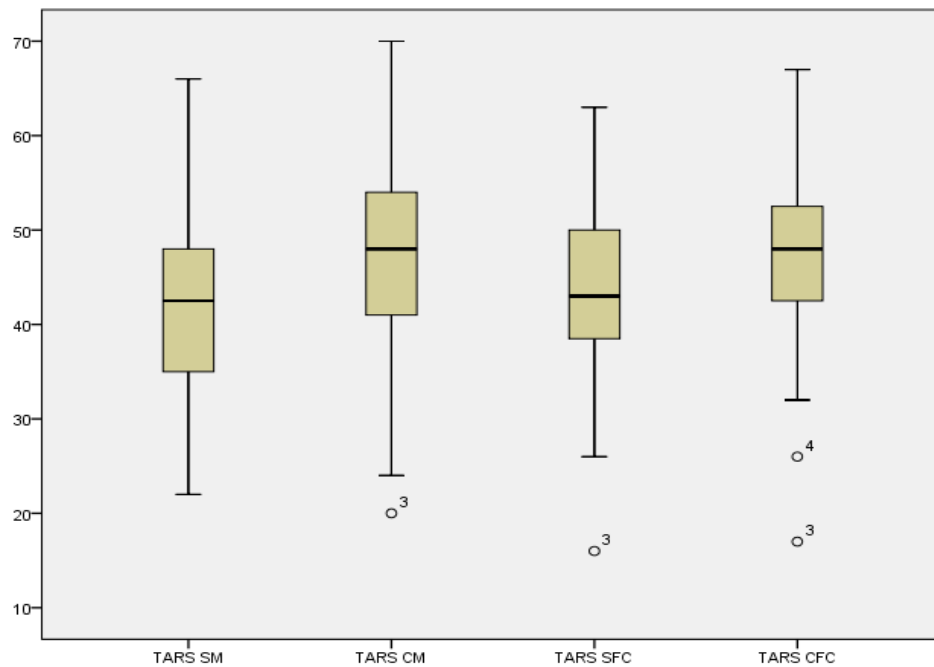


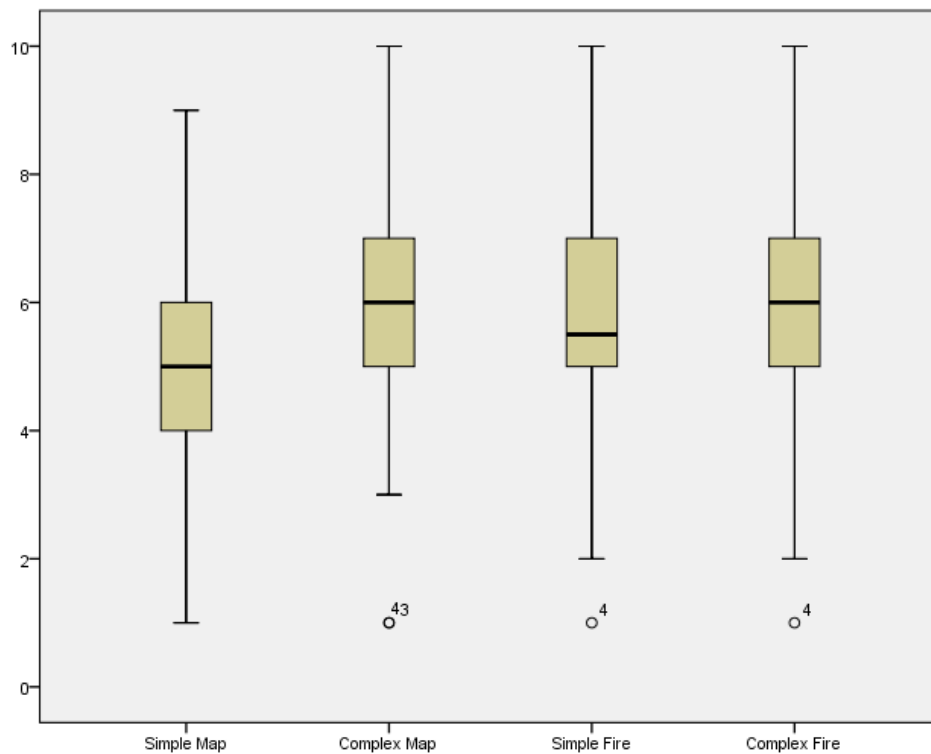
Table 15. One-way ANOVA of the Anxometer Scores.

Tasks	<i>M</i>	<i>SD</i>	<i>95% CI</i>	
			<i>LB</i>	<i>UB</i>
SM	5.241	1.71	4.913	5.568
CM	5.972	1.84	5.617	6.327
SFC	5.593	1.83	5.243	5.943
CFC	6.019	1.76	5.683	6.355

N= 108

\*Reminder that SM is Simple Map is SM, Complex Map is CM, Simple Fire Chief is SFC, and Complex Map is CM.

Figure 13. Boxplots of the Anxometer scores.





## Second Language Anxiety and Task Complexity

Table 16. Pairwise Comparison of Complexity: TARS Scores

(I) Task	(J) Task	M Diff (I-J)	Std. Error	Sig.b	95% CI
Simple Map	CM	<b>-5.41*</b>	<b>0.84</b>	<b>0</b>	<b>[-7.67, -3.15]</b>
	SFC	-1.48	0.71	0.25	[-3.41, 0.45]
	CFC	<b>-5.24*</b>	<b>0.81</b>	<b>0</b>	<b>[-7.41, -3.06]</b>
Complex Map	SM	<b>5.41*</b>	<b>0.84</b>	<b>0</b>	<b>[3.15, 7.67]</b>
	SFC	<b>3.93*</b>	<b>0.77</b>	<b>0</b>	<b>[1.88, 5.99]</b>
	CFC	0.17	0.77	1	[-1.89, 2.24]
Simple Fire Chief	SM	1.48	0.72	0.25	[-0.45, 3.41]
	CM	<b>-3.93*</b>	<b>0.77</b>	<b>0</b>	<b>[-5.99, -1.88]</b>
	CFC	<b>-3.75*</b>	<b>0.62</b>	<b>0</b>	<b>[-5.41, -2.10]</b>
Complex Fire Chief	SM	<b>5.24*</b>	<b>0.81</b>	<b>0</b>	<b>[3.06, 7.41]</b>
	CM	-0.17	0.77	1	[-2.24, 1.89]
	SFC	<b>3.75*</b>	<b>0.62</b>	<b>0</b>	<b>[2.11, 5.41]</b>

\*.The mean is significant at the .05 level.

b. Adjustments for multiple comparisons: Bonferroni.

Table 17. Pairwise Comparison of Complexity: Anxometer Scores

TASK (I)	TASK (J)	M-D (I-J)	Std. Error	Sig.b	95% CI
Simple Map	CM	<b>-.73*</b>	<b>0.17</b>	<b>0</b>	<b>[-1.19, -0.28]</b>
	SFC	-0.35	0.15	0.12	[-0.75, 0.05]
	CFC	<b>-.78*</b>	<b>0.17</b>	<b>0</b>	<b>[-1.22, -0.33]</b>
ComplexMap	SM	<b>.73*</b>	<b>0.17</b>	<b>0</b>	<b>[0.28, 1.19]</b>
	SFC	0.38	0.18	0.22	[-0.11, 0.86]
	CFC	-0.05	0.18	1	[-0.54, 0.45]
Simple Fire Chief	SM	0.35	0.15	0.12	[-0.05, 0.75]
	CM	-0.38	0.18	0.22	[-0.86, 0.11]
	CFC	<b>-.43*</b>	<b>0.14</b>	<b>0.02</b>	<b>[-0.81, -0.04]</b>
Complex Fire Chief	SM	<b>.78*</b>	<b>0.17</b>	<b>0</b>	<b>[0.33, 1.22]</b>
	CM	0.05	0.18	1	[-0.45, 0.54]
	SFC	<b>.42*</b>	<b>0.14</b>	<b>0.02</b>	<b>[0.04, 0.81]</b>

\*.The mean is significant at the .05 level.

b. Adjustments for multiple comparisons: Bonferroni.

*6.3.2. Results for analyses of part (B): Task sequencing.*

A repeated-measure ANOVA was conducted to evaluate if there were any effects of sequencing found through the participants' state anxiety scores, with regards to the sequence in which the tasks were performed. The table below is once again presented in order to remind the reader of the order in which the tasks were sequenced.

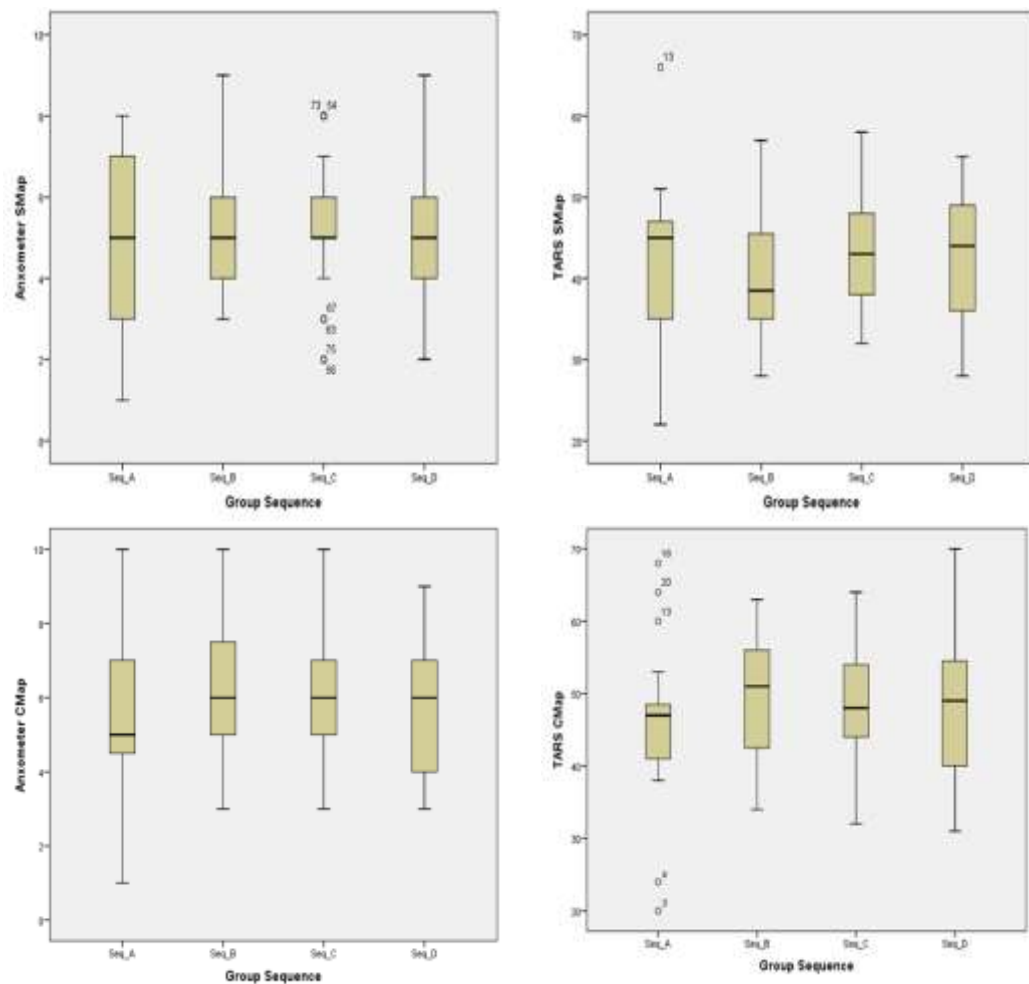
<b>SEQUENCE</b>	<b>TASK 1</b>	<b>TASK 2</b>	<b>TASK 3</b>	<b>TASK 4</b>
<b>A</b>	SM	CM	SFC	CFC
<b>B</b>	CM	SM	CFC	SFC
<b>C</b>	SFC	CFC	SM	CM
<b>D</b>	CFC	SFC	CM	SM

Table 18 and Figures 14 and 15 present the results from the ANOVAs based on task sequencing effects. The information found here demonstrates that this factor evidently had an effect on the state anxiety scores during this study. As can be seen in Table 18, all of the highest means (TARS) scores appeared when the task was performed first in the series, with the exception of the Simple Map task. However, the highest means for the Simple Map occurred when it was performed as the first of the Map tasks, upon switching the task-type, followed by the Complex Fire Chief task. Conversely, the lowest mean scores were found either when the task was performed at the end of the sequence of four tasks, or upon following a complex task.

