THE ROLE OF FL APTITUDE AND THE EXECUTIVE FUNCTIONS OF
WORKING MEMORY AND INHIBITION IN FL VOCABULARY ACQUISITION BY
YOUNG GREEK LEARNERS OF ENGLISH

by

Lia Efstathiadi

A thesis submitted for the degree of
Doctor of Philosophy in Applied Linguistics

at

Aristotle University of Thessaloniki
Department of Theoretical and Applied Linguistics
School of English, Faculty of Philosophy

December 2013
One language sets you in a corridor for life. Two languages open every door along the way.

Frank Smith
This thesis is dedicated to all young foreign language learners
ABSTRACT
The research reported in this thesis took place in Thessaloniki during the academic years 2010-11 and 2011-12, in two primary schools. The mainstream school introduces English as a foreign language from Grade 3, while the experimental one follows an intensive programme from Grade 1 that aims to develop children’s oracy in the L2.

In an attempt to explain individual differences in the acquisition of L2 English vocabulary, the study mainly explores whether the early introduction of English as a foreign language affects positively the cognitive functioning of young Greek learners. The main cognitive variables investigated in the study are Foreign Language Aptitude (Alexiou, 2005), Phonological Short-term Memory and the central executive of Working Memory (Baddeley & Hitch, 1974). To my knowledge, this has never before been attempted, as Early Foreign Language Learning research so far has primarily investigated the linguistic and affective outcome of this enterprise (García Lecumberri & Gallardo 2003; Mihaljevic Djigunovic & Krevelj 2009; Mihaljevic Djigunovic & Lopriore 2010; Muñoz 2006, 2010; Nikolov 2009). The main hypothesis of the thesis is that the learners’ early as well as intensive exposure to L2 English will beneficially affect their cognitive skills.

The research tools that were used are all well-established in developmental literature. The non-verbal skills of informants were assessed by the Young Learners’ Aptitude Test (YLAT) (Alexiou, 2005) and the Stop-signal test (Logan & Cowan, 1984). The verbal tests examined participants’ Phonological Short-term Memory via the Digit recall_Forward task (Wechsler, 1991), the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996), and the Test of Nonword Repetition for Greek-speaking children (Maridaki-Kassotaki, 1998). The L1 verbal intelligence of participants was also tested via the sub-tests of the two versions of the Diagnostic Test of Verbal Intelligence (DVIQ I & II) (Stavrakaki & Tsimpi, 2000) that pertain to vocabulary and metalinguistic knowledge. The central executive of Working
Memory was explored twice. Participants took the Digit recall_Backward task (Wechsler, 1991) and the Listening span and Recall task (Daneman & Carpenter, 1980). Finally, a FL vocabulary test was designed by the researcher, which tested both the receptive and the productive skills of the experimental group. This was based on the material that was covered in the two years of the group’s FL schooling.

By means of these research tools, the study attempts to answer the four research questions. The first one examines whether the early introduction of L2 English enhances any of the aptitude skills that are associated with young learners (Alexiou, 2005). The second research question wishes to find out whether at this early stage of foreign language learning there is another factor apart from Phonological Short-term Memory that can strongly predict L2 vocabulary performance. The third research question examines whether the experimental group, after the two-year intensive L2 programme followed, will display a firmer control than the control group, over their response inhibition mechanism. Finally, the last question of this study explores whether Early Foreign Language Learning can be associated with a gender effect, regarding the above-mentioned cognitive skills.

Overall, the results are more than promising, as they strongly suggest that Early Foreign Language Learning can enhance certain cognitive skills in young learners as well as their complex working memory. This may have a positive impact on their language learning skills, be these of their native language or of a foreign one.

**Keywords:** Early Foreign Language Learning, Foreign Language Aptitude, Working Memory, Response Inhibition, FL vocabulary acquisition.
ACKNOWLEDGEMENTS

First and foremost, I am deeply indebted to my supervisor, Marina Matheoudaki, for her insightful and fruitful comments and continuous encouragement throughout this academic endeavour. I would also like to thank the other two members of my supervising committee, Thomai Alexiou and Areti Okalidou, for their support and suggestions that contributed immensely to the making of this thesis.

I also thank the Ministry of Education, Lifelong Learning, and Religious Affairs, in particular the Pedagogical Institute of Athens, for granting me permission to visit the two schools to collect my data.

The present thesis would not have been possible without the substantial contribution of several eminent colleagues and friends, who provided me with their help and feedback. At the risk of leaving out several who deserve mention, I wish to thank in particular Mariette Huizinga for offering me guidance and advice on the technicalities of the stop-signal task. I feel the need to thank Katerina Nikolaidis for letting me use the facilities of the Phonetics lab to record the two pseudoword tests as well as for her comments on the first presentation of my work to my supervising committee. Also, I would like to express my gratitude to Ed Joycey, for helping me with the recordings and the scoring of the English nonword repetition test. Sincere thanks go to Nikos Amvrazis, Georgia Fotiadou, Popi Katsika, Despina Papadopoulou, Eleni Peristeri, and Costas Tzanis, for enlightening me on issues relating to SPSS. My appreciation also goes to Tasos Paschalidis for his technical assistance and to Fotini Stavrou for her help in tracing bibliographical resources. Iantiki Maria Tsimpli deserves special recognition for being an important influence on my academic thinking. Iantiki provided me with the encouragement, as well as the critical commentary, required to generate and present my own ideas. Finally, I would like to express my thanks to those scholars whose work has stimulated my own thinking during this research.
I am truly indebted to all the children of the 3\textsuperscript{rd} Model Experimental school and the 2\textsuperscript{nd} primary school in Evosmos for the enthusiasm they exhibited during the two years this research lasted. If it wasn’t for them, I couldn’t have possibly gathered the data of the study. But first, I wish to thank their parents for consenting to their children’s participation in this research. Last, I would like to express my gratitude to the school directors and teachers, for their assistance, understanding and excellent cooperation.

This thesis proved to be a fascinating journey to self-discovery as well as to academic knowledge. I would not have succeeded, however, in achieving anything really without the immense help, love, and psychological support I received from my husband Mitsos, my two daughters Evi and Mary, my father, Stathis Efstathiadis, and from the closest of my friends. I thank them all so very dearly.

Parts of this thesis were presented in two 2013 conferences and will be published in their proceedings. First, in the 21\textsuperscript{st} International Symposium of Theoretical and Applied Linguistics that was held in Thessaloniki (5-7 April), which was organised by the Department of Theoretical and Applied Linguistics of the School of English, Aristotle University of Thessaloniki. Second, in the 7th Postgraduate Conference, that was held in Athens (16-18 May), organised by the Faculty of Philology of the National and Kapodistrian University of Athens. I thank the audiences of both for their fruitful comments and the particular interest they demonstrated in my research. Needless to say that any errors still found in this thesis are my own.

It is my deeper hope that this study will contribute in some small way to the understanding of the cognitive processes that are involved in the FLL process, in particular in the early FL vocabulary acquisition by young learners.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xi</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>xv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xvii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xviii</td>
</tr>
</tbody>
</table>

PART I: THE THEORETICAL BACKGROUND OF THE STUDY

CHAPTER 1: INTRODUCTION

1.1 Learner differences: A brief introduction                               1
1.2 The rationale of the study                                              1
1.3 The aim of the present thesis                                           2
1.4 The research tools used                                                4
1.5 The four research questions                                             5
1.6 The outline of the thesis                                              5

CHAPTER 2: SECOND LANGUAGE LEARNING THEORIES

2.1 Some background information                                            7
2.2 Differences between L1 acquisition and SLA/FLL                         9
2.3 SLA Theories                                                          12
   2.3.1 Behaviourism                                                      12
   2.3.2 Innatism                                                         13
   2.3.3 The Interaction Approach                                           22
   2.3.4 Other SLA approaches                                              25
2.4 Chapter Summary                                                       26

CHAPTER 3: THEORIES OF CHILD COGNITIVE DEVELOPMENT

3.1 Introduction                                                          28
3.2 Theories of child cognitive development                               28
   3.2.1 Constructivism                                                    29
5.5 A reconceptualisation of FL aptitude: The post-Carrollian era .................................. 105
5.6 FL aptitude and SLA achievement .................................................................................. 111
5.7 Chapter Summary ............................................................................................................. 113

CHAPTER 6: EXECUTIVE FUNCTIONS: WORKING MEMORY, RESPONSE INHIBITION AND THEIR CONTRIBUTION TO FL SUCCESS

6.1 Introduction .................................................................................................................... 114
6.2 Executive Functions (EFs) ............................................................................................. 114
6.3 Defining Working Memory (WM) .................................................................................... 119
  6.3.1 WM and its involvement in the LL process and FL success .................................. 123
6.4 Defining Response or Motor Inhibition ........................................................................ 128
6.5 Chapter Summary .......................................................................................................... 131