Table 18. Descriptive statistics of TARS scores according to sequence.

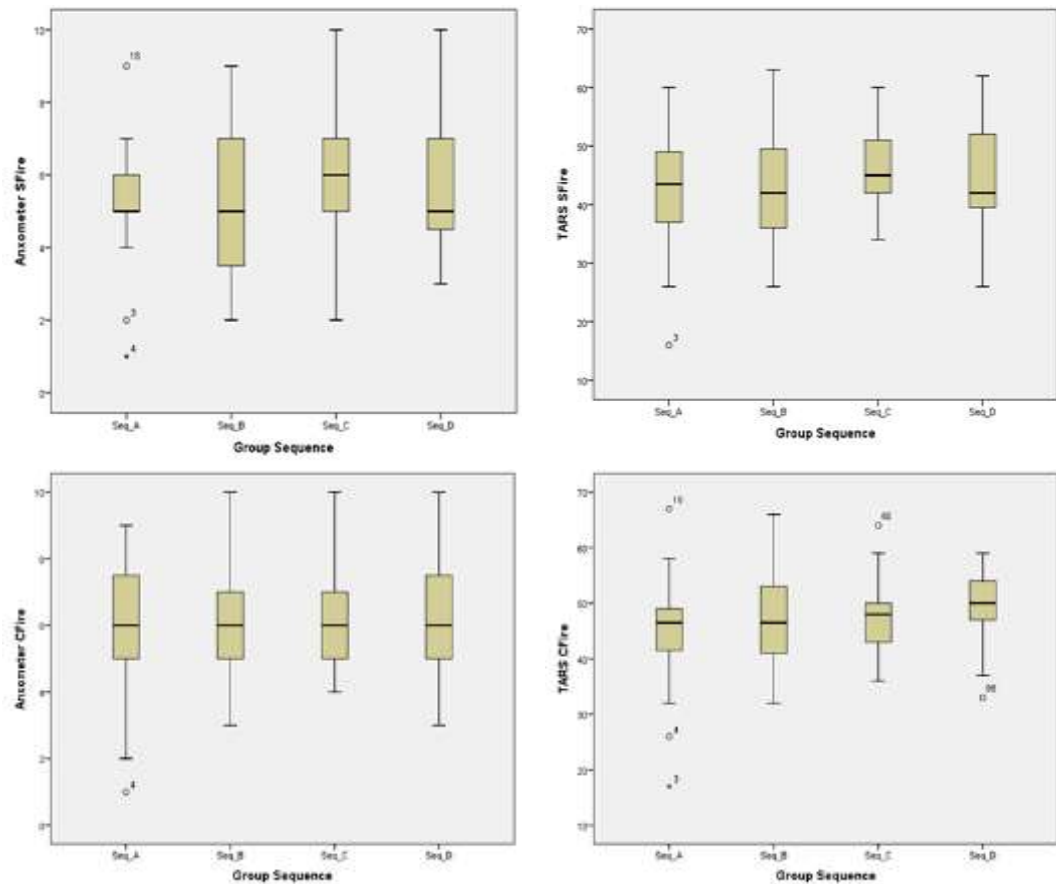
<i>TARS-TASK</i>	<i>Group Sequence</i>	<i>M</i>	<i>SD</i>	<i>N</i>
SM	Seq_A	41.96	9.494	24
	Seq_B	40.68	8.340	28
	Seq_C	43.72	7.587	29
	Seq_D	42.52	7.542	27
	Total	42.24	8.190	108
<i>TARS-TASK</i>	<i>Group Sequence</i>	<i>M</i>	<i>SD</i>	<i>N</i>
CM	Seq_A	45.71	10.369	24
	Seq_B	49.32	7.538	28
	Seq_C	47.72	7.874	29
	Seq_D	47.59	9.520	27
	Total	47.66	8.791	108
<i>TARS-TASK</i>	<i>Group Sequence</i>	<i>M</i>	<i>SD</i>	<i>N</i>
SFC	Seq_A	42.13	10.084	24
	Seq_B	41.96	9.118	28
	Seq_C	46.10	6.366	29
	Seq_D	44.41	8.395	27
	Total	43.72	8.579	108
<i>TARS-TASK</i>	<i>Group Sequence</i>	<i>M</i>	<i>SD</i>	<i>N</i>
CFC	Seq_A	45.13	10.543	24
	Seq_B	47.71	7.920	28
	Seq_C	47.66	6.201	29
	Seq_D	49.15	6.520	27
	Total	47.48	7.879	108

Figure 14. Boxplots for Map Tasks: Anxometers and TARS.



As can be seen in the boxplots of Figures 14 and 15, there were slightly varied findings between the Anxometer and TARS mean ratings. For example in Figure 14, if we look closer at the means variance of Sequence D, we can see that there is a larger range of means variance in the TARS scores for the CM. As well, it was interesting to see that there were outliers found, at both the high level and low levels of SEQ A (TARS CM) and SEQ C (Anxometer SM), which were the simple-complex sequences of both tasks. Nevertheless, it appears that the majority of the participants fell within the middle range for these sequences.

Figure 15. Boxplots for Fire Chief Tasks: Anxometers and TARS.

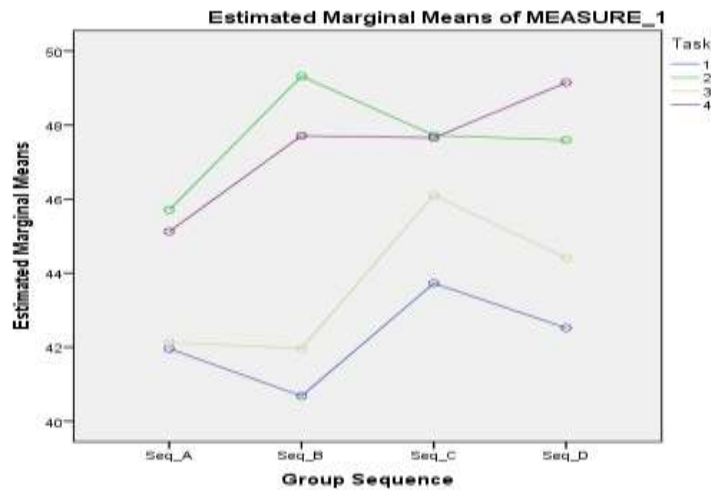


In Figure 15, the boxplots capture a more varied image of the scores yielded from the anxiety ratings of the Fire Chief tasks, especially when comparing the TARS and the Anxometer scores. Similarly to the Map tasks, the highest Anxometer ratings were found to be when the Fire Chief tasks were the first tasks of the series. Conversely, this was not the same for the TARS. According to these results, there appears to be only a slight difference seen when comparing the TARS scores of the Simple Fire Chief, regardless of the sequence. Finally, the results displayed in Figure 15 show that the Anxometer scores for the Complex Fire Chief task yielded the highest anxiety scores of all four tasks.

*Second Language Anxiety and Task Complexity*

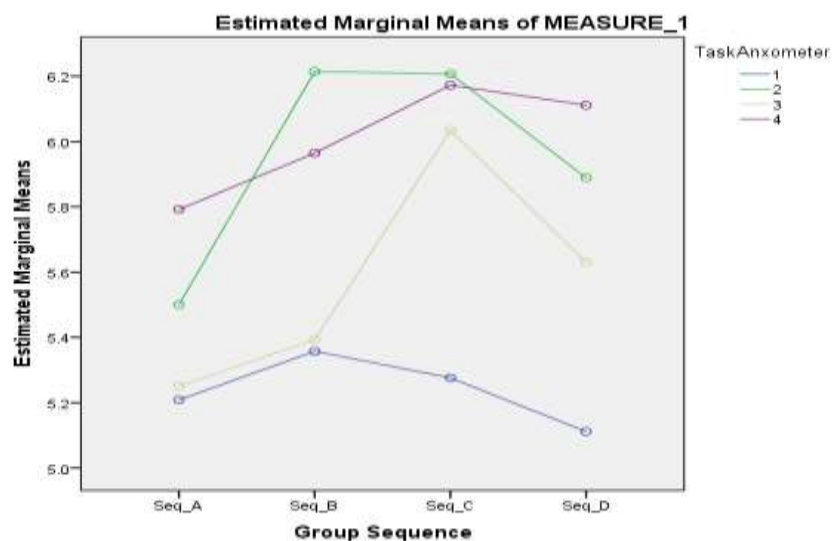
Additionally, Figures 16 and 17 visually display the results of the repeated measures ANOVA based on the sequence in which the tasks were carried out.

**Figure 16. Repeated Measures ANOVA for all four TARS.**



\*The numbers represent the tasks in the following way: 1=SM; 2=CM; 3=SFC; 4=CFC

**Figure 17. Repeated Measures ANOVA for all four Anxometers.**



\*The numbers represent the tasks in the following way: 1=SM; 2=CM; 3=SFC; 4=CFC

These Figures strongly indicate that Sequence A (SM-CM-SFC-CFC) produced the least amount of anxiety overall, especially in terms of TARS scores. Also, there is a clear distinction made here between those sequences that began with a simple task in contrast with those that began with a complex task. In other words, once again, the sequences that began with either of the Complex tasks appear to have produced more anxiety.

No further analyses were conducted with respect to answering RQ1 (C), which concerned whether there were any differences in the participants' anxiety levels regarding the task-types. Although minimal differences can be seen, according to the analyses conducted thus far. However, the differences appear to occur with regards to task complexity and task sequence rather than task-type. Once again, the Simple Map task seemed to produce the lowest levels of state anxiety. Whereas, the mean scores generated during the two Complex tasks had a similar effect (higher anxiety). Thus, there was only a difference in terms of task-type found between the Simple Map and the Simple Fire Chief. The Simple Fire Chief appears to have provoked more anxiety.

In sum, with respect to the type of task performed, it is clear that both of the Complex tasks created the most anxiety, regardless of when in the series it was performed. Whereas, the least amount of state anxiety was demonstrated upon performing the Simple Map task. Therefore, there did not appear to be a significant difference of state anxiety scores based on task-type in this study.

## **6.4. Results for research question 2.**

*To what extent do FLCAS scores predict state anxiety, as measured by the TARS and Anxometer scales completed upon oral L2 task performance?*

### *6.4.1. Descriptive exploratory analysis for FLCAS scores.*

First, a descriptive statistics analysis of all FLCAS scores was conducted according to each of the four sequences. There were 86 participants included in this part of the analysis. The data were found to be within normal limits according to the Shapiro-Wilk's test ( $p > .05$ ). Additionally, upon a visual inspection of the boxplots (Figure 18) and Q-Q plots the means appeared to be approximately normally distributed and no outliers were found. Table 19 presents the means, Standard Deviation, and the Skewness and Kurtosis for each of the four groups. Figure 18 also visually presents the means variance for each of the four sequences. The median appears to be more or less equal throughout, with the largest variance occurring within the Sequence A group. Additionally, histograms are presented in Figure 19 to highlight that although the means score are more or less the same, the distribution of scores is quite varied.



Table 19. Descriptive statistics of all FLCAS.

FLCAS SCORES (N=86)	<i>M</i>	<i>SD</i>	<i>S</i>	<i>K</i>
SEQ A <i>n</i> =21	96.04	19.77	.247	-.654
SEQ B <i>n</i> =22	94.31	23.39	-.630	-.197
SEQ C <i>n</i> =22	92.63	15.47	-.192	-1.01
SEQ D <i>n</i> =21	96.143	18.81	-.262	-1.02

Figure 18. Boxplots for FLCAS means scores by sequence.