PART II: RESEARCH METHODOLOGY AND THE FINDINGS OF THE STUDY

CHAPTER 7: THE PRESENT STUDY, RESEARCH METHODOLOGY & DESIGN

7.1 The study and its goal ....................................................................................................... 133
7.2 A brief account of the theoretical framework of the study ........................................... 133
7.3 The research questions .................................................................................................... 135
7.4 The two schools .............................................................................................................. 140
7.5 Recruitment of participants ........................................................................................... 141
7.6 The testing procedure ...................................................................................................... 142
7.7 The tools used and the skills tested ................................................................................ 144
  7.7.1 Non-verbal tasks: Measures of cognitive processing .............................................. 144
    7.7.1.1 The Young Learners’ Aptitude Test (YLAT) ..................................................... 144
    7.7.1.2 The stop-signal task ......................................................................................... 148
  7.7.2 Verbal tasks and WM assessments ......................................................................... 152
    7.7.2.1 Measures of PSTM: The forward digit span and recall test ......................... 152
    7.7.2.2 Measures of PSTM: The two nonword repetition tests .................................. 153
    7.7.2.3 Measures of CWM ......................................................................................... 155
      7.7.2.3.1 The backwards digit span and recall test .................................................. 155
      7.7.2.3.2 The Listening span and Recall task .......................................................... 156
    7.7.2.4 Measures of L1 verbal intelligence ................................................................. 158
    7.7.2.5 The English vocabulary test .......................................................................... 160
CHAPTER 8: DATA ANALYSIS

8.1 The testing waves in the two schools ............................................................. 164
8.2 Elimination of outliers .................................................................................... 165
8.3 Descriptive Statistics ....................................................................................... 165
8.4 Research Methodology ..................................................................................... 168
8.5 The independent samples t-tests ..................................................................... 170
  8.5.1 Grade 1 ........................................................................................................ 170
  8.5.2 Grade 2 ........................................................................................................ 173
  8.5.3 The performance of the two groups across the two grades ....................... 175
  8.5.4 Experimental group: Paired samples t-tests ................................................ 176
8.6 Correlation matrices: Grades 1 and 2 ............................................................. 179
  8.6.1 Experimental data: Grade 1 ........................................................................ 181
  8.6.2 Control data: Grade 1 ................................................................................ 184
  8.6.3 Experimental data: Grade 2 ........................................................................ 187
  8.6.4 Control data: Grade 2 ................................................................................ 191
8.7 Regression analyses .......................................................................................... 192
  8.7.1 Grade 1 ........................................................................................................ 192
  8.7.2 Grade 2 ........................................................................................................ 194
    8.7.2.1 FL_Voc_Total as the dependent variable .................................................... 194
    8.7.2.2 FL_Receptive as the dependent variable .................................................... 195
    8.7.2.3 FL_Productive as the dependent variable .................................................. 196
    8.7.2.4 L1 verbal intelligence as the dependent variable ..................................... 197
    8.7.2.5 FL aptitude as the dependent variable ..................................................... 198
    8.7.2.6 PSTM as the dependent variable .............................................................. 199
8.8 Two factor analyses of variance (ANOВAs) ...................................................... 200
8.9 Multivariate analyses of variance (MANВAs) .................................................... 202
8.10 The Listening span and Recall task results ..................................................... 202
  8.10.1 The experimental data ................................................................................ 202
  8.10.2 The control data ......................................................................................... 203
    8.10.3 Correct (XX+OX) vs incorrect word recall (XO+OO) (%) per group, grade, and level ........................................................................................................... 204
8.11 The English Vocabulary Test .......................................................................... 206
CHAPTER 10: CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH

10.1 The significance of the present study ......................................................... 270
10.2 An early start is compatible with all FLL and child cognitive theories .......... 277
10.3 Future directions of research ................................................................. 280

REFERENCES ................................................................................................. 283
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Period</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>the 40s and 50s</td>
<td>1941-1960</td>
</tr>
<tr>
<td>the 60s</td>
<td>1961-1970</td>
</tr>
<tr>
<td>the 70s</td>
<td>1971-1980</td>
</tr>
<tr>
<td>the 90s</td>
<td>1991-2000</td>
</tr>
</tbody>
</table>

ADHD: Attention Deficit Hyperactivity Disorder
ALAT: Army Language Aptitude Test
ANOVA(s): Analysis(-es) of variance
BAF: Barcelona Age Factor
BICS: Basic Interpersonal Communications Skills
C: Classification
CAH: Contrastive Analysis Hypothesis
CALP: Cognitive Academic Language Proficiency
Canal-F: Cognitive Ability for Novelty in Acquisition of Language (Foreign)
CI: Confidence interval
CLIL: Content and Language Integrated Learning
CNRep: Children’s Test of Nonword Repetition
CPH: Critical Period Hypothesis
CPS: Concurrent processing-storage
CREED: Construction-based, Rational, Exemplar-driven, Emergent, and Dialectic
CUCB: Common Underlying Conceptual Base
CWM: Complex Working Memory
D: Differences
df: degree of freedom
DLAB: Defense Language Aptitude Battery
DLAT: Defense Language Aptitude Test
DRB: Digit recall_Backwards
DLI: Defense Language Institute
DVIQ: Diagnostic Test of Verbal Intelligence
EC: European Commission
EF(s): Executive Function(s)
EFL: English as a Foreign Language
EFLL: Early Foreign Language Learning
ELL: Early Language Learning
ELLiE: Early Language Learning in Europe
ELT: English Language Teaching
EU: European Union
FLLD: Foreign Language Learning Disability
FL(s): Foreign Language(s)
FLL: Foreign language learning
FoF: Focus on Form
Go-RT(s): RT(s) to the Go signal
GLL: Green: left presented-left pressed
GLR: Green: left presented-right pressed
GRL: Green: right presented-left pressed
HD: High Definition
IDs: Individual differences
IL: Interlanguage
IP: Information-processing
IQ: Intelligence Quotient
ITM: Intermediate-term memory
L1: Mother tongue/1st language
L2: Second language
L3: Third language
LAD: Language Acquisition Device
LASS: Language Acquisition Support System
LAT: Language Aptitude Test
LCDH: Linguistic Coding Differences Hypothesis
LL: Language Learning
LLAMA: Llama Language Aptitude Test
LR: Listening Recall
LTM: Long-term memory
M capacity: Mental capacity
MANOVA(s): Multivariate analysis(-es) of variance
MENYET: Magyar Egyetemes Nyelvérzékmérő Teszt (Hungarian General Aptitude Test)
Met.: Metalinguistic
MLAT: Modern Language Aptitude Test
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLPS</td>
<td>Modern Languages at Primary School</td>
</tr>
<tr>
<td>M-Q3 range</td>
<td>between the median and lower limit of the upper quartile</td>
</tr>
<tr>
<td>ms</td>
<td>milli-seconds</td>
</tr>
<tr>
<td>n</td>
<td>Number</td>
</tr>
<tr>
<td>NRGreek</td>
<td>Test of Nonword Repetition for Greek-speaking children</td>
</tr>
<tr>
<td>NNS(s)</td>
<td>Non-native speaker(s)</td>
</tr>
<tr>
<td>NS(s)</td>
<td>Native speaker(s)</td>
</tr>
<tr>
<td>ns</td>
<td>non-significant</td>
</tr>
<tr>
<td>PA</td>
<td>Paired Associates</td>
</tr>
<tr>
<td>PIN</td>
<td>Personal Identification Number</td>
</tr>
<tr>
<td>PLAB</td>
<td>Pimsleur’s Language Aptitude Battery</td>
</tr>
<tr>
<td>PM</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>PSTM</td>
<td>Phonological short-term memory</td>
</tr>
<tr>
<td>RT(s)</td>
<td>Reaction time(s)</td>
</tr>
<tr>
<td>SCT</td>
<td>Sociocultural Theory</td>
</tr>
<tr>
<td>SD(s)</td>
<td>Standard deviation(s)</td>
</tr>
<tr>
<td>SI</td>
<td>Semantic Integration</td>
</tr>
<tr>
<td>sig.:</td>
<td>significance</td>
</tr>
<tr>
<td>SL</td>
<td>Second Language</td>
</tr>
<tr>
<td>SLA</td>
<td>Second Language Acquisition</td>
</tr>
<tr>
<td>SLL</td>
<td>Second Language Learning</td>
</tr>
<tr>
<td>SLLs</td>
<td>Second language learners</td>
</tr>
<tr>
<td>SOA(s)</td>
<td>Stimulus-onset asynchrony(-ies)</td>
</tr>
<tr>
<td>STM</td>
<td>Short-term memory</td>
</tr>
<tr>
<td>SS</td>
<td>Story Sequences</td>
</tr>
<tr>
<td>SSRT(s)</td>
<td>Stop-signal reaction time(s)</td>
</tr>
<tr>
<td>TALD</td>
<td>Theoretical and Applied Linguistics Department</td>
</tr>
<tr>
<td>TFT</td>
<td>Thin-film-transistor</td>
</tr>
<tr>
<td>TL(s)</td>
<td>Target language(s)</td>
</tr>
<tr>
<td>ToM</td>
<td>Theory of Mind</td>
</tr>
<tr>
<td>TPR</td>
<td>Total Physical Response</td>
</tr>
<tr>
<td>UG</td>
<td>Universal Grammar</td>
</tr>
<tr>
<td>Voc.:</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>WM</td>
<td>Working Memory</td>
</tr>
</tbody>
</table>
YLAT: Young Learners’ Aptitude Test
ZPD(s): Zone(s) of proximal development
# LIST OF APPENDICES