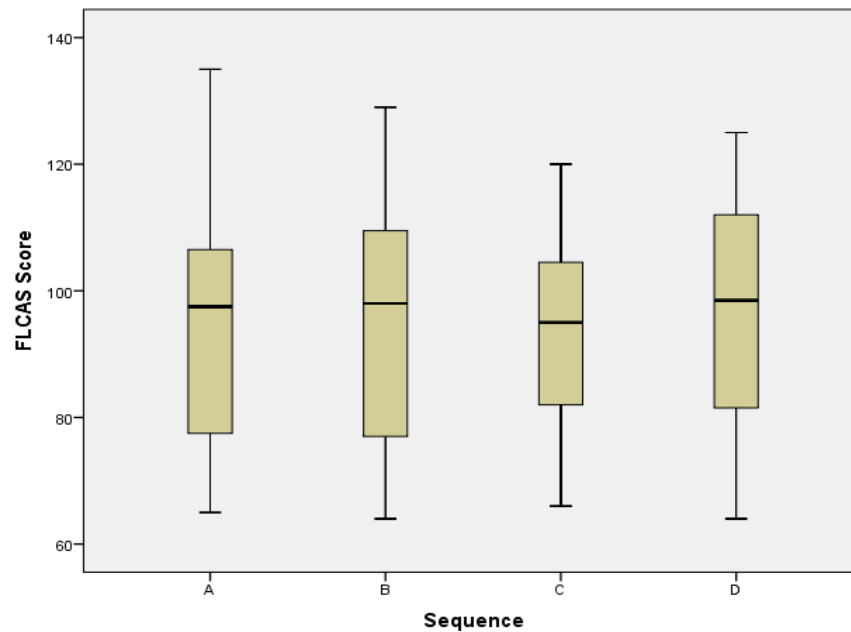
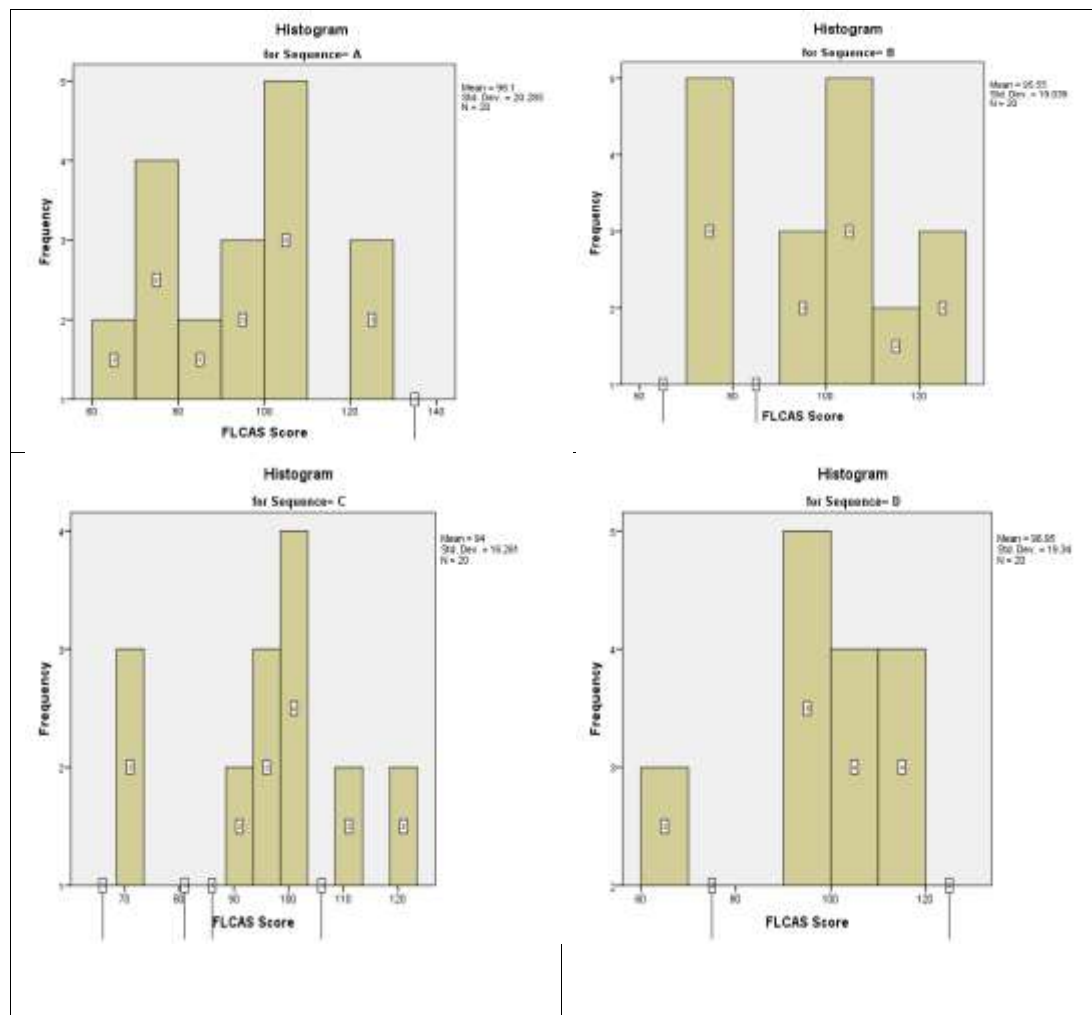


Figure 19. Histograms of FLCAS by sequence.



As displayed in the Figures and Tables here, the FLACS (overall FLA) median scores appeared to be similar for each sequence. Interestingly, however, the highest means range presented were found in Sequence A and Sequence D, which if we look back at the answer to RQ (A), had been deemed to be the least and most anxiety provoking, respectively. Conversely, referring back to the state anxiety scores that were found at these levels (see Table 18), Sequence A yielded the lowest mean scores for both of the TARS and the Anxometer scales.

#### *6.4.2. Inferential statistics.*

Next, a 2-tailed Pearson's correlation was conducted to compare if the FLCAS was, in fact, a good predictor of the state anxiety scores captured by the TARS and Anxometer. The analyses revealed a weak positive correlation between FLCAS and the TARS:  $r(84) = .25, p = .02$ ; TARS Complex Map  $r = .20$ , TARS Simple Fire Chief  $r = .26$  and TARS Complex Fire Chief  $r = .18$ . For further analysis, see Table 22, which includes a correlation of all the FLCAS scores with the TARS and Anxometers, regardless of task-type. These results indicated that the FLCAS was, therefore, not a reliable predictor for the level of state anxiety that occurred during the participants' performance of the L2 tasks included in this study.

#### **6.5. Qualitative analysis.**

In order to triangulate the quantitative analysis, an 'additional comments' option was included after each of the TARS items. Table 20 presents a list of the comments that were made by the participants, according to each item from Sequence B, (see Appendix 6, 7, and 8 for comments made during the other three sequences). Although the comments were voluntary, several participants included reflection statements that directly supported the findings for RQ1. This was especially true when the tasks were manipulated for complexity.

Table 20. Examples of participants' comments from the TARS: SEQ B.

<i>SEQ B</i>		
<i>Task construct</i>	<i>TARS item</i>	<i>Participants' statements</i>
<b>Complex Map</b>	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática. (It was difficult for me to do the task due to lack of vocabulary/grammar)	On the practice task
<i>OD</i>		masa, amasar y las escaleras mecánicas y el dibujo del mapa no era del todo claro (dough, to knead and escalator and the map drawing wasn't very clear)
<i>MD</i>		Me ha resultado difícil los mapas o dibujos que no eran claros (It was difficult for me because the pictures weren't clear)
<i>MD</i>	No estaba seguro/a de lo que tenía que hacer. (I wasn't sure what I had to do)	el mapa es un poco confuso (the map was confusing)
<i>MD</i>		la task 2 era complicado entender el mapa (task 2 was difficult to understand the map)
<i>PD</i>	He tenido tiempo suficiente para terminar la tarea. (I had sufficient time to finish the task)	no habia un tiempo limite no? Al menos yo no me he enterado (there wasn't a time limit, right? At least I didn't know)
<i>OD</i>	Sabía lo que quería decir pero no tenía las palabras para expresarme bien. (I knew what I wanted to say, but didn't have the words to express myself well)	las escaleras mecánicas (the escalator)
<i>OD/PD?</i>	Estoy convencido/a de que me he expresado bien. (I am convinced that I expressed myself well)	No me he expresado bien, además me he saltado pasos creo (I didn't express myself well, I also skipped steps, I think)
<i>MD</i>	No he tenido ningún problema para entender las instrucciones. (I didn't	he tardad un ratito en leerlas (It took me a while to read

	have any problem understanding the instructions)	the instructions)
<i>MD/CD?</i>		la task uno era bastante fácil pero la segunda el mapa era difícil de entender (the first task was quite easy whereas the second one was difficult to understand)
<i>MD</i>	El material era difícil de seguir. (the material was difficult to follow)	la task 2 si (task 2 yes)
<b>Simple Map</b>	No he tenido ningún problema para entender las instrucciones. (I didn't have any problems understanding the instructions)	Esta vez ha sido mas fácil (This time was easier)
<b>Complex Fire Chief (OD)</b>	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática. (It was difficult for me to do the task due to lack of vocabulary/grammar)	Extintor (Fire extinguisher)
<i>MD/CD?</i>	No me ha costado decidir cómo responder. (It wasn't difficult for me to decide how to respond)	the drawings were quite complicated for me to decipher
<i>CD</i>		Sometimes expressing yourself correctly is not as much of a problem as trying to resolve the issue in the most adequate way, and then making it have sense when you tell it.
<b>Simple Fire Chief (OD)</b>	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática. (It was difficult for me to do the task due to lack of vocabulary/grammar)	habian mas elementos que antes y he fallado en mucha gramatica. (there were more elements than there had been before and so I made a lot of mistakes with my grammar)
<i>OD</i>		Extintor (Fire extinguisher)
		the more tests I take the less stressed I feel
<i>CD</i>	No me ha costado decidir cómo responder. (It wasn't difficult for me to decide how to respond)	Yeah, it's like I said. You just get used to the kind of issues you're presented to (in this case, fires), and with time you start getting the grasp on it.

## 6.6. Additional analysis: Factor analysis of TARS.

As a method of assessing the reliability of the constructs that the TARS scale intended to capture, a factor analysis was conducted. The four constructs were originally meant to tap into specific areas of task-based performance that could potentially affect the state anxiety levels of the participants. To review, the four areas were: Performance Demands (PD; e.g., the anxiety that manifests while being recorded), Material Demands; the anxiety that may occur when interpreting or understanding the materials (MD), Outcome Demands (OD; anxiety cause by lacking in vocabulary/grammar to carry out task), and finally, Conceptual Demands (CD; e.g., not being able to decide how to respond, regardless of the language). For ease of understanding there are symbols next to each of the items, according to the relevant construct it was intended to fall under. We can see in the results presented in the Pattern Matrix (Table 21) that not all of the items loaded onto the proposed construct. However, a definite pattern was found.

As previously mentioned, a factor analysis was used to explore if the 16 items loaded onto the four projected factors. The analysis was conducted using an Oblimin and Kaiser Normalization method to extract the factors, the results yielded four components with an eigenvalue greater than 1. These four factors explained 60% of the total variance. First, KMO and Bartlett's test was .825 and a  $p < .001$  significance. Looking at the Pattern Matrix, which is presented in Table 21 we can see which items comprised each of the four constructs, which we will discuss further in the Discussion chapter.