## CHAPTER 7

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WISC-III: Digit span task</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>The children’s test of nonword repetition (Gathercole &amp; Baddeley, 1996)</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>The test of nonword repetition for Greek-speaking children (Μαριδάκη-Κασσωτάκη, 1998)</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>The Listening Span and Recall Test</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>DVIQ I_Voc.: Vocabulary (Productive skills tested)</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>DVIQ I_Met.: Metalinguistic concepts (Receptive skills tested)</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>DVIQ II_Met.: Metalinguistic concepts (Receptive skills tested)</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>FL Vocabulary Test (Receptive)</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>FL Vocabulary Test (Productive)</td>
<td>25</td>
</tr>
</tbody>
</table>

## CHAPTER 8

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Levene’s test for equality of variances Experimental group, Grades 1 &amp; 2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Control group: Differences attested in the measures between Grades 1 &amp; 2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Correlation matrices (per group &amp; grade)</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Digit recall_Backwards: Two-way ANOVA</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>Digit recall_Backwards Independent samples t-test: Grades 1 &amp; 2</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Paired Associates: Two-way ANOVA</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Paired Associates Independent samples t-test: Grades 1 &amp; 2</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Digit recall_Forward, Grade 1 - Two-way ANOVA</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Multivariate Analyses of Variance: Grade 2</td>
<td>14</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1:</td>
<td>The current multi-component model of WM</td>
<td>122</td>
</tr>
<tr>
<td>Figure 2:</td>
<td>Experimental group, Grades 1 and 2 (no. of students per grade &amp; level)</td>
<td>203</td>
</tr>
<tr>
<td>Figure 3:</td>
<td>Control group, Grades 1 and 2 (no. of students per grade &amp; level)</td>
<td>204</td>
</tr>
<tr>
<td>Figure 4:</td>
<td>FL_Receptive: Time duration and score</td>
<td>207</td>
</tr>
<tr>
<td>Figure 5:</td>
<td>FL_Productive: Time duration and score</td>
<td>208</td>
</tr>
<tr>
<td>Figure 6:</td>
<td>The overall Go-RT\textsubscript{mean} values (per group &amp; grade)</td>
<td>212</td>
</tr>
<tr>
<td>Figure 7:</td>
<td>The overall SSRT\textsubscript{mean} values (per group &amp; grade)</td>
<td>213</td>
</tr>
<tr>
<td>Figure 8:</td>
<td>Go-RT\textsubscript{mean} across the two grades (with respect to group)</td>
<td>227</td>
</tr>
<tr>
<td>Figure 9:</td>
<td>SSRT\textsubscript{mean} across the two grades (with respect to group)</td>
<td>227</td>
</tr>
<tr>
<td>Figure 10:</td>
<td>Error rates on Go- and Stop-signal task trials (per group &amp; grade)</td>
<td>228</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1:  Compulsory start age for Foreign Language Learning:
The European context ................................................................. 75
Table 2:  Skehan’s proposal of SLA stages and aptitude constructs .............. 110
Table 3:  The testing waves in both schools ........................................ 164
Table 4:  Descriptive Statistics - Means and Standard Deviations (SDs), Grade 1 ...... 166
Table 5:  Descriptive Statistics - Means and Standard Deviations (SDs), Grade 2 ...... 167
Table 6:  Independent samples t-tests, Grade 1 .................................... 171
Table 7:  Independent samples t-tests, Grade 2 .................................... 173
Table 8:  Differences attested in the independent samples t-tests, in favour of the experimental group .................................................. 176
Table 9:  Experimental group, differences (%) found in the means of all tests between Grade 1 and Grade 2 ................................................ 177
Table 10: Experimental group (n=47): Correlations between the measures, Grade 1 ................................................................. 180
Table 11: Control group (n=49): Correlations between the measures, Grade 1 ................................................................. 183
Table 12: Experimental group (n=47): Correlations between the measures, Grade 2 ................................................................. 186
Table 13: Control group (n=49): Correlations between the measures, Grade 2 ................................................................. 190
Table 14: Correct (XX+OX) vs Incorrect (XO+OO) word recall:
response types (%) and SDs .......................................................... 205
Table 15: The combined performance of the experimental group in the two FL tests 209
Table 16: Mean and SDs of Go_RT and SSRT (per group & grade) .................. 212
Table 17: Experimental group (n=46): Correlations between the totality of the measures, Grade 1 ................................................................. 214
Table 18: Control group (n=49): Correlations between the totality of the measures, Grade 1

Table 19: Experimental group (n=46): Correlations between the totality of the measures, Grade 2

Table 20: Control group (n=49): Correlations between the totality of the measures, Grade 2

Table 21: Descriptive statistics for SSRT\textsubscript{mean} (per group & grade)

Table 22: t\textsubscript{-}test results for SSRT\textsubscript{mean} (per group & grade)

Table 23: Go-RT\textsubscript{mean}, SSRT\textsubscript{mean} and their error rates (%) (Grade 1, across the groups), *CI (=Confidence Interval)

Table 24: Go-RT\textsubscript{mean}, SSRT\textsubscript{mean} and their error rates (%) (Grade 2, across the groups), *CI (=Confidence Interval)
PART I: THE THEORETICAL BACKGROUND OF THE STUDY
1.1 Learner differences: A brief introduction

Nowadays, abundant is the SLA research that focuses on the numerous ways individuals differ and on the cognitive, affective, and social variables that are at work during the FL learning process (Ehrman, Leaver, & Oxford, 2003; Lightbown & Spada, 1999; Skehan, 1989). This indirectly demonstrates the complexity of the process. Ellis (2003) counts a number of learner interrelated differences that are generally acknowledged, such as learners’ beliefs about language learning (LL), their affective states (e.g. anxiety) during the L2 learning process and general factors that influence LL such as age, language aptitude, motivation, cognitive styles, learning strategies, and lately, the capacity of Working Memory (WM) (Baddeley & Hitch, 1974).

1.2 The rationale of the study

Since 2004 the Greek primary public schools introduced the teaching of English at Grade 3 to thus align with the recommendation of the European Commission Action Plan of 2004-06: ‘member states should move towards ensuring that foreign language learning at primary school is effective’ (Commission of the European Communities, 2003, as cited in Enever, 2009: 16).

Since 2005 the Theoretical and Applied Linguistics Department has had under its custody the 3rd Model Experimental School in Evosmos. The department supervises the school’s operations with respect to the teaching of English from Grade 1 and is responsible for issues that relate to curriculum design, staff selection, teaching methods and materials. Students are strongly discouraged to attend any private English lessons, because of the intensity of the FL programme followed. English is taught for 5 hours per week in Grades 1 and 2 and for 8 hours per week from Grade 3 to Grade 6. From 2010-11, Content and
Language Integrated Learning (CLIL) is also practised in History (Grade 3), Environmental Study (Grade 4), Arts and Crafts (Grade 4), Science (Grade 6), Geography (Grades 5-6), Religious Education (Grades 5-6), via the medium of English.

As of September 2010, the Ministry of Education and Religious Affairs is running a pilot program in 800 primary schools throughout Greece, where English is taught for 2 hours per week from Grade 1. Even so, the case of the 3rd Model Experimental School is a unique one all over Greece, because of the intensity of its programme. For this reason, it was considered imperative to conduct a research that would measure for the first time the outcome of this experimental initiative. This is so, because the Ministry of Education is expected to gradually introduce the FL subject (from Grade 1) to the totality of the Greek primary state schools, to fully align with the European recommendation mentioned above.

Presently, the teaching of FLs to young learners is gaining popularity all over the world. The age for learning a FL has been reduced in many countries, driven by the Critical Period Hypothesis’ main assumption, i.e. that learning a FL is at its highest peak around the ages of 6-9 and by ‘the younger the better in the long run’ view of Singleton (1995), who suggested that the sooner a child is exposed to the L2, the better the results in the long run, provided that exposure to the FL is substantial.

The study was conducted in Greece, Thessaloniki, during the academic years of 2010-11 and 2011-12, in two primary schools: a mainstream school (hereafter, the control group) which introduces English as a FL from Grade 3, and the 3rd Model Experimental School (hereafter, the experimental group) mentioned earlier.

1.3 The aim of the present thesis

The main goal of this thesis is to explore whether the early introduction of the FL in the school curriculum, can be directly linked to a more efficient cognitive functioning of young FL learners.
CHAPTER 1: INTRODUCTION

To this end, the study concentrates on two cognitive constructs, namely FL aptitude and WM. Previous work on young learners’ FL aptitude (Alexiou, 2005), demonstrates a close link between this construct and FL performance. Recently, WM, i.e. the mental space where numerous executive processes take place, has been identified as a key component of language aptitude (Chan, Skehan, & Gong, 2011; Dörnyei, 2005; Robinson, 2005). Also, it is well-established that the phonological store of WM plays a central role in L1 (Adams & Gathercole, 1996; Baddeley, 1986; Gathercole & Adams, 1994; Gathercole & Baddeley, 1993), child and adult L2 vocabulary development (Baddeley, Gathercole, & Papagno, 1998; Cheung, 1996; Gathercole & Adams, 1994; Gathercole & Baddeley, 1993; Masoura & Gathercole, 1999; Μασούρα, Gathercole, & Μπαμπλέκου, 2004; Service, 1992).

In addition, this thesis will further explore the existence of a relation between FL vocabulary performance and the central executive of WM. According to the WM model of Baddeley and Hitch (1974), the central executive is responsible for numerous high-level regulatory executive processes, such as the abilities to direct attention, plan actions, solve problems, etc. (Baddeley, 1986; Kimberg, D'Esposito, & Farah, 1997). The allocation of attention, in particular, is inextricably linked to the FL process (Randall, 2007), because of the explicit and conscious nature of the process, which demands a lot of effort on the part of the learner before the new FL routines become automatic.

Apart from WM, the study will additionally examine the second of the three most frequently discussed Executive Functions (EFs), that of Inhibition (for a review see Sawyer & Ranta, 2001). Inhibition has been defined as the ability to resist interference and control the suppression of responses, when the context changes and these are no longer considered appropriate or relevant to the demands of a task (Logan & Cowan, 1984; Miyake et al., 2000). WM and Inhibition were selected for further investigation, on the grounds of their emerging early in one’s life (Beveridge, Jarrold, & Pettit, 2002; Wiebe et al., 2011) and of their sharing
a close association with children’s academic progress during their school years (see Chapter 6). Given the limited storage capacity of WM, a firm control of one’s Inhibition mechanism, will filter out all irrelevant information from intruding into WM, contributing thus in its more efficient functioning, which, in turn, will facilitate the FL learning process.

To the best of my knowledge, all these cognitive variables have never before been examined together, in relation to FL vocabulary performance in young learners. The majority of the EFLL studies have investigated the linguistic or affective outcome of this endeavour (García Lecumberri & Gallardo 2003; Mihaljevic Djigunovic & Krevelj 2009; Mihaljevic Djigunovic & Lopriore 2010; Muñoz 2006, 2010; Nikolov 2009).

1.4 The research tools used

The research tools that were used, non-verbal as well as verbal, are all well-established in developmental literature. The non-verbal tests of the study are the Young Learners’ Aptitude Test (YLAT) (Alexiou, 2005), that assesses the FL aptitude of learners and the Stop-Signal test (Logan & Cowan, 1984), which examines their Response Inhibition mechanism. Regarding the verbal tests of the study, these tap participants’ Phonological Short-term Memory (PSTM) via the Digit recall_Foward task (Wechsler, 1991), the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996), and the Test of Nonword Repetition for Greek-speaking children (Maridaki-Kassotaki, 1998). The L1 verbal intelligence of informants was also tested, via the sub-tests of the Diagnostic Test of Verbal Intelligence I & II (DVIQ I & II) (Stavrakaki & Tsimpli, 2000), which assess learners’ vocabulary skills (receptive, productive) and metalinguistic knowledge. The central executive of WM was tested twice, by means of the Digit recall_Backward task (Wechsler, 1991) and the Listening span and Recall task (Daneman & Carpenter, 1980). Finally, a FL vocabulary test was designed by this researcher, which was based on the material and the thematic units that were covered during the two years of the experimental group’s FL schooling.
1.5 The four research questions

Through the administration of the above-mentioned research tools, the study sets out to answer four research questions. The first research question examines whether the early introduction of EFL enhances any of the aptitude skills that are associated with young learners (Alexiou, 2005). The second research question explores whether, apart from PSTM, there is another factor that can strongly predict the FL vocabulary performance at this early stage of FL learning (FLL). The third research question seeks to find out whether the experimental group, after the two-year intensive L2 programme, will outperform the control group, with regard to a better functioning of their response inhibition mechanism. Finally, the last research question explores whether EFLL can be associated with a gender effect, regarding the cognitive skills examined in the study.

Overall, the results are encouraging as they suggest that EFLL can enhance certain cognitive skills in young learners as well as their complex working memory. This is truly important as cognitive skills are not language-specific but constitute part of one’s common pool of cognitive resources. The speculation of this study is that once these skills are positively affected, even boosted, by the FL experience, their enhancement will activate a ‘domino’ effect on the learners’ overall language learning process, be this the mother-tongue or the FL.

1.6 The outline of the thesis

The thesis is divided into two parts. Part I is comprised of six chapters that establish the theoretical framework of the research. More specifically, Chapter 1, the present chapter, constitutes the introductory part of this thesis. It briefly informs the reader of the goals of the study, displays the main research questions and provides an outline of the thesis. Chapter 2 discusses the most prominent Second Language Learning theories, while Chapter 3 is concerned with displaying the prevailing theories of child cognitive development. Chapter 4
discusses one of the IDs that differentiate FL learners, namely Age, and fully portrays the European picture with regard to EFLL. Chapter 5 relates to the cognitive construct of FL aptitude, while Chapter 6 gives a detailed description of the two EFs that are investigated in the study, i.e. Working Memory and Inhibition.

Part II consists of four chapters. Chapter 7 discusses the goal of the study and presents in detail the four research questions along with the initially formulated hypotheses, which were formed on the basis of the literature review that was conducted in Part I. It also gives information as to the schools that took part in the study and a full description of the research methodology and design. Chapter 8 is a comprehensive report of the study findings. Chapter 9 revisits the four research questions, checks whether the initial hypotheses were borne out or not, and attempts to interpret these in relation to their pedagogical value. Finally, Chapter 10 presents some concluding remarks, explains the contribution of the thesis, and finds that EFLL is compatible with all FL learning and child cognitive theories, and offers directions for future research.

In the enclosed CD, the reader will find two appendices. The appendix to Chapter 7 is a presentation of the research tools, while in the appendix to Chapter 8 the reader will find adequate information and extra details on the statistical tests that were performed for the purposes of this thesis.
CHAPTER 2
SECOND LANGUAGE LEARNING THEORIES

2.1 Some background information

Language acquisition is one of the most impressive and fascinating aspects of human development. What is perhaps remarkable in the LL process is that there is a high degree of similarity with which children all over the world learn their mother tongue (L1), a second language (SL), or a foreign language (FL). This has been found to hold true with respect to the sequence of developmental stages in the acquisition of certain syntactical features (Dulay & Burt, 1974). Research findings suggest that L2 learners, just like L1 acquirers, carry an ‘internal syllabus’ (Johnson, 2001: 71).

This thesis focuses on EFL and, in particular, on very young learners (6-8 years of age) who learn a FL in a formal context during the first two years of primary schooling. The terms Foreign Language Learning (FLL) and Second Language Acquisition (SLA) will be used interchangeably for reasons of convenience, although they differ in nature. Unlike the case of SLA where the SL is also the language of the community, the learning of the additional language (the FL) takes place within an instructed setting, the school, outside which one has no immediate need to use the language in order to satisfy one’s basic communicative needs.

Since the end of the 60s, SLA research has developed into a wide-reaching field of scientific enquiry that attempts to decipher the mystery of L2 learning. It is interdisciplinary in nature, with an orientation which ranges from the structure of language, to psycholinguistic issues, to pragmatic aspects that also take into account the social factors that may influence language development (Ortega, 2007). Its inherent diversity is what makes research in this field a fascinating enterprise.
SLA is not a uniform and predictable process but, rather, a complex one with a wide variation attested among learners. This is because it involves many interrelated factors. Depending on the perspective adopted, numerous theories have attempted to explain SLA. This may explain the lack of consensus that exists among researchers and the variety of contradictory positions they have held. For example, SLA has been seen as: a) the business of the individual or an internal cognitive process that is, in part, specified by the learner’s innate characteristics, b) a psycholinguistic process, “which ultimately resides in the mind-brain, where also lie its secrets” (Long & Doughty, 2003: 866), c) a process that takes place in a social setting which is extremely sensitive to environmental variation, i.e. the setting and the interlocutors (Ellis, 1985).

Even casual observations in one’s daily life reveal that people learn FLs differently, with some learning a FL in an easier, faster or better way than others. From the late 60s (for a review see Sparks & Ganschow, 2001), an extensive research on individual differences (IDs) and affective variables has been conducted, because IDs were seen as “the most consistent predictors of L2 learning success” (Dörnyei, 2005: 2). Individual differences such as age, aptitude, intelligence, attitude, motivation, LL strategies, cognitive styles or personality traits were found to be responsible for and influence the route along which learners pass, as well as the rate and their eventual success (Clément & Gardner, 2001; Dörnyei, 1998; for a review see Gardner, 1990; Gardner & Lambert, 1972; O’Malley & Chamot, 1990; Oxford, 1990, as cited in Dörneyi & Skehan, 2003: 608; Oxford & Anderson, 1995, as cited in Dörneyi & Skehan, 2003: 605; Skehan, 1989, 1998).

What is perhaps of utmost importance is the pedagogic value of SLA research findings. Its outcomes touch upon issues such as language teaching and child language acquisition. They have the potential to trigger further action and provide the inspiration to those who take pedagogic decisions, launch new (inter)national language policies, and design new,
innovative textbooks. The people involved in this field, namely politicians, policy designers, materials and textbook writers, school teachers and principals, all need to serve one common purpose: to facilitate the process of FLL.