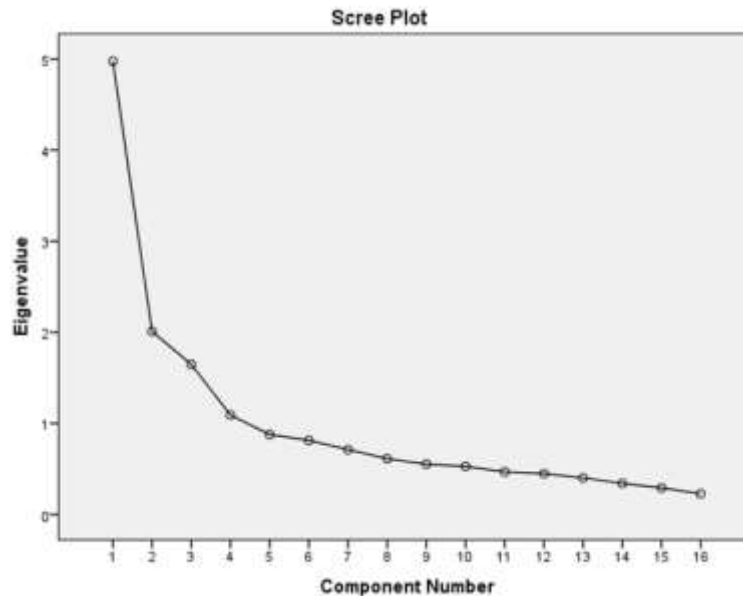
Table 21. Pattern Matrix: TARS.

Construct of item		Components			
		CD	MD	PD	OD
OD	Item1				.764
PD	Item2			.862	
PD	Item3			.568	.330
MD	Item4		.600		
CD	Item5	.729			
MD	Item6		.787		
PD	Item7		.573		
CD	Item8	.513			
OD	Item9				.816
PD	Item10			.920	
OD	Item11	.617			
MD	Item12		.759		
OD	Item13	.528			.424
CD	Item14	.764			
MD	Item15		.758		
CD	Item16	.757			

\*PD is Performance Demands, MD is Materials Demands, OD is outcome, and CD is conceptual demands

As it can be further seen in Figure 20, the scree plot illustrates that the factor loadings predominantly load onto the first construct which has been considered the Conceptual Demands. Whereas, the remaining three constructs present a clear slope, albeit slight, in relationship to one another.

Figure 20. Scree Plot Factor Analysis: TARS



### 6.7. Additional analyses: Correlation of anxiety scales.

Next, the FLCAS scores were grouped into four levels: Low (46-78), Low-Medium (79-96), High-Medium (97-109) and High (110-135). In this way, a correlation was calculated in order to analyze the relationship between the different general FLA levels as determined by the FLCAS scores, with the task-specific state anxiety levels which were measured by the TARS and the Anxometer scores, for each of the four tasks. The analysis was organized in the order in which the tasks were performed rather than the grouping the results by task-type or sequence.

The level of significance was set at  $p < .05$  and the Pearson Correlation (PC) within -1.0 and +1.0. As can be seen in Table 22, once again, there appears



to be only a mild positive correlation between the FLCAS scores and the results from each of the four TARS and Anxometer state anxiety scales.

Table 22. Correlation of all Anxiety scale scores.

<u>Tars 1</u> $r(84) = .26, p=.02$	<u>Anxometer 1</u> $r(84) = .23, p=.02$
<u>Tars 2</u> $r(84) = .27, p=.01$	<u>Anxometer 2</u> $r(84) = .35, p<.01$
<u>Tars 3</u> $r(84) = .12, p=.28$	<u>Anxometer 3</u> $r(84) = .23, p=.04$
<u>Tars 4</u> $r(84) = .26, p=.07$	<u>Anxometer 4</u> $r(84) = .28, p=.01$

## **PART IV: DISCUSSION AND CONCLUSIONS**

## **CHAPTER VII: Discussion**

### **7.1. Introduction.**

This chapter aims to elaborate upon and discuss the results of the quantitative and qualitative data presented in Chapter six. First, we will address how the analyses have answered each of the posed research questions and hypotheses.

Additionally, we will contemplate how these results compare to the theoretical and empirical studies of the task-based sequencing models with regards to the effects that task complexity, sequencing, and task-type had on task performance, as proposed by Robinson (2001, 2003, 2005, 2007; Robinson and Gilabert, 2007; Robinson, 2011, 2015) and Skehan (1996; Skehan & Foster, 1999, 2001; Skehan, 2003, 2009). Next, we will further examine the tools that were used for assessing anxiety in this study and how effective they were in measuring what they were designed to measure. Most specifically, we will explore how these results have contributed to the research gaps that were brought to light in Part I of this dissertation. Finally, the limitations as well as the implications for future studies in this area will be addressed.

## **7.2. Addressing results to research question 1.**

*How do the participants' state anxiety scores generated by the Anxometer and the Task Anxiety Reflection Scale (TARS) compare, according to:*

*(A) The complexity of the tasks (i.e., simple vs. complex)?*

The results presented in Chapter six demonstrated that task complexity appeared to have a strong effect on the participants' state anxiety levels as measured by the TARS and the Anxometer scales. As we have seen, both the Complex Map and Fire Chief tasks yielded similar overall high state anxiety ratings, especially when they were performed first in the series. Therefore, we are able to reject the null hypothesis as we had predicted.

These findings are in line with the task-based complexity research studies that have used affective measures as part of their research designs, and therefore, the predictions that were made prior to carrying out this study. Nevertheless, those studies generally have used affective measures in order to empirically support task design decisions in terms of manipulating cognitively demanding features of the L2 tasks. However, for example (Baralt & Gurzynski-Weiss, Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication, 2011), investigated whether participants experienced more state anxiety while carrying out speaking tasks comparing the modalities of a live interlocutor and computer-mediated interactive tasks. Similarly, the instrument used in that study aimed at examining the effects that both complexity and context could have on state anxiety. Again, they found no effects on the

modality; however, they found that in both contexts task complexity did appear to have an effect on the state anxiety levels of the participants. They claimed that the qualitative open-ended questions were especially revealing. This is similar to the impressions found in the study carried out here.

The participants' comments gathered during this study further supported the findings generated by the quantitative analyses of the TARS and Anxometer scores, specifically for task-complexity and sequencing. Let us consider the comments taken from the responses to the open-ended questions from Sequence A, seen in Table 23. The first one seems to suggest that the participant knew the task was more 'complex' than the previous task. Additionally, the second one provides support for a positive task repetition effect; even though the participant had just completed the complex version of the task, she/he was not anxious about it. This point had also been suggested in some of the post-pilot interviews. This contributes to the research that proposes task repetition as a possible method of avoiding or alleviating stress. (See Appendix 1 for examples of these types of studies).

Table 23. Examples of participants' comments.

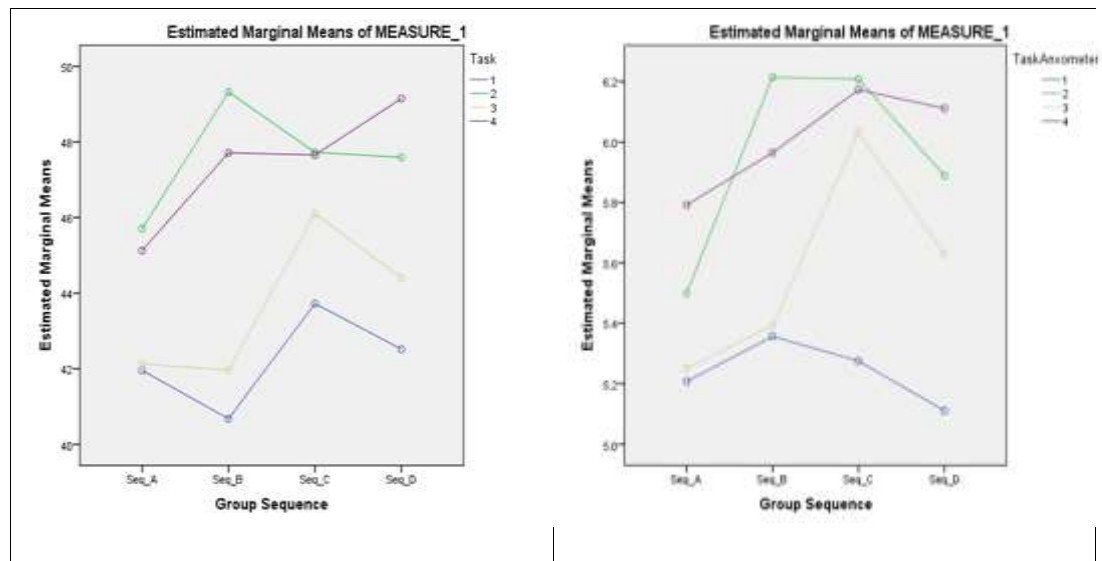
Item Statement	Participant's comment
<p>“He ido cambiando de idea sobre cómo responder.” (I kept changing my mind about how to respond.)</p>	<p>“Esta vez que me he liado más que la anterior.” (This time I got a bit more confused than during the previous one.)</p>
<p>No me ha costado decidir cómo responder. (It wasn't difficult to decide how to respond.)</p>	<p>“Yeah, it's like I said. You just get used to the kind of issues you're presented to (in this case, fires), and with time you start getting the grasp on it”.</p>

*RQ1: How do the participants' state anxiety scores generated by the Anxometer and the Task Anxiety Reflection Scale (TARS) compare, according to:*

*(B) The sequence in which the tasks were conducted (i.e., simple-complex, or complex-simple)?*

The results from the ANOVAs, which compared sequencing effects, also demonstrated that there was a strong relationship between the order in which the tasks were sequenced and the participants' responses on the state anxiety scales used for this study. Beginning with the simple task followed by the more complex task proved to be the least anxiety-provoking. Conversely, when the complex task was performed first out of the four tasks in the sequence, or first in the sequence of task-types, the anxiety ratings were at their highest for all tasks apart from the Simple Map task. These findings are also in accordance with the task-based sequencing literature that has employed 'task difficulty' rating scales (e.g., Gilabert, 2007b; Gilabert et al. 2009; Robinson, 2001; Robinson, 2007b). Furthermore, this empirically contributes to the premise that the tasks used for this study are in fact seen by the participants as differing in terms of complexity.

Figure 21. Comparison graphs of the Anxometer and TARS: Sequencing



\*The numbers represent the tasks in the following way: 1=SM; 2=CM; 3=SFC; 4=CF

Now, let us compare the difference between the Anxometer and the TARS scales in terms of task sequencing. As we can noticeably see in the two graphs displayed in Figure 21 above, which were generated from the results of the Repeated Measures ANOVA, there appears to be an effect of sequencing on the anxiety ratings which is exhibited through both scales. However, as we look closer we can see that there are slight differences when interpreting and comparing the results generated from the Anxometer and the TARS. This brings us to the topic that will be addressed when reflecting on RQ2, by further exploring the correlation between the different scales used to assess anxiety for this study. But first, we will review the effects of task-type and anxiety ratings.

*RQ1: How do the participants' state anxiety scores generated by the Anxometer and the Task Anxiety Reflection Scale (TARS) compare, according to:*

*(C) The task-type (i.e, the Map tasks or the Fire Chief tasks)?*

According to the TARS and Anxometer ratings, the effects of task- type on anxiety levels were not significant. However, when considering the qualitative feedback, several of the participants expressed a strong emotional reaction towards one task or the other. These were seen through either the written comments or verbal comments made to the researcher during post-study conversations.

Interestingly, in the stimulated-recall sessions and the post-pilot interviews, participants commented about the way they felt about the tasks, even if they did not display a high anxiety rating on the quantitative scales. Several stated that they could never see themselves being in the Fire Chief situation, and others, even gave suggestions of what they would do instead. For example, one participant stated that she would jump out the window or that she would ask someone else to take over the volunteer position and that was her complete response to the task. She was encouraged to just 'imagine' what steps she would take, however she refused to elaborate. In fact, this raised an interesting question for several participants. They began to discuss that they would not know what to do if a fire had occurred in the building. This, therefore, poses an interesting



question about the effect that task-type and cultural awareness, and moreover, the practical need for the language necessary to carry out the task.

As the Fire Chief had been developed from a ‘needs-analysis’ (Gilabert, 2005), it had served a purpose in a particular language learning context. However, as this need did not seem practical to some participants in this study, it actually made them ‘anxious’ to think about such a situation, and therefore, they stated that they felt it had affected their L2 performance. This could be seen as problematic and, in fact, in some cases cause anxiousness in L2 learning situations. These comments have strengthened the argument for using qualitative feedback to support or enhance quantitative results.

In sum, these results to RQ1 have provided evidence that the participants’ levels of state anxiety increased, especially as affected by task complexity and task sequencing. Therefore, this further contributes to the empirical findings that the tasks used in this study have been manipulated on two levels of complexity. However, what is not clear is if the instruments (the Anxometer and the TARS) are actually testing for anxiety *per se*, which poses the underlying question of how robust these scales are, and, what exactly they are testing. We will now take a further look at RQ2 in order to analyze the instruments used for this study in more depth.

### 7.3. Addressing results to research question 2.

*To what extent do the situation-specific anxiety scores of the Foreign Language Classroom Anxiety Scale (FLCAS) predict state anxiety during oral task-based performance?*

The results showed that there was only a mild positive correlation between the overall FLA of the participants (i.e., FLCAS scores) and the state anxiety scores (i.e., TARS and Anxometer) exhibited during the L2 task performance of this study. To review the FLCAS and the TARS correlation results were: SM  $r=.25$ ; CM  $r=.20$ , SFC  $r=.26$ ; CFC  $r=.18$ . Therefore, we cannot immediately reject the null hypothesis. It had been predicted that there would most likely be a high positive correlation, as we have defined situation-specific anxiety to be a stable or trait-like anxiety that would most likely occur in the L2 performance context.

Thus, even though FLCAS has been a widely accepted and utilized instrument for assessing the construct of FLA in general, it seems through the results found in this study that it may not be a strong predictor of the state anxiety during performance on these specific L2 tasks in the task-based context.

Moreover, it could be argued that even if there was a positive correlation of results at some point, the FLCAS does not provide *enough* information about the causes, e.g., understanding which aspects of the L2 task features caused the momentary or state anxiety. This poses the question of whether the FLCAS scale is a useful tool for assessing task-based anxiety. Horwitz (2001; 2010) has even

cautioned that the context in which the FLCAS is implemented is important to identify when reporting empirical results.

#### **7.4. Additional results and justification for further analysis.**

A factor analysis was conducted to see if the tailor-made scale used for this study could be empirically supported for construct validity. As a whole, the results from the factor analysis demonstrated that the 16 items loaded on four separate constructs which accounted for 60% variance. However, let us look at a more specific breakdown of the four constructs.

Table 24. Table Extraction Sums of Squared Loadings

Construct	Cumulative %
Concept Demands	31.10
Material Demands	12.57
Performance Demands	10.30
Output Demands	6.82

As can be seen in the cumulative % of the constructs demonstrated in Table 24, it appears that Concept Demands yielded the highest anxiety ratings, followed by Material Demands. If we reflect upon the context of the study, this makes sense in that the performance demands (e.g., timing or being recorded)

were not strict, i.e. there was no grade result given nor was there any real time limit. Furthermore, it could be argued that the low percentage of Output Demands was influenced by the fact that the participants in this study had a high level of English, and therefore, this did not really affect them.

Let us consider the quantitative results compared with the qualitative comments taken from Sequence D, for example, in response to the first item on the TARS scale which was the Fire Chief (see Table 25). It is interesting to see that, although 45% of the participants responded that it was difficult for them to do the task for lack of vocabulary/grammar, their comments added additional information. It is possible that because it was the first question on the anxiety reflection scale as well as the first of the sequence, they reacted with a high anxiety rating. However, here it is an example of how additional comments can give further insight into the reasons why the learner felt anxious. This further extends the research of having a mixed-method design when testing affective factors, which was a similar finding to that of Baralt & Gurzynski-Weiss (2011).

Table 25. Quantitative and qualitative comparison for Task 1 SEQ D

Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática. ( <i>It was difficult for me to do the task due to lack of vocabulary/grammar</i> ).		
Answer Options	Response Percent	Response Count
1 Muy de acuerdo ( <i>I completely agree</i> )	3.6%	1
2 De acuerdo ( <i>I agree</i> )	42.9%	12
3 Indiferente ( <i>Neither agree nor disagree</i> )	25.0%	7
4 En desacuerdo ( <i>I disagree</i> )	25.0%	7
5 Totalmente en desacuerdo ( <i>I completely disagree</i> )	3.6%	1
Comentarios (opcional):		4
<i>answered question</i>		(N=28)

## Comentarios (opcional):

- “Probablemente hubiera resultad mas fácil con un vocabulario mas extenso, pero la dificultad reside en que la descripcion interfiere con la tarea de hablar en ingles”.  
(*It probably would have been easier with a wider range of vocabulary but the difficulty lies in that the description interfered with the task of speaking in English*).

---

- “Un poco con el vocabulario del incendio y por los nervios de que los demás me escucharan”.  
(*A little bit with the vocabulary about a fire, but I was more worried that the others were listening to me*).

---

- “también es una situación difícil de describir. Sería más fácil actuar que describir los pasos de mi actuación”.  
(*It is also a difficult situation to describe. It would be easier to act out rather than to describe the steps of my actions*).

---

- Era una tasca inesperada y por eso me ha resultado difícil.  
(*It was an unexpected task and for that reason it was difficult*).

However, at this point the question should be posed, once again, about how robust the TARS is in accurately assessing *anxiety*, or, perhaps something else is being tested here. Take, for example, studies that have utilized these types of affective scales either during, or immediately after, task- based performance Baralt (2010) and Robinson (2007). Both studies included scales that had the constructs of both ‘difficulty’ and ‘anxiety’, which are arguably similar to what was captured by the TARS. However, with the combination of the Anxometer and TARS scores, as well as qualitative comments perhaps, enough information is captured. Thus, this needs further investigation.

## **CHAPTER VIII: CONCLUSIONS**

### **8.1. Limitations.**

One of the biggest limitations to this study was the sample size. A larger sample size, especially at the piloting stages could have provided further support for construct validation of all the scales used. Additionally, a more comprehensive understanding of the participants would have provided a stronger support for generalizing results. For example, the background questionnaire could have addressed the language learning experiences of participants in detail, in order to gain more knowledge about what the participants had been exposed to, and what they were doing in their current classes (e.g., task-based or communicative approaches).

Furthermore, although the participants claimed to have a high level of English, a base-line proficiency level test may have been useful to have a better indication of how this could have affected the results. Whilst they were expected to have a high level, perhaps this could have caused some anxiety. The participants themselves may have high expectations of their own language abilities, and therefore, when placed in this unexpected situation where they have not been given time to prepare, they may have reacted in an anxious way.

It would also be interesting to have replication studies or similar task-based complexity studies implementing a systematic way of selecting the tasks in combination with a consistent form of analyzing the various IDs (e.g., anxiety and working memory or motivation) that could interact with L2 performance. Take for

example the student who finds the prompts of the speaking task ‘difficult to follow’, or they feel ‘awkward or strange about being recorded’, these are insightful thoughts and helpful for those who are implementing the tasks; however, it is not clear what they tell us about the way in which the learners will react to these posed situations. For example, if one learner says that something is more ‘difficult’ or more ‘complex’, are we to consider that this is to produce more anxiety?

## **8.2. Implications for future research.**

Throughout this study both theoretically and practically, the question has been raised about which instruments for assessing anxiety are the most reliable, and in which contexts (and for which types of anxiety). Upon reflection of the tools used here, it could be argued that perhaps we have ‘fallen into the trap’ of using what Ellis (2008) cautioned against by considering quantitative measures the ‘standardized’ (i.e., FLCAS) method of assessing anxiety. If this study were to be replicated it seems that the IPOAS, which had been used in a study by Robinson (2007), may be more revealing about the different components of the language learning process that could potentially cause the participants’ state anxiety while carrying out L2 tasks. Furthermore, this appears to be in line with the task-based cognitive models that have guided this study. In personal communication with Peter Robinson, he argued that the FLCAS may not provide much insight when predicting the anxiety that manifests in the task-based



situation, whilst the IPOAS potentially could, as it is geared to test a more cognitive approach to language learning.

A further suggestion, as previously mentioned, would be to employ a more comprehensive battery of initial learner assessments that would include other IDs: proficiency, working memory, motivation, and willingness to communicate. Additionally, whilst adhering to the main constructs of the TARS, (i.e., conceptual, performance, material, and output demands), perhaps the individual items could be adjusted based on the specific features posed by the L2 task being used. For example, if the task calls for interaction, or receptive activities, there could be items added to include these aspects.

Finally, outcome measures were not included in this study as the idea was not to measure L2 performance, but to specifically examine how participants reacted to the L2 tasks based on the results of the proposed state anxiety scales. The next step, therefore, would be to see how the anxiety levels demonstrated here have affected L2 performance. Consider the following comment made by Eysenck:

Since task- irrelevant cognitive activities such as worry preempt some of the limited capacity of working memory, it is clear that they will produce decrements in the quality of performance. However, the extent to which such decrements occur should depend on the demands placed on the capacity of working memory by task-relevant information. (1979, p. 356).

Thus, using the proposed models of task sequencing in order of ‘easy’ to more ‘difficult’ tasks, as Derakshan & Eysenck (2009) and Ellis (2003) have suggested; or ‘simple’ to ‘complex’, as Robinson and Skehan have proposed, it could be assumed that this would work as a method to support those learners who become anxious during L2 task-based performance specifically. That is to say, if there are negative outcomes found in performance measures.

With regards to the TOH and CH, it could also be assumed that in terms of task complexity levels, the participants would become anxious when carrying out a more complex task. However, these models are intended to predict performance in terms of the CAF outcome measures and not affective measures; hence, it is something that should be considered. In fact, there are very few studies that look specifically at FLA and task complexity, and furthermore, how performance is affected through CAF measurements.

### **8.3. Pedagogical implications.**

The study conducted by Révész et al. (2015) set out to empirically test the effects of Cognitive Load (mental effort; task difficulty) on learner performance. The design employed a quantitative (dual-task and self-ratings for mental effort) and qualitative questionnaires for teachers to analyze the construct of complexity. This is a good example of how researchers are addressing the issue of teacher awareness of task complexity in order to support research design choices that are set forth by the task-based sequencing models. Additionally, these types of studies

could include an aspect of teachers' awareness of the anxiety that these tasks may produce.

Although the study here has proposed that the task-based sequencing models presented in Chapter I could work in supporting IDs in SLA, namely FLA, or state anxiety that may be caused by task features or situations, it is clear that this is only part of the puzzle. By working to identify this phenomenon before it becomes a situation-specific anxiety seems to be something that needs to be addressed immediately as the language learning process begins. For example, Young (1992) has suggested including some teaching strategies in the L2 classroom to help prevent and to alleviate some of the language anxiety that occurs in the language classroom.

First, it would be interesting to find out how teachers interpret task complexity and FLA, and ways in which they deal with it in the classroom. In a similar way as some have done with task 'difficulty' (e.g., Tavakoli, 2009). If teachers implement the use of a state anxiety scale during classroom activities, they could potentially gain insight on what specific task features or situations may cause these emotions. Furthermore, if a system such as those proposed by the frameworks of Robinson and Skehan are implemented systematically, then perhaps the phenomenon of FLA may even be prevented or somewhat lessened. In addition, if it is found that state anxiety occurs during performance of specific L2 tasks then the situations could be appropriately and immediately dealt with, in order to support further development and avoid possible negative outcome measures such as poor grades.

#### **8.4. Summary of findings.**

In sum, this study set out to analyze if the use of quantitative and qualitative measures could capture the state anxiety levels that participants felt during L2 task-based performance. First, the effects of sequencing in terms of task-type and task complexity on the levels of anxiety were investigated. The results demonstrated that by using a quantitative and qualitative method of analyzing state anxiety, we were able to see a difference in the levels of anxiety as affected by both task complexity and the manner in which L2 tasks were sequenced.

Secondly, the analyses of different measures of anxiety were compared. The findings indicated that the widely used FLA scale (i.e., the FLCAS) was not a strong indicator of the state anxiety levels demonstrated during the task-based performance of this study. In addition, a factor analysis was conducted to investigate if the TARS state anxiety scale, designed for this study, captured the intended constructs. Upon analyzing the results, it was found that they did comprise the four constructs which explained 60% of the total variance. Further research, however, needs to be conducted in order to verify that these scores are in fact testing anxiety in these situations and not a different construct.

## **8.5. Conclusions.**

This study set out to analyze whether task-based sequencing, in terms of task-type or task complexity, had any effect on the participants' state anxiety levels.