2.2 Differences between L1 acquisition and SLA/FLL

Research has shown that there are important differences as well as similarities attested in the ways children acquire their L1 and learners (children or adults) learn an L2 (an additional language) in a formal instructional context. Child-adult differences have been addressed in the research literature with regard to rate, level of ultimate attainment (see the section that discusses the Critical Period Hypothesis (CPH)) and processes of acquisition (e.g. nativist accounts, work on developmental sequences and language transfer). These issues have led to research comparing L1 acquisition with child L2 acquisition, and child SLA with adult SLA (Meisel, 2008; Unsworth, 2008). The following are some of the most frequently discussed findings that relate to the differences between L1 acquisition and SLA/FLL (Ellis, 2003; Krashen, 1982b; Lightbown & Spada, 1999):

- In L1 acquisition the initial and final states are the same across individuals, regardless of the language to be learned. In SLA, there is great variation attested in both the initial state (as L1 grammars vary) and the final state.

- L1 acquisition is an effortless, unconscious process that takes place implicitly, in a naturalistic environment. On the contrary, L2 learning takes place, in most cases, in a scholastic environment and LL is explicit and conscious.

- L2 learners have access to a previously acquired language or languages. This consequently means they a) have an already established knowledge of the world, b) are more cognitively mature than L1 acquirers, as they usually learn the L2 later in life, and c) have already developed a metalinguistic awareness of how language works.
There is a huge difference in the nature of the input and the feedback received in the two LL processes. In L1 acquisition children need no corrective feedback, i.e. no negative evidence, to discover the meanings of words (Bloom, 2002). Even if they are corrected, this makes little difference to their speech patterns (Harley, 2001). Caretaker speech is, in general, message- and not form-focused (Johnson, 2001). It is intelligible, simplified and more tied to the child’s reality. It contains a lot of repetition and is roughly tuned to satisfy the current needs of the L1 child. Consequently, the vocabulary used is rather restricted, bound by the child’s everyday routines, while the sentences are quite short (Harley, 2001). Usually, the exact opposite happens in L2 instructional contexts. The teacher or more advanced peers correct learner errors implicitly or explicitly. The input is structured and carefully organised and the focus is primarily on form, sometimes at the expense of meaning. The material is usually authentic and may, at times, sound unintelligible to the ears of the novice L2 learner. As for vocabulary, this usually covers a wide range of everyday activities, which may not always be related to the L2 learner’s immediate life experiences.

In child L1 acquisition, errors are viewed as productive, transitional forms that pass unnoticed, unless they affect the overall meaning. Likewise, in adult native speaker (NS) talk errors are treated as ‘slips of the tongue’, whereas in SLA errors are undesirable forms, a deviation from the norm.

While in L1 acquisition a silent period is something natural, expected and therefore excused when it happens, this is not the case with SLA. FL learners are expected to produce language from the very beginning, a fact that may increase their anxiety. Even within L2 learners important differences do exist: a silent period is expected and thus excused in child SLA (Krashen, 1985), due to the child’s cognitive immaturity, while in adult SLA a silent period is not tolerated because cognitive maturity is at its peak.
In the case of instructed SLA, affective variables such as anxiety, attitude, and motivation play a decisive role to ultimate attainment and learner performance. In L1 acquisition these factors are simply non-existent as it is a natural process. Young children acquire their mother tongue not because they hold positive attitudes to it or because they are highly motivated and anxious-free but because they are pre-programmed to acquire language (see the nativist account below), which, in fact, distinguishes humans from the other species. In the case of instructed L2 learning, however, a lot of effort is required to learn the FL. The differences between the two learning contexts are numerous. To mention only one, there are no testing conditions in the case of L1 acquisition, whereas these are a ‘must’ in L2 instructed learning. Dörnyei (2005: 199) cites Horwitz, Horwitz and Cope’s (1986) idea of ‘foreign language anxiety’. Very often negative reactions and the feeling of worry are aroused in L2 learners, due to their inherent L2 linguistic deficit. Also, Dörnyei (2005) acknowledges that only highly motivated learners with an intense desire to learn an L2 will accomplish long-term goals with respect to the target language (TL) and cope with the tedious learning process. Motivation, along with the attitude held towards the TL, culture, and the social setting where this is spoken, may strongly affect the L2 learning outcome, for the worse or for the better.

Formulaic speech plays a critical role in FLL, much more so than in L1 acquisition (Wong Fillmore, 1976, as cited in McLaughlin, 1984: 168). This is because learners need to rely on prefabricated chunks in everyday life, especially in the very beginning of the learning process. These ready-made chunks give them a good starting point when their linguistic resources are still naïve or poor. As a consequence, L2 learners produce longer utterances from the beginning, which are semantically and structurally simplified (Ellis, 2003).

Perhaps the most significant difference of all is the time available within which L1 and L2 learning take place and the contact with NSs of the TL. In L1 acquisition the child is
surrounded by NSs who use the language, which is also the medium of communication among the speakers of the community the child lives in. In the latter case, exposure to the L2 input is rather limited as the L2 itself is one of the school subjects, taught for just a few hours a week. The TL is not used outside the school to satisfy the communicative needs of the child, whereas full- or part-time contact with proficient speakers or NSs of the TL is usually scarce or simply non-existent.

The section that follows discusses some of the most prominent theories that have been proposed to account for SLA in the classroom context.

2.3 SLA Theories

2.3.1 Behaviourism

Behaviourism was an influential psychological theory of learning of the 40s and 50s, with Bloomfield, Pavlov, Sapon and Skinner among its strongest proponents (Ellis, 1985). The theory dominated both L1 acquisition and SLA up to the 60s but studied only the ‘observable’ (Johnson, 2001: 44), failing to discuss and explain any mental or internal processes taking place during the LL process. All behavior was explained with reference to conditioning (i.e. responses to external stimuli) and subsequent reinforcement (VanPatten & Williams, 2007). Learning in general, with LL being one of its manifestations, was a question of habit formation. The theory held that children learn their L1 out of imitation of what they see, hear, or experience. LL takes place through repetition, practice and reinforcement. Positive feedback provided by the parents or caretakers upon successful L2 attempts encourages the formation of new L2 habits. The more positive the experience and the feedback received, the more likely learning will take place.

Lack of success and the commitment of errors were attributed to ‘interference’ or the negative transfer of L1 habits into L2 production (Ellis, 2003). In this respect, Behaviourism (or Empiricism) was closely linked to Contrastive Analysis Hypothesis (CAH). Behaviourists
held that the learner’s L1 is likely to influence L2 acquisition. When the two languages are similar, the transfer is positive and target-like structures are easily acquired and learning is facilitated. Otherwise, the learner faces difficulties and learning itself is impeded by negative transfer. Thus, errors, the result of negative transfer of L1 habits, were undesirable (Ellis, 1985). CAH was also considered to be an inadequate research tool because it only explained learners’ errors without having any predictive power (Johnson, 2001). Even when errors were predicted, many were never actually committed by L2 learners. Error Analysis, originally associated with the work of Corder (1967), was also viewed as insufficient as it did not give the complete picture of learner language. It only examined learner errors, failing to view the systematic nature of learners’ ‘Interlanguage’ (Selinker, 1972) which includes what they do wrong, what they do right, what they avoid doing altogether in cases of difficulty or fear of more error production.

Vygotsky (1960, as cited in Wertsch, 1985: 28), criticised Behaviourism for reducing human behavior to animal behavior, as it tried to explain the former solely in terms of a “collection of habits worked out through the method of ‘trial and error’”. Behaviourists viewed the learner as a passive medium with no significant role to play in the language acquisition process. For them the child was just a ‘tabula rasa’ onto which experience “writes its messages” (Johnson, 2001: 43).

2.3.2 Innatism

The inadequacy of Behaviourism to account for the complexity involved in LL opened a fierce debate around the 60s between the proponents of this theory and Innatists. Chomsky’s attack (1959) on Skinner’s *Verbal Behaviour* (1957) is well-known. Chomsky, one of the most prominent figures of the Innatist school of thought, strongly believed that imitation and practice alone could not possibly explain the creativity of children in their language production (Lightbown & Spada, 1999). He stressed the active contribution of the learner and
minimised the importance of imitation and reinforcement. He severely questioned Behaviourism on the grounds of the so-called ‘logical problem of language acquisition’ or ‘the poverty of the stimulus’ problem (Chomsky, 1986). This claims that children learning their L1 (and by extension language learners in general) are creative and innovative, as they can understand and produce new forms and uses of words never before heard. Very often, there is a mismatch between the degenerate input they receive which is incomplete, full of false starts and slips of the tongue, and the kind of linguistic output they produce. Children seem to ‘know by feel’ that certain structures are ungrammatical without anyone ever teaching them this. They ‘know’ that certain interpretations of sentences are just not possible in certain contexts (Ellis, 1985; VanPatten & Williams, 2007). In sum, they “come to know more than they could reasonably be expected to learn on the basis of the samples of language which they hear” (Lightbown & Spada, 1999: 15). On these grounds, Chomsky concluded that the complexity of the linguistic system children produce in such a relatively short time (usually by the age of 2), cannot be explained solely by imitation and analogy.

Innatists like Fodor (1983) held that all human beings are born with the same computational systems, called ‘modules’. These are genetically pre-structured to allow humans to make sense of the world. The specified modules in the brain are different subsets of neural networks, that process and accumulate different kinds of information which, in their turn, make up different ‘domains’, such as language, physics, music, mathematics, etc. The task of modules is to pass on information to the central executive, which is the attentional controller, and if this input is attended to and selected for further processing, it will eventually be built up in the permanent long-term memory (LTM), where all crystallised knowledge resides. Viewed from the innatist perspective, development is domain-specific and modular: maturation in one domain is independent of maturation in another. This contrasts the domain-general development proposed in Piaget’s theory (see Chapter 3).
CHAPTER 2: SECOND LANGUAGE LEARNING THEORIES

According to Karmiloff-Smith (1992, as cited in Das Gupta & Richardson, 1995: 41), a neo-Piagetian, infants are born with a fairly limited amount of innate, domain-specific predispositions that are meant to give inexperienced children a ‘flying start’ by constraining the classes of inputs their mind computes. In the light of new experience these specifications are progressively restructured to develop into better-adapted modules.

Chomsky (1976, 2011) believes that the human mind determines human behavior. The language faculty is part of the learner’s cognitive mechanism and modular in nature. It is an organ of the mind, an altogether separate faculty, a sub-system of the body like the visual, the sensorimotor, the digestive or the immune one (Chomsky, 2000, 2005). In this sense, language acquisition for Chomsky (2000: 8) “is something that happens to the child, not something that the child does”. In the same vein, Pinker (1994, as cited in Singleton & Ryan, 2004: 188) defines language as “a specific instinct” with an “identifiable seat in the brain” and perhaps a “special set of genes”.

Chomsky’s views have been highly influential in the study of L1 acquisition as well as FLL (even though he made no explicit reference to FLL as such). It was his firm conviction that research needed to go ‘beyond the observable’. On these grounds, Chomsky (1976: 2) developed the theory of Universal Grammar (UG): “What many linguists call ‘universal grammar’ may be regarded as a theory of innate mechanisms, an underlying biological matrix that provides a framework within which the growth of languages proceeds”. UG is a theory of the ‘initial state’ of the language faculty, before any linguistic experience takes place. Although it is present from the day a baby is born, the whole of UG does not manifest itself immediately (Ellis, 1985). The primary determinant of L1 acquisition is the Language Acquisition Device (LAD), ‘a good starting point’ according to Ellis (ibid: 14), the initial state common to all species with the exception of pathological cases, a ‘piece of machinery’ that “contains a kind of blueprint of how language works” (Johnson, 2001: 47).
Chomsky’s (1986) central claim is that children are biologically pre-programmed for LL. They are equipped with universal principles that hold for all natural languages and possess a special ability to discover the rules of their L1 or of an additional language. He conceived the initial state of the language faculty as a fixed network connected to a switch box. The network consists of the principles of the language while the switches are the options to be determined by experience and are set on the basis of the very limited information available to the child. Each possible human language is identified as a particular setting of the switches, i.e. the UG parameters. By means of hypothesis-testing procedures the child/learner gradually discovers the rules of the TL grammar and converts experience into knowledge of the language of exposure (Ellis, 2003). The LAD takes experience as ‘input’ and gives language as ‘output’ that is internally represented in the mind (Chomsky, 2000). As learners progress in the active construction of the L2, errors are bound to happen (Ellis, 2003). For Chomsky (2000), input plays only a limited role in the acquisition process. He holds that sufficient interaction with others (parents, caretakers, siblings, relatives, etc.) is necessary for the device to be activated within certain time limits. In this, he agrees with Lenneberg (1967) and his discussion of the biological prerequisites of language (his nativist account will again be discussed in Chapter 4, with relation to the CPH). Chomsky’s (1976) quotation that follows summarises what has been said so far:

There is a fixed, genetically determined initial state of the mind, common to the species with at most minor variation apart from pathology. The mind passes through a sequence of states under the boundary conditions set by experience, achieving finally a “steady state” at a relatively fixed age, a state which then changes only in marginal ways. (p. 3)

Following the Chomskyan line of thought with regard to L1 acquisition, VanPatten and Williams (2007) further extend his theory to talk about all kinds of LL, be this L1, L2, or additional/foreign languages, which are all internally driven by UG. They view that SLA
resembles L1 acquisition in that L2 learners acquire many grammatical structures in relatively consistent sequences that do not vary according to child, context, caregiver behavior, prior knowledge of an L1, or any other external influence. Many of their errors are similar to those made by L1 learners. L2 learners make only certain kinds of errors and not the range of errors Behaviourism predicted. Dulay and Burt’s study (1974) on morpheme acquisition indicated a striking similarity in learners’ errors and suggested that most of L2 learner errors are developmental in nature.

Although Chomsky made no specific claims about the implications of his theory for SLA, many linguists who work within the UG framework, support the full-access view, i.e. they believe that UG should be accounted for both L1 acquisition and SLA (Borer, 1996, as cited in Doughty, 2003: 257; Long & Robinson, 1998; White, 2003). They base their claims on evidence coming from L2 post-pubertal learners who may fail to achieve ‘complete’ mastery of the TL but, still, they know so much more about the language in question, despite the limited time of exposure or the insufficient input they receive. Thus, they infer that learner Interlanguage (IL) is still constrained by UG (Martohardjono & Flynn, 1995; White, 2007). On the other hand, there are others who believe that UG is no longer available in L2 learning. Bley-Vroman (1990) proposed the Fundamental Difference Hypothesis which claims that child L1 acquisition is handled by UG, whereas adult SLA relies on more general problem-solving abilities. In addition, there is a third kind, like DeKeyser (2000) for instance, who believes that what L2 learners lose is not access to UG but the capacity to learn implicitly, because this ability deteriorates when L2 learning begins after puberty. Others claim that the innate knowledge of universal linguistic principles survives; what is lost is the ability to reset the parameters of UG (Eubank & Gregg, 1999). Scholars who agree about the existence of maturational constraints in language acquisition often disagree over their scope, nature and the precise moment these are activated. Some of the questions that have been posed are the
following: Is it phonology only or phonology as well as other domains that are affected? Are these constraints at work by the age of 6, puberty or perhaps around the mid-teens? Is there a cut-off point and a one-time catastrophic loss or a gradual sequence of losses in different domains? All these questions will be addressed to in Chapter 4 that deals with the Age Factor and EFL.

Krashen (1981b, 1985), also an innatist, predicted that native-like mastery comes simply from exposure to comprehensible language samples. In the 70s and 80s, he developed the first theory to be developed specifically for SLA, the Input Hypothesis. He was very much influenced by Chomsky’s theory of LL and believed that humans already know a great deal when they come to the task of language. What they need is simply the triggering data in the input for language acquisition to take place. He believed his theory could explain why not everything taught is learned, why what is learned may not have always been taught, and also, the ways IDs among learners and learning situations relate to the variable outcome of SLA. His Input Theory consists of the following five hypotheses:

i. The Acquisition-Learning Hypothesis: Krashen (1987) held a non-interface position as he believed in the existence of two independent and distinct ways to develop ability in the L2 that never interact. The first is implicit, a subconscious ‘picking up’ of a language, a ‘feel’ for correctness called ‘acquisition’ (ibid: 10). It is identical to the process children utilise in L1 acquisition which emerges spontaneously when they engage in normal L2 interaction that is meaning- and not form-driven (VanPatten & Williams, 2007). The second is ‘learning’ and takes place when learners gain explicit knowledge about language, what Krashen (1981a: 155) called, a ‘meta awareness’ of language, such as the learning of rules and patterns through formal instruction where learning is characterised by error correction and rule isolation. Krashen (1985) believed that during the very early stages of acquisition,
language classes and the conscious learning of rules in formal contexts seem to work best when they are the primary source of comprehensible input.

ii. The Natural Order Hypothesis was first proposed by Corder (1967) to be further elaborated by Krashen. The hypothesis states that just like L1 acquisition, L2 learners acquire grammatical structures such as questions, negation, relative clauses, etc. in a fairly predictable and natural order. This is quite independent of their L1, age, background, conditions of exposure or schooling, the order in which these are taught in language classes, or the complexity of the structure to be acquired. This, he explained, is because all language acquisition is guided by the innate LAD which helps the acquirer ‘pick up’ what (s)he hears.

iii. The Comprehensible Input Hypothesis: Krashen (1985) believed that people acquire an L2 in an amazingly simple way – by understanding messages. Through considerable exposure to ‘comprehensible input’ learners improve and progress along the natural order, if they receive enough input that is one step beyond \((i+1)\) the current level of their linguistic ability \((i)\). In this way, they can understand language that contains grammatical structures not acquired yet, with the help of the context and the extra-linguistic information that includes pictures, titles, short descriptions, etc., one’s knowledge of the world and previously acquired linguistic knowledge. One of the sources of comprehensible input is the ‘undirected reading for pleasure’. Even though there may be individual variation on the surface, deep down, the ‘mental organ’ for language (Chomsky, 1975), produces one basic product, a human language. However, if the input is insufficient or of the wrong type (no \(i+1\)), then, instead of acquisition, fossilisation may take place (1985: 43). For Krashen, comprehensible input is not just a necessary condition for SLA, it is the sufficient condition: in the presence of comprehensible input SLA is an inevitable result. A very different view was advanced by Schmidt (1990), who claimed that incidental learning is
insufficient for adult SLA inside or outside the classroom. Noticing is the necessary and sufficient condition for L2 acquisition. L2 forms must first be noticed to be registered in the memory store, for any learning to occur.