Through the analyses conducted here it would appear that anxiety levels were, in fact, affected by task sequencing and task complexity but there were no clear findings about task-type. This was demonstrated through both quantitative and qualitative measures. These findings, therefore, contribute to the literature providing evidence that sequencing in terms of complexity has an effect on learner factors, namely, state anxiety. This is in accordance with the theoretical models that have been postulated in this dissertation and have been described in Chapter I. Specifically, this study has taken into account how L2 task features and learner factors may interact during L2 performance.

Additionally, there were several methods of assessing anxiety in this study, and therefore, this has extended the research in this area. Most importantly, by providing evidence of results from a widely used general scale of FLA (i.e., FLCAS) in comparison with the more timely or specific state anxiety scales (i.e., the Anxometer and the TARS) used to identify learner factors during L2 task performance. It has been demonstrated here that several types of tools are needed in order to be confident that what is being tested is, in fact, what the design of the tools are meant to measure in terms of constructs. It is also clear that further studies need to be conducted which compare different types of measures, and

moreover, deciding which ones would be more appropriate for the context in which they are being used.

This study has further contributed to the research on how using different types of affective tools may be useful in increasing the knowledge we have about this topic in SLA. However, as was strongly suggested throughout the first part of this study, it is pertinent to understand what the different types of tools available (or that are tailor-made) are indeed testing. Therefore, this study has extended the research that has been conducted thus far in terms of identifying and analyzing IDs (in this case anxiety), in order to help understand the effects that task features may have on learner performance, and consequently, second language acquisition. To date, there have been only a few published studies in this area of research that have attempted to examine learners' state anxiety levels through a questionnaire of some kind, while at the same time looking at the effects of specific L2 task features. It appears that at this point, we need to be more systematic about the analysis of this construct in the context of task-based instruction and SLA. Therefore, more research of this kind is recommended at this time.

## PART V: REFERENCES

- American Psychological Association. (2010). *Manual of the American Psychological Association*. Washington, DC: American Psychological Association.
- Baralt, M. (2010). *Task complexity, the cognition hypothesis, and interaction in CMC and FTF environments*. (Unpublished doctoral dissertation): Georgetown University, Washington, DC.
- Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, 35, 689-725.
- Baralt, M., & Gurzynski-Weiss, L. (2011). Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Language Teaching Research*, 15(2), 201-229. doi:10.1177/0265532210388717
- Brennan, K. (2009). *Research synthesis of 'tasks as target' in SLA research*. (Unpublished Master's Thesis). Universitat de Barcelona, Barcelona.
- Chen, I.-J., & Chang, C.-C. (2009). Cognitive load theory: an empirical study of anxiety and task performance in language learning. *Electronic Journal of Research in Educational Psychology*, 7(2), 729-746.
- Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance. *European Psychologist*, 14(2), 168-176. doi:10.1027/1016-9040.14.2.168
- Dewaele, J.-M. (2009). Individual differences in second language acquisition. In W. C. Richie, & T. K. Bhatia (Eds.), *The new handbook of second language acquisition*. (pp. 623-646). Bingley: Emerald.
- Dewaele, J.-M. (2012). Learner internal psychological factors. In *The Cambridge handbook of second language acquisition*. (pp. 159-179). Cambridge: Cambridge University.
- Dewaele, J.-M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, 58, 911-960.

- Dörnyei, Z. & Ryan, S. (2015). *The psychology of the language learner revisited*. London: Routledge.
- Dörnyei, Z. (2003). *Questionnaires in second language research: construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Dörnyei, Z. & Csizér, K. (2012). How to design and analyze surveys in SLA research? In A. Mackey & S. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 74-94). Malden, MA: Wiley-Blackwell.
- Dörnyei, Z. & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M.H. Long (Eds.), *Handbook of second language acquisition* (pp.589-630). London: Blackwell.
- Doughty, K. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition in second language instruction*. (pp. 206-57). Cambridge: Cambridge University Press.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2008). *The study of second language acquisition (2nd edition)*. Oxford: Oxford University Press.
- Eysenck, M. W. (1997). *Anxiety and Cognition*. East Sussex: Psychology Press.
- Eysenck, M. W. (1979). Anxiety, learning, and memory; A reconceptualization. *Journal of Research in Personality*., 363-385.
- Eysenck, M. W. (1985). Anxiety and cognitive-task performance. *Personality and Individual Differences*, 6(5), 579-586.
- Ganschow, L., Sparks, R., Anderson, R., Javorsky, J., Skinner, S., & Patton, J. (1994). Differences in anxiety and language performance among high-and-low-anxious college foreign language learners. *Modern Language Journal*, 78, 41-55.
- Gardner, R. C., & MacIntyre, P. D. (1992). A student's contribution to Second Language Learning: Part 1, Cognitive Factors. *Language Teaching*, 25, 211-220.



- Gardner, R. C., & MacIntyre, P. D. (1993a). A student's contribution to Second Language Learning: Part II, Affective Factors. *Language Teaching*, 26, 1-11.
- Gardner, R. C., & MacIntyre, P. D. (1993b). On the measurement of affective variables in second language learning. *Language Learning*, 43, 157-194.
- Gilabert, R. (2005). *Task complexity and L2 narrative oral production*. (Unpublished doctoral dissertation). Universitat de Barcelona, Barcelona.
- Gilabert, R. (2007a). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, 45, 215-240.
- Gilabert, R. (2007b). The simultaneous manipulation of task complexity along planning time and [+/- here-and-now]: effects of L2 production. In M. Garcia Mayo (Ed.), *Investigating tasks in formal language learning*. (pp. 44-68). Clevedon: Multilingual Matters.
- Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning performance* (pp. 105-28). Amsterdam: John Benjamins.
- Gilabert, R., Barón, J., & Llanes, Á. (2009). Manipulating cognitive complexity across task types. *International Review of Applied Linguistics in Language Teaching*, 47, 367-95.
- Hewitt, E. K., & Stephenson, J. (2011). Foreign Language Anxiety and Oral Exam Performance: A Replication of Phillips's MLJ Study. *The Modern Language Journal*. doi:10.1111/j.1540-4781.2011.01174.x
- Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112-126. doi:10.1017/S0267190501000071
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132.
- Jackson, D. O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A synthesis and a meta-analysis of research on second language task complexity. *Language Learning*, 2, 330-67.
- Jane, W. (1996). *A framework for task-based learning*. London: Collins.

- Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System*, 37, 254-68.
- Kim, Y., & Tracy-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).
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Reading: Addison-Wesley Longman Ltd.
- Larsen-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(S1), 127-159.
- Larsen-Hall, J. (2012). How to run statistical analyses. In A. Mackey, & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide, first edition*. Oxford: Blackwell Publishing Ltd.
- Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26, 27-55.
- Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Long, M.H. (2005). *Second language learning needs analysis*. Cambridge: Cambridge University Press.
- MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal*, 79(1), 90-99.
- MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of anxiety in language learning: A review of the literature. *Language Learning*, 41, 85-117.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*(44), 283-305.
- MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, 47, 265-287.

- MacIntyre, P. D., & Gregersen, T. (2012). Affect: The role of language anxiety and other emotions in language learning. In *Psychology for language learning: Insights from research, theory and practice*. (pp. 103-18). New York: Palgrave Macmillan.
- Mackay, J. (2015). *An ideal L2 self intervention: Implications for self-concept, motivation, and engagement with the target language*. (Unpublished doctoral dissertation). Universitat de Barcelona, Barcelona.
- Malicka, A. (2014). The role of task sequencing in monologic oral production. In P. Robinson (Ed.), *Task sequencing and instructed second language learning*. (pp. 71-93). London: Bloomsbury Academic.
- McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication research reports*, 5, 108-113.
- Michel, M. (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. 144-174). Amsterdam: John Benjamins.
- Norris, J. M. (2010, May). Understanding constructs in second language acquisition. *Seminario de investigacion conducted at the University of Spain*. Spain.
- Norris, J.M., & Ortega L. (2003). Defining and measuring SLA. In C. J. Doughty, & M.H. Long (Eds.), *Handbook of second language acquisition* (pp.716-761). London: Blackwell.
- Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris, & L. Ortega (Eds.), *Synthesizing research on language learning and teaching*. Philadelphia: John Benjamins.
- Norris, J.M., & Ortega L. (2009). Towards and organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

- Perez-Paredes, P., & Martínez Sanchez, F. (2000-2001). A Spanish version of the FLCAS: Revisiting Aida's Factor Analysis. *RESLA*, *14*, 337-352.
- Pica, T., Kang, H., & Sauro, S. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition*, *28*(2), 305-35.
- Porte, G. K. (2010). *Appraising Research in Second Language Learning: A practical approach to critical analysis of quantitative research (2nd ed.)*. Amsterdam: John Benjamins.
- Prabhu, N. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437-470.
- Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *Modern Language Journal*, *95*, 168-181.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62* (1), 93-132.
- Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, *35*, 87-92.
- Révész, A., Michel, M., & Gilabert, R. (2015). Measuring cognitive task demands using dual-task methodology, subjective self-ratings and expert judgements. *Studies in Second Language Acquisition*, 1-35.
- Robinson, P. (2001a). *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interaction in a componential framework. *Applied Linguistics*, *22*, 27-57.
- Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, *21*(2), 45-105.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Journal of Applied Linguistics*, *43*(1), 1-32.

- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, and perceptions of task difficulty. *International Journal of Applied Linguistics in Language Teaching*(45), 191-213.
- Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of task sequencing. In M. Putz, & L. Sicola (Eds.), *Inside the Learner's Mind: Cognitive Processing in Second Language Acquisition*. (pp. 243-268). Amsterdam/Philadelphia: John Benjamins.
- Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, 61, 1-36.
- Robinson, P. (2015). The Cognition Hypothesis, second language task demands and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and Directions in TBLT: Plenaries of a Decade of the International Conference* (pp. 87-122). Amsterdam: John Benjamins.
- Robinson, P., & Gilabert, R. (2007). Task Complexity, the Cognition Hypothesis and second language and performance. *International Review of Applied Linguistics in Language Teaching*, 45, 161-175.
- Samuda, V., & Bygate, M. (2005). Integrative planning through the use of task repetition. In *Planning and task performance in a second language*. (pp. 119-165). Amsterdam: John Benjamins.
- Sasayama, S., Malicka, A., & Norris, J. (forthcoming). *Cognitive task complexity: A research synthesis and meta-analysis*. Amsterdam: John Benjamins.
- Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety literature. *Language Learning*, 28(1), 129-142.
- Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1-14.
- Skehan, P. (2007, Sept. 20th – 22nd). Tradeoff and Cognition: Two hypotheses

regarding attention during task-based performance. Second International Conference on TBLT. University of Hawaii.

- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30, 510-532.
- Skehan, P., & Foster, P. (1997). Task type and 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, 93-120.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction*. (pp. 183-205). Cambridge: Cambridge University Press.
- SurveyMonkey Inc. (1999-2016). Retrieved from [www.surveymonkey.com](http://www.surveymonkey.com).
- Tavakoli, P. (2009). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, 19(1), 1-25.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439-73.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In *Planning and task performance in a second language*. Amsterdam: John Benjamins.
- Tobias, S. (1990). *Anxiety and cognitive processing of Instruction*. New York: City College, City University of New York.
- Van den Branden, K. (Ed.). (2006). *Task-based language education: From theory to practice*. Cambridge: Cambridge University Press.
- VanPatten, B. (1990). Attending to form content in the input. *Studies in Second Language Acquisition*, 12, 287-301.
- Wilkins, D. (1976). *The lexical syllabus: A new approach to language teaching*. London: Collins.
- Willis, J. (1996). *A framework for task-based learning*. London: Collins.

- Yerkes and Dodson, H. -D. (2007). The Temporal Dynamics Model of Emotional Memory Processing: A Synthesis on the Neurobiological Basis of Stress-Induced Amnesia, Flashbulb and Traumatic Memories, and the Yerkes-Dodson Law. *Nueral Plasticity*, 33. doi:doi:10.1155/2007/60803. PMID 17641736., CC0,  
<https://commons.wikimedia.org/w/index.php?curid=34030428>
- Young, D. (1990). An investigation of students' perspecitves on anxiety and speaking. *Foreign Language Annals*, 23(6), 539-553.
- Young, D. (1992). Creating a low-anxiety classroom environment: What does the anxiety research suggest? *The Modern Language Journal*, 426-439.