iv. **The Monitor Hypothesis** explains how acquisition and learning are used in real-time. The Monitor helps learners filter their language, given they know the rules, they have sufficient time to think, and are consciously concerned with correctness, i.e. they focus on form. The acquired system is responsible for the speakers’ utterances and their fluency, i.e. the quick and spontaneous use of the language. The learned system, on the other hand, can only be used as a *Monitor* (Krashen, 1981a). It serves as the *Editor* that makes minor changes and ‘polishes’ (before any actual speaking or writing takes place) what the acquired system produces. A fundamental claim of this hypothesis is that conscious learning cannot initiate performance. Speech will emerge when the child/learner feels ready. Depending on the use of their Monitor, acquirers are distinguished into over-users, under-users and optimal users or ‘good language learners’ who supplement acquisition with learning (ibid: 171). Although monitoring generally improves accuracy levels, the gain achieved is fairly modest (Krashen, 1981a, 1985). Hulstijn and Hulstijn (1984) reported that the informants of their study needed 30% more time when they paid attention to form, which resulted in the transmission of 14% less information. In 1975 Krashen suggested that the source of the Monitor is Piaget’s *Formal Operations* (see Chapter 3), a stage that many but not all people reach at about the age of 12.

v. **The Affective Filter Hypothesis**: The affective filter is a mental block that prevents acquirers from fully utilising the comprehensible input they receive for language acquisition. The learners’ emotional state (motives, needs, attitudes) is like a filter that either lets input essential to acquisition to freely pass or blocks it. The acquirer needs to be comfortable and ‘open’ to the input in the learning environment. The affective filter is low
CHAPTER 2: SECOND LANGUAGE LEARNING THEORIES

when the acquirer is not concerned with the possibility of failure in language acquisition. In cases of unmotivated, anxious learners who also lack self-confidence or just are ‘on the defensive’, the affective filter is high. This may partly explain the variable outcome of SLA across L2 learners. When the learners’ affective filter is low and appropriate comprehensible input is presented and comprehended, acquisition is inevitable. On this, Krashen (1985) agreed with Chomsky (1975), that the LAD will function just as automatically as any other organ.

Krashen’s theory has severely been criticised for not being adequately supported by empirical research (Long, 2007), with Krashen (1985) being aware of such a possibility:

‘Proof’ of any hypothesis is difficult, and may be impossible. To prove conclusively the correctness of the Input Hypothesis would require examination of every case of successful language acquisition, making sure it was done via comprehensible input, and eliminating all possible alternative hypotheses. (p. 67)

Despite the criticism, his work was and still is influential because it implicitly supports Communicative Language Teaching, i.e., the use of language for the sake of meaningful interaction and the accomplishment of everyday tasks and not for the sake of a ‘dry’ and rote learning of the rules. Krashen claimed that supportive evidence for his Input Hypothesis could be found in caretaker speech, where there is large amount of exposure to the L1 that contains a lot of repetition and is finely tuned (semantically, syntactically, lexically or otherwise) to match the child’s needs and current linguistic capacities. In fact, Bruner (1981: 160) held that parents are very sensitive to this as “they get down to the level on which their children are operating and move ahead with them at a rate that shows remarkable sensitivity to their child’s progress”.
2.3.3 The Interaction Approach

A third approach to (F)LL places emphasis on communicative episodes. The *Interaction Approach* holds that LL is the result of the interaction between children’s innate capacities and the environment in which they develop. Pica (1998: 1) focused on the social aspect of interaction, and viewed this “as the context and process through which language can be learned”. In the same vein, Gass and Mackey (2007) discuss the role of input, interaction, and output in LL. They hold that learners need opportunities for comprehensible input, interaction, and output. During interaction, they negotiate for meaning and with the help of the feedback received they proceed smoothly along the LL path.

Hatch (1978: 404) claimed that “language learning evolves out of learning how to carry on conversations”. She (1983) studied the impact of interaction on comprehension and believed that ‘foreigner talk’, i.e. the conversational interactions between learners and NSs by means of simplified input, serves the same functions of ‘motherese’ talk, i.e. mother-child interactions: it promotes communication, establishes an affective bond between the NS and the non-native speaker (NNS), and serves as an implicit teaching mode (as cited in Ellis, 1985: 136-7).

The *Interaction Approach* is also associated with Long (1983, as cited in Johnson, 2001: 95), who mainly focused on SLA in naturalistic settings. According to Long (1983, 1996), SLA takes place through conversational interactions between learners and NSs by means of foreigner talk that is tuned to the current linguistic abilities of the first. The *Interaction Approach* is the theoretical underpinning of Long’s *Focus on Form* (FoF) idea (Long & Robinson, 1998). Long (1991) defined FoF as the interactional moves directed at raising learner awareness of forms by drawing students’ attention to linguistic elements (words, collocations, grammatical structures, pragmatic patterns, etc.), in context, as these arise incidentally in communicatively oriented lessons that primarily focus on meaning. Extensive
SLA research has shown that communicative activities with an exclusive focus on meaning do not suffice for LL and that there is also need for a FoF (Doughty & Williams, 1998; Long, 1996).

Interactionists hold that learners acquire the L2 through talking with more competent speakers of the TL. As the former are still in the process of LL development, it is only natural that their communicative attempts may go wrong at times. If such a breakdown in communication occurs, this can actually serve a facilitative purpose by giving rise to all kinds of interactional modifications such as self-repetition, paraphrases/recasts, clarification requests, comprehension and confirmation checks, which all mean to trigger comprehensible input (Gass & Mackey, 2007; Lightbown & Spada, 1999).

When a communication problem emerges the interlocutors need to ‘repair’ and ‘negotiate for meaning’. Pica (1994) defined negotiation for meaning as the interactional modifications (e.g. the rephrasing of an unclear message or the alteration of its syntax) that need to occur when a message fails to be fully perceived or comprehended by any of the interlocutors. During negotiation the input received is tailored to match the learners’ particular strengths, weaknesses and communicative needs, providing them with a language sample that aligns with their current developmental level while it gives them opportunities to attend to L2 form and to form-meaning relationships.

Long (1996) equated interaction with negotiation for meaning, because it triggers interactional adjustments by the NS or a more competent interlocutor which facilitate acquisition. It also connects input and internal learner capacities, in particular selective attention to problematic areas of linguistic knowledge. It creates opportunities for noticing the gaps, comprehending unfamiliar input which, in turn, enables the elaboration of the learners’ linguistic repertoire. Van den Branden (2008) holds that the capacity to negotiate for meaning in the L2 is already established by the age of 11-12.
During negotiation for meaning feedback also occurs. Its role is to draw, explicitly or implicitly, the learners’ attention to problematic constructs in their IL and give them additional opportunities to focus on their comprehension or production (Gass & Mackey, 2007). Ellis, Loewen, and Erlam (2006) as well as Lyster and Ranta (1997) distinguish between explicit and implicit feedback. The first directly guides the learners’ attentional resources on particular discrepancies between what they think they know about the L2 and what the TL feature in question actually looks like. Its ultimate goal is the elicitation of the correct form. On the whole, they believed that explicit corrective feedback is more effective than implicit one, usually provided by means of recasts (Sheen, 2004). Gass and Mackey (2007) argue that interactions facilitate L2 development “because it is in this context that learners receive information about the correctness and, more important, about the incorrectness of their utterances”, to continue: “overt correction or negotiation is one way of alerting a learner to the possibility of an error in his or her speech”. The two authors hold that the Interaction Approach tried to explain how interaction is linked to and contributes to learning. In doing so, they used concepts that derive from psychology such as the allocation of attention and of focus on some features only from the external stimuli, the conscious noticing of certain aspects in the input (verbal or non-verbal) or the contribution of the cognitive construct of working memory, wherein the noticed features are temporarily stored to be then sent for further processing.

Merrill Swain (1985) also supported the facilitative effect of noticing but viewed things from a different angle. She drew particular emphasis on the importance and the role comprehensible ‘pushed’ output plays to the development of certain grammatical features in learners. Her Output Hypothesis was based on the idea that understanding and producing language are two different skills: the latter can only be developed if the learner is forced to produce more adequate output, i.e. to say and write things accurately. Negotiation for
meaning also promotes more target-like production that requires a very good command of the grammatical processing of the language (Gass & Mackey, 2007). By signaling incomprehension the more competent speakers alert learners to the fact that their output is not fully comprehended. By helping language learners to solve a communication breakdown, their interlocutors ‘force’ them to integrate more TL formulations in their output. Thus, on their way towards mutual understanding, language learners are supported by their interlocutors to notice certain gaps in their IL. Gass and Mackey (ibid) note that output production also promotes automaticity, as language use undergoes a ‘routinisation’ phase, i.e. a “consistent mapping of the same input to the same pattern of activation over many trials” (McLaughlin, 1987: 134), an idea that was very much supported by the information processing (IP) approach that will be discussed in the Chapter 3.

### 2.3.4 Other SLA approaches

There are many more competing accounts and explanations of various aspects of SLA, such as Ellis’s (2007) Associative-Cognitive CREED theory, Schumann’s Acculturation Model, Andersen’s Nativization Model, or Giles’ Speech Accommodation Theory, as cited in Giles, Coupland and Coupland (1991: 5), in Ellis (2003: 128). However, most of them address naturalistic SLA in the TL environment and not instructed FLL and thus will not be further discussed. The interested reader may visit Long (2007) and VanPatten and Williams (2007) for a detailed discussion of the current SLA theories. Long (2007) reports there has been such a theory proliferation (they count as many as 60) witnessed since the 80s and the 90s that the field of SLA has become increasingly fragmented. These differ in form, hypotheses and generalisations made that are of greater or lesser certainty. Some consider this a healthy sign others believe it constitutes a serious problem. As to day, there is no agreement on a ‘complete’ theory of SLA.
2.4 Chapter Summary

It would not be an exaggeration to say that the epitome of humans is the precious tool of language we have been endowed with by nature. Apart from being an expressive tool of thoughts, ideas, etc., language has determined our very existence in this planet and our evolution as a species and will continue to do so. To put it in Ortega’s words (2007: 235): “the language faculty is a unique form of intellectual organization” that “enters crucially into every aspect of human life, thought, and interaction” and “is largely responsible for the fact that alone in the biological world humans have a history, cultural evolution and diversity of any complexity and richness”.

Theories of language learning that have been proposed have moved from what is observable in people’s behaviour (Behaviourism) to the driving force(s) that shape this, what Chomsky called “the inner mechanisms of the mind that enter into thought and action” (1986: 3) and “the ways these operate in executing actions and interpreting experience” (Chomsky, 2000: 7).

Current SLA research places great emphasis on the study of the cognitive processes that are involved during FLL, in particular the study of attention and awareness and L2 development through interaction (Long, 1996; Robinson 1995a, 2001; Schmidt, 2001). The literature suggests an interesting link between Working Memory (WM) (which will be discussed extensively in Chapter 6), noticing and interaction-driven L2 development (Mackey et al., 2002).

From what has been said so far, it is clear that language acquisition is an amazing achievement in one’s life. Therefore, the scientific study of the very process, be this L1 acquisition or a second or FLL, is an exciting field of research. Nevertheless, it is reminded to the reader that the present study focuses on the process of EFLL and in this sense is mostly interested in SLA studies.
Chapter 3 that follows will discuss the dominant cognitive theories that have influenced scientific thought so far and will present the major cognitive achievements witnessed in young children during the earliest years of middle childhood (i.e. 6- to 8-years of age).
CHAPTER 3
THEORIES OF CHILD COGNITIVE DEVELOPMENT

3.1 Introduction

The emergence and development of the ability to communicate through speech and move beyond the ‘here-and-now’ through logical reasoning are among the most striking cognitive achievements of early childhood, which according to Cole and Cole (2001) falls between 2½ and 6 years of age, while middle childhood falls roughly from 6 to 12 years of age.

Before moving any further, it would be wise to define Cognition. Das Gupta and Richardson (1995: 3) referred to Cognition as a ‘collective form’ for all the processes and types of mental activity that are involved in the organisation, handling and use of knowledge, such as “remembering, understanding, problem-solving, relating, imagining, creating, fantasizing, using symbols, etc.”. Broadly speaking, the development of cognition covers both the acquisition and the development of knowledge as well as the cognitive processes involved during this development, such as language, memory, and problem-solving. These relate to the operations of the human mind and as such they are the principal objects of inquiry of Cognitive Science (Long & Doughty, 2003).

3.2 Theories of child cognitive development

In Chapter 2 some of the most influential SLA theories that relate to the instructed context were presented. To gain a basic understanding of the cognitive, social and emotional development of the age group under investigation (6-8 years of age), though, it is important we consider two of the most influential theories of cognitive development, namely Constructivism and Social or Cultural Constructivism.

One of the most prominent proponents of the former theory was Piaget while Vygotsky and Bruner worked within the framework of the latter. The two theories differ as to where they place special emphasis on. Unlike Innatism, which holds that language is
encapsulated in a discrete innate module, both schools of thought hold that LL is a cognitive process just like any other, one of the skills to be learnt during childhood.

3.2.1 Constructivism

Jean Piaget is one of the most influential figures in developmental psychology and one of the most well-known proponents of *Constructivism*. He defined intelligence as a basic life function that helps individuals adapt to their environment. He viewed that language acquisition can be accounted for in terms of general cognitive developmental processes. He believed that children understand the external world and develop their cognitive abilities in a sequence of well-defined stages. To reach a certain stage of development and achieve a balance between themselves and the environment, children must go through all the preceding stages. Even though these are universal and the same for all humans, Piaget believed that development can be speeded up or slowed down by environmental variations such as, for instance, the presence or absence of schooling (as cited in Cole & Cole, 2001: 36). He acknowledged the importance of social interaction but believed the social setting alone cannot explain the remarkable intellectual development witnessed in children. Therefore, he placed more emphasis on the role children play in their own cognitive development. He viewed them as curious explorers or little scientists who experiment a lot and construct higher levels of knowledge out of their physical interaction with the environment. Children make and test hypotheses and successively create mental structures or schemas (Piaget & Inhelder, 1956, as cited in Pinter, 2011: 8). Piaget believed that rational thought develops as the child gradually adapts to the environment and it is cognitive development that prepares the ground for language development (as cited in Cole & Cole, 2001: 316).

His theory offers an outline of children’s cognitive development at different ages. According to Piaget, all children pass through four successive stages of thinking, each being qualitatively different from all the rest (as cited in Das Gupta & Richardson, 1995: 7):
1. **The Sensorimotor stage** (from birth to about 2 years): During this stage, children discover the world by means of the physical actions they can perform. They learn to coordinate their sensory perceptions (i.e. the input they receive via their senses) and their simple motor behaviours, to gain knowledge of the surrounding world. One of the main achievements of this period is object permanence, i.e. the understanding that objects have substance and continue to exist when they go out of sight. In addition, children of this age are egocentric as they tend to consider the world entirely in terms of their own self, they tend to imitate the actions of adults while they also possess an active curiosity for their surroundings (Cole & Cole, 2001; Pinter, 2011).

2. **The Pre-operational stage** (from about 2 to about 7 years). This stage is further subdivided into the *pre-conceptual period* (from 2-4 years) and the *intuitive thought period* (from 4-7 years). The first sub-stage is still characterised by egocentric thought. Children lack decentration because they cannot yet focus on more than one quality of an object. They have not formed a Theory of Mind (ToM), as they are not yet able to make sense of the behaviours of others on the grounds of their mental and emotional states, points of view and beliefs⁹ (Carlson, Mandell, & Williams, 2004). Pre-conceptual children engage a lot in ‘collective monologues’ (Cole & Cole, 2001: 328) rather than true dialogues in play-like situations with other children. They focus on what they are doing by themselves without really worrying about their partners or without having any apparent need to communicate. Their thinking is dominated by what they directly perceive from the external world, but gradually they show the first signs of symbolic functioning. This is a transitional period in cognitive development, during which spoken language emerges. In the second part of this stage, the *intuitive thought period*, children begin to rely more on intuition to solve problems. They start to form logical categories, distinguish relations among things, and

---

⁹ To have a ToM means that one can attribute thoughts, desires, and intentions to others, in order to predict or explain their actions. It enables one to understand that mental states can be the cause of people’s behaviour.
handle the first arithmetical concepts. They leave behind the ‘here-and-now’ and start to think about objects they cannot see, hear, or touch and about people that are not physically present.

3. **The Concrete Operational stage:** (from 7 to about 11-12 years). According to Wood, (1998: 23), this is the age of ‘intellectual revolution’. Children of this age group are able to adopt alternative viewpoints (i.e. ToM) and handle two perspectives of the same situation simultaneously. Also, they can hold one characteristic of a situation in mind while they compare it with another. Cole and Cole (2001) attribute this to the increased memory abilities attested during this period. The concept of conservation is fully accomplished around the age of 7-8, although it emerges in some rudimentary form two or three years earlier. This enables children to predict the operations in the physical world, as they now understand that certain properties of an object, such as its size, density, length, and number remain the same even when the object’s outward appearance changes in some superficial way. Children’s thinking becomes more organised and flexible and they gradually become interested in the ‘what if’ type of problems. Their knowledge is organised in terms of categories (e.g. the category of animals: cats, dogs, cows, etc.) and hierarchical relations and they are now able to perform hierarchical classifications. Also, they intuitively make up connections between two or more things they need to remember. In addition, their thought becomes reversible as they come to realise that one operation can be negated or reversed by the effects of another. Operational thought, i.e. the ability to think logically, can now be applied but only to concrete objects, as children’s mental actions are still directed toward real objects that are present in their everyday activities. They can competently use analogies of the kind ‘if A is smaller than B and C is smaller than B, then …’. This period witnesses the full emergence of symbolic thought (where one thing may stand for another), decentration, gradual loss of egocentricity, relational logic (i.e. the
ability to mentally order a set of stimuli along a dimension). To sum up, operational thought means flexible, reversible reasoning which allows children to conserve, organise, sort, classify, seriate, coordinate perspectives and overcome misleading perceptual impressions (Pinter, 2011).

4. **The Formal operational stage** (from 11-12 to about 15 years): The cognitive capacities of early adolescents largely correspond to those of adults. Older children can think logically about abstract ideas or hypothetical situations. They proceed from hypotheses to systematic