## **PART V: APPENDICES**



**Appendix 1.** Task as Target (TAT) Studies: Planning and Task Repetition.

<b>Researcher(s)</b>	<b>Tasks used in study</b>	<b>Theory/Researcher driving task selection</b>	<b>Constructs or aspects targeted</b>
<b>Planning Studies</b>			
Foster & Skehan (1996)	Information gap, narrative, decision-making	Crookes (1989), 'Limited capacity' Kahneman (1973); Skehan (1996)	SP, task type, guided vs. unguided
Skehan & Foster (1997)	Narratives, Decision-making	Continued working from results/ideas in previous study	SP, trade-off effects, task complexity and accuracy, post-task
Mehnert (1998)	Instruction task, Exposition task	Clark & Clark (1977); Faerch & Kasper (1983;1986); Crookes (1989); and above	Easy and difficult, different planning times
Foster & Skehan (1999)	Decision-making: 'Balloon debate'	Ellis (1987); Crookes (1989); Mehnert (1998); Foster & Skehan (1996)	FoM vs. FoF, guided vs. unguided
Ortega (1999)	Monologic Narrative re-tell	Ellis (1987); Crookes (1989); Foster & Skehan (1996); Doughty and Williams; (1998); Long (1991)	SP, FoF, Learner strategies
Yuan & Ellis (2003)	Picture-based Narratives	Wendel, (1997), Foster & Skehan, (1996;1999)	SP vs. OLP methods
Yuan & Ellis (2005)	Narratives	Baddeley (1986); Ellis (1987); Crookes (1989); Foster & Skehan (1999); Wendel (1997); Tarone; Levelt; Kellog; etc.	OLP, Pressured vs unpressured, writing vs. speaking
Ortega (2005)	Monologic Narrative	Ellis (1987); Crookes (1989); Foster & Skehan (1996); Ortega (1999)	Learner driven, FoM vs. FoF
Kawauchi (2005)	Picture-based Narratives	Wigglesworth (1997); Foster & Skehan (several studies by these)	Different learner proficiency, type of sp strategies
Sanguran (2005)	Instruction, Argumentative	Levelt (1989); Foster & Skehan (1996); Skehan (1996); Mehnert (1998)	SP, FoM vs. FoF
Skehan & Foster (2005)	Decision-making tasks	Same as above	OLP, FoF, group planning
Gilbert, R. (2005); (2007)	Narratives: here-&-now vs. there-&-then	Foster & Skehan (1996); Mehnert (1998); Ortega (1999); Skehan & Foster (1997)	Unplanned vs. planned, complexity
Guará-Taveras (2009) ( <i>also complexity &amp; Ind. Diff</i> )	There -& -then, picture narratives	Mehnert (1998); Robinson (1995)	+/- planning time, WM and speech production

<b>Task-Repetition studies: (or familiarity)</b>			
Gass, <i>et al.</i> (1999)	Narrative, video retell	Bygate (1996), IPT-Levelt, Skehan & Foster, (1997;1999)	FoF
Bygate (1999)	Argumentative, Narrative	Skehan & Foster; IPT (1989)	Task Type; grammatical complexity
Bygate (2001)	'Narrative set', 'Interview set'	IPT-Levelt (1989)	
Lynch and Mclean (2000)	'Poster Carousel'	Bygate (1996); Foster (1998); Skehan (1998); & above of Foster & Skehan; Lynch & Mclean (1994), Plough & Gass (1993)	Time pressure, short intervals between tasks
Samuda (2001)	Argumentative, Narrative	Above	FoF and FoM
Bygate & Samuda (2005)	Cartoon narration	McCloughlin (1990), Skehan (1998)	
Bygate et al., (2007)	Ditogloss and Jigsaw	Previous studies	Task repetition and effects on oral performance
<b>Interaction Studies:</b>			
Pica, T., Kanagy, R., & Falodun, J. (1993)	See Table 2	Long, Duff, Doughty, & many others	Evaluate features of communicative tasks & tasks types
Lee, L. (2002)	Online essay writings, open discussions	Pica <i>et al.</i> (1993) & others	Compare to live communication, are negotiation processes similar?
Gonzalez-Lloret, M. (2003)	3D simulation, Information gap, one-way giving instr.	Daughty & Long (2002); Chappelle (1998), Daughty (2000), Robinson	Task structure & how it affects communication
Hardy, and Moore (2004)	Video, computer, observation and true/false pairwork,	Robinson (2001), Skehan (1996)	Familiar Content vs. Unfamiliar Content; High vs. Low support task
Pica, T., Kang, H., & Sauro, S. (2006)	Information gaps: grammar communication, spot the diff and jigsaw	Loschky and Bley-Vroman, (1993), & others	Assessing and comparing tasks types, attentional & interactional processes
Lampert & Engler, (2007)	Narratives/picture seq., Determining Who's responsible solving, arranging times to meet.	Long (1989); Robinson (2001), Skehan (1996), Duff (1986)	Task structure (6 different versions), Sequencing
Fernandez Garcia, M. (2007)	Information gap; one-way vs. two-way	Pica (1993); Shrum & Glisan (2000)	Task types and how negotiation occurs

(From Brennan, 2009)

*Second Language Anxiety and Task Complexity*

**Appendix 2. FLCAS-Castellano**

*Instrucciones: Las siguientes afirmaciones se refieren a diversas situaciones frecuentes en el aprendizaje de un idioma. Su tarea consiste en valorar su grado de acuerdo o desacuerdo con cada una de las siguientes afirmaciones, utilizando para ello la escala puesta.*

	1= Estoy totalmente de acuerdo	2= Estoy de acuerdo	3= No sé	4= No estoy de acuerdo	5= Estoy totalmente en desacuerdo
1. Nunca estoy completamente seguro de mí mismo cuando hablo en la clase de idioma extranjero.	1. 0	2. 0	3. 0	4. 0	5. 0
2. No me preocupa cometer errores en clase.	1. 0	2. 0	3. 0	4. 0	5. 0
3. Tiemblo cuando sé que me van a preguntar en clase	1. 0	2. 0	3. 0	4. 0	5. 0
4. Me asusta no entender lo que el profesor está diciendo en idioma extranjero.	1. 0	2. 0	3. 0	4. 0	5. 0
5. No me molestaría en absoluto asistir a más clases de idioma extranjero.	1. 0	2. 0	3. 0	4. 0	5. 0
6. Durante la clase, me doy cuenta pienso en cosas que no tienen nada que ver con la clase.	1. 0	2. 0	3. 0	4. 0	5. 0
7. Pienso que a los otros compañeros se les dan mejor idiomas que a mí.	1. 0	2. 0	3. 0	4. 0	5. 0
8. Normalmente estoy a gusto cuando hago exámenes en clase.	1. 0	2. 0	3. 0	4. 0	5. 0
9. Me pongo muy nervioso cuando tengo que hablar en clase y no me he preparado bien.	1. 0	2. 0	3. 0	4. 0	5. 0
10. Me preocupa las consecuencias que pueda traer el suspender.	1. 0	2. 0	3. 0	4. 0	5. 0
11. No entiendo por qué alguna gente se siente tan mal por las clases de idioma extranjero.	1. 0	2. 0	3. 0	4. 0	5. 0
12. En clase, me pongo tan nervioso que se me olvidan algunas cosas que sé.	1. 0	2. 0	3. 0	4. 0	5. 0
13. Me da corte salir voluntario en clase.	1. 0	2. 0	3. 0	4. 0	5. 0
14. Creo que no me pondría nervioso si hablara el idioma extranjero con una persona nativa.	1. 0	2. 0	3. 0	4. 0	5. 0
15. Me irrita no entender lo que el profesor está corrigiendo	1. 0	2. 0	3. 0	4. 0	5. 0
16. Aunque vaya con la clase preparada, me siento nervioso.	1. 0	2. 0	3. 0	4. 0	5. 0
17. A menudo no me apetece ir a clase.	1. 0	2. 0	3. 0	4. 0	5. 0
18. Me siento seguro a la hora de hablar en la clase.	1. 0	2. 0	3. 0	4. 0	5. 0
19. Me da miedo que mi profesor corrija cada fallo que cometo.	1. 0	2. 0	3. 0	4. 0	5. 0
20. Siento cómo mi corazón palpita cuando sé que me van a pedir que intervenga en clase.	1. 0	2. 0	3. 0	4. 0	5. 0

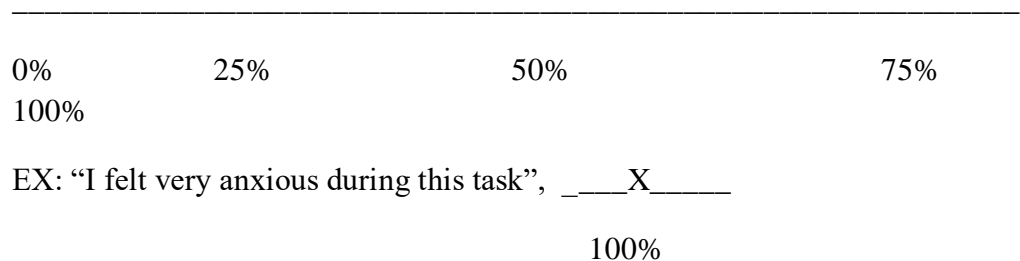
*Second Language Anxiety and Task Complexity*

21. Cuanto más estudio, más me lío.	1. O	2. O	3. O	4. O	5. O
22. No tengo ninguna presión ni preocupaciones para prepararme bien las clases.	1. O	2. O	3. O	4. O	5. O
23. Tengo la sensación de que mis compañeros hablan el idioma extranjero mejor que yo.	1. O	2. O	3. O	4. O	5. O
24. Me da mucho corte hablar en la lengua extranjera delante de mis compañeros.	1. O	2. O	3. O	4. O	5. O
25. Las clases transcurren con tal rapidez que me preocupa quedarme atrasado.	1. O	2. O	3. O	4. O	5. O
26. Comparativamente, estoy más tenso y me siento más nervioso en la clase de idioma extranjero que en otras clases o que en mi propio trabajo.	1. O	2. O	3. O	4. O	5. O
27. Me pongo nervioso mientras hablo en clase.	1. O	2. O	3. O	4. O	5. O
28. Antes de entrar a clase, me siento seguro y relajado.	1. O	2. O	3. O	4. O	5. O
29. Me pongo nervioso cuando no entiendo cada una de las palabras que mi profesor dice.	1. O	2. O	3. O	4. O	5. O
30. Me abruma la cantidad de cosas que hay que aprender para poder hablar otro idioma.	1. O	2. O	3. O	4. O	5. O
31. Temo que mis compañeros de clase se rían de mí cuando hablo en otro idioma.	1. O	2. O	3. O	4. O	5. O
32. Creo que me sentiría a gusto hablando entre nativos que hablan el idioma que estudio.	1. O	2. O	3. O	4. O	5. O
33. Me pongo nervioso cuando el profesor pregunta cosas que no me he podido preparar.	1. O	2. O	3. O	4. O	5. O

*(From Pérez-Paredes & Martínez Sanchez 2000-01)*

**Appendix 3.** Anxometer and TARS from the pilot phase.

A) Mark along the line how anxious you felt while completing the task:



B) Rate the following statements according to the task you have just completed by circling the corresponding number indicated below. There is no right/wrong answer, so respond as honestly as possible.

1. I found this task challenging because I didn't have the vocabulary/grammar to accurately complete it.	1	2	3	4	5
2. I was confused about what I was supposed to do in order to complete the task correctly.	1	2	3	4	5
3. I knew all the related vocabulary but I was anxious about the recording and therefore forgot how to say some words.	1	2	3	4	5
4. I understood all the vocabulary included in the tasks.	1	2	3	4	5
5. It was easy to decide which candidate(s) would be the best for the position(s).	1	2	3	4	5
6. I felt prepared to complete the tasks.	1	2	3	4	5
7. I couldn't decide the best candidate(s) and this affected my response.	1	2	3	4	5
8. I knew what I wanted to say but I could not express myself well.	1	2	3	4	5
9. I didn't mind being recorded.	1	2	3	4	5

*Second Language Anxiety and Task Complexity*

Reflection Statements	Totally Agree	Partially Agree	Neither agree or disagree	Partially Disagree	Totally Disagree
10. I'm confident that I expressed myself accurately.	1	2	3	4	5
11. I had no problem understanding the instructions given.	1	2	3	4	5
12. I'm afraid that I made many mistakes during this task.	1	2	3	4	5
13. Some of the vocabulary given was difficult and therefore I had to guess.	1	2	3	4	5
14. I think I made a few mistakes which might make it difficult to understand my response.	1	2	3	4	5
15. I felt the goal of the activity was clear and I knew exactly what was expected of me.	1	2	3	4	5
16. The task was easy to do because I knew exactly what to say.	1	2	3	4	5
17. I had problems understanding the materials which affected my response.	1	2	3	4	5
18. I think what I said on the recording could be easily understood.	1	2	3	4	5

**Appendix 4. Item analysis Correlation Matrix from Pilot of Spanish TARS**

		Correlation Matrix															
		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16
Correlation	Item 1	1.000	.220	-.011	-.107	-.258	-.095	.296	.609	.053	-.524	-.229	.096	.404	-.277	.039	.374
	Item 2	.220	1.000	.299	-.425	-.203	-.295	.320	.250	-.245	-.173	-.423	.211	.134	-.359	.363	.293
	Item 3	-.011	.299	1.000	-.074	-.043	-.273	.036	.050	-.717	-.091	-.228	.415	.132	-.150	.248	.241
	Item 4	-.107	-.425	-.074	1.000	.183	.239	-.230	-.118	.193	.194	.506	-.138	.004	.311	-.344	-.096
	Item 5	-.258	-.203	-.043	.183	1.000	.254	-.227	-.283	.069	.306	.176	-.212	-.388	.327	-.316	-.556
	Item 6	-.095	-.295	-.273	.239	.254	1.000	-.126	-.131	.300	.128	.162	-.451	-.162	.193	-.231	-.223
	Item 7	.296	.320	.036	-.230	-.227	-.126	1.000	.365	-.098	-.321	-.264	.062	.127	-.350	.051	.310
	Item 8	.609	.250	.050	-.118	-.283	-.131	.365	1.000	-.068	-.326	-.110	.229	.505	-.235	.110	.456
	Item 9	.053	-.245	-.717	.193	.069	.300	-.098	-.068	1.000	.088	.249	-.493	-.125	.145	-.242	-.300
	Item 10	-.524	-.173	-.091	.194	.306	.128	-.321	-.326	.088	1.000	.361	.082	-.171	.514	.143	-.299
	Item 11	-.229	-.423	-.228	.506	.176	.162	-.264	-.110	.249	.361	1.000	-.059	-.060	.474	-.243	-.211
	Item 12	.096	.211	.415	-.138	-.212	-.451	.062	.229	-.493	.082	-.059	1.000	.402	.009	.474	.350
	Item 13	.404	.134	.132	.004	-.388	-.162	.127	.505	-.125	-.171	-.060	.402	1.000	-.123	.375	.585
	Item 14	-.277	-.359	-.150	.311	.327	.193	-.350	-.235	.145	.514	.474	.009	-.123	1.000	.035	-.320
	Item 15	.039	.363	.248	-.344	-.316	-.231	.051	.110	-.242	.143	-.243	.474	.375	.035	1.000	.460
	Item 16	.374	.293	.241	-.096	-.556	-.223	.310	.456	-.300	-.299	-.211	.350	.585	-.320	.460	1.000

## Appendix 5. The Task Anxiety Reflection Scale (TARS)

**TASKS SEQ B**

**Escala de reflexión de la tarea oral- Task 4**

**Califica las siguientes declaraciones con la numeración correspondiente en relación a las tareas realizadas. No existe una respuesta correcta o incorrecta, puntúa los comentarios con la mayor franqueza posible.**

1. Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.

1     2     3     4     5  
 Muy de acuerdo    De acuerdo    Indiferente    En desacuerdo    Totalmente en desacuerdo

Comentarios (opcional):

2. El hecho de que me estuvieran grabando me ha puesto nervioso/a.

1     2     3     4     5  
 Muy de acuerdo    De acuerdo    Indiferente    En desacuerdo    Totalmente en desacuerdo

Comentarios (opcional):

3. Me he sentido bajo presión para completar la tarea a tiempo.

1     2     3     4     5  
 Muy de acuerdo    De acuerdo    Indiferente    En desacuerdo    Totalmente en desacuerdo

Comentarios (opcional):

4. No estaba seguro/a de lo que tenía que hacer.

1     2     3     4     5  
 Muy de acuerdo    De acuerdo    Indiferente    En desacuerdo    Totalmente en desacuerdo

Comentarios (opcional):

5. Me ha resultado fácil decidir cómo responder.

1     2     3     4     5  
 Muy de acuerdo    De acuerdo    Indiferente    En desacuerdo    Totalmente en desacuerdo

Comentarios (opcional):



6. He entendido todo de los materiales.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

7. He tenido tiempo suficiente para terminar la tarea.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

8. He ido cambiando de idea sobre cómo responder.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

9. Sabía lo que quería decir pero no tenía las palabras para expresarme bien.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

10. No me importaba el hecho de que me grabaran.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

11. Estoy convencido/a de que me he expresado bien.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

12. No he tenido ningún problema para entender las instrucciones.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

13. No creo que haya cometido errores que puedan dificultar que se entienda mi respuesta.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

14. Sabía exactamente qué quería decir para completar la tarea.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

15. El material era difícil de seguir.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

16. No me ha costado decidir cómo responder.

- 1 Muy de acuerdo     2 De acuerdo     3 Indiferente     4 En desacuerdo     5 Totalmente en desacuerdo

Comentarios (opcional):

**Appendix 6.** Participants' Comments on the TARS: SEQ A

SA	TARS Statement	Comment
SM	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Solamente ha ocurrido esto en la primera prueba, al ser un vocabulario algo más específico.
		Estaba más preocupada por los programas informáticos que por el vocabulario
		Aunque había palabras como la masa de la pizza o delantal que no conocía.
	El hecho de que me estuvieran grabando me ha puesto nervioso/a.	anteriormente no me había grabado
		Yo soy una persona que incluso en una clase llena a reventar habla en inglés en voz alta. No obstante, me ha parecido un tanto ridículo. Me ha puesto más nervioso que un profesor real.
	Estoy convencido/a de que me he expresado bien.	En la primera task sí pero en la segunda no.
	Sabía exactamente qué quería decir para completar la tarea	no EXACTAMENTE
CM	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	tenía claro el vocabulario y la gramática
	No estaba seguro/a de lo que tenía que hacer.	No se ve nada bien la imagen (no se localizar donde están las cosas en el dibujo)
	Me ha resultado fácil decidir cómo responder.	he tenido que pensar mucho la respuesta antes de responder porque no sabía cómo expresarme claramente aunque sabía lo que tenía que decir.
	He entendido todo de los materiales.	algunas cosas no sabía cómo decir las
	He ido cambiando de idea sobre cómo responder.	he tenido que reestructurar varias veces mi respuesta
CFC	Me ha resultado fácil decidir cómo responder.	he tenido que pensar bastante cómo estructurar mi respuesta correctamente
	No creo que haya cometido errores que puedan dificultar que se entienda mi respuesta	podría haber sido más clara en mis instrucciones
	Sabía exactamente qué quería decir para completar la tarea	he tenido que reestructurar varias veces mi respuesta

**Appendix 7.** Participants' Comments on the TARS: SEQ C

SEQC	TARS Statement	Participants' statements
SFC	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Ha sido uno de los factores, pero no el más importante
		I realized one mistake I made
	Me he sentido bajo presión para completar la tarea a tiempo.	Es una de las cosas que me ha puesto más nerviosa
		He hecho lo que he podido, pero al final no lo he encontrado tan difícil como esperaba, simplificando la historia.
	He ido cambiando de idea sobre cómo responder.	He ido hablando sobre la marcha.
	Sabía lo que quería decir pero no tenía las palabras para expresarme bien.	Tenía las palabras pero a veces no me salían a tiempo por los nervios
	Estoy convencido/a de que me he expresado	No estoy segura porque de los nervios no recuerdo exactamente como me he expresado (espero que bien)
	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	No creo que haya cometido errores que puedan dificultar que se entienda mi respuesta.
		No estoy segura porque de los nervios no recuerdo exactamente como me he expresado (espero que bien)
		Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.
	Sabía exactamente qué quería decir para completar la tarea.	Tenía las ideas básicas en mi cabeza, lo complejo era explicarlas
CFC	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Hay palabras que no me salían en este momento.
	Me he sentido bajo presión para completar la tarea a tiempo.	Esto es lo que me ha puesto más nerviosa porque he empezado a hablar muy tarde, me he bloqueado por los nervios y no leía las instrucciones!

## Second Language Anxiety and Task Complexity

		Al parecerse a la task anterior ya sabía de qué iba el tema
	He ido cambiando de idea sobre cómo responder.	Esta vez que me he liado más que la anterior
	Sabía lo que quería decir pero no tenía las palabras para expresarme bien.	Algunas no me salían en el momento indicado
		smoke,breathe...las escalerasdelosbomberos ?
	Estoy convencido/a de que me he expresado bien.	Esta vez considero que me he expresado bastante peor
	No creo que haya cometido errores que puedan dificultar que se entienda mi respuesta.	Quizá la forma de expresarme haya dificultado la comprensión
SM	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Pese a ser vocabulario típicamente enseñado en la escuela, al no haberlo puesto en práctica lo he ido olvidando
	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Siempre me ha costado un poco dar indicaciones, quizás por falta de vocabulario, pero en mi opinion me falta un poco de base en esto.
	Sabía lo que quería decir pero no tenía las palabras para expresarme bien.	Me ha faltado recordar vocabulario específico de indicaciones
	Estoy convencido/a de que me he expresado bien.	Creo que me he hecho entender (o eso espero)
	No creo que haya cometido errores que puedan dificultar que se entienda mi respuesta.	Quizá la falta de vocabulario específico haya dificultado ligeramente la comprensión
	He entendido todo de los materiales.	Era un poco difícil de entender en que planta estaba la comida de perro, los gráficos no se entendían muy bien.
	Sabía lo que quería decir pero no tenía las palabras para expresarme bien.	Me he quedado atascada unas cuantas veces, intentando procesar lo que tenía que decir.
		again 'floor'

**Appendix 8.** Participants' Comments on the TARS: SEQ D

SEQD	TARS Statement	Comment
CFC	Me ha resultado difícil hacer la tarea por falta de vocabulario/gramática.	Probablemente hubiera resultado más fácil con un vocabulario más extenso, pero la dificultad reside en que la descripción interfiere con la tarea de hablar en inglés.
		Un poco con el vocabulario del incendio y por los nervios de que los demás me escucharan
		también es una situación difícil de describir. Sería más fácil actuar que describir los pasos de mi actuación.
		Era una tarea inesperada y por eso me ha resultado difícil.
	No estaba seguro/a de lo que tenía que hacer.	Al principio
	Me ha resultado fácil decidir cómo responder.	No entendía muy bien el dibujo
		Me ha costado porque realmente no se que es lo que hay que hacer durante un incendio y los dibujos no erae muy claros
	He entendido todo de los materiales.	Una vez he terminado
		Aunque no mucho
	He ido cambiando de idea sobre cómo responder.	Si, porque cada vez que grababa la tarea, veía que no había respondido del todo bien
	Sabía lo que quería decir pero no tenía las palabras para expresarme bien.	También me sentía observada
	Estoy convencido/a de que me he expresado bien.	Podría haberme expresado mejor
	Sabía exactamente qué	No se me ocurría nada
	quería decir para completar la tarea.	
CM		Me he olvidado de como decir una "mazana de casas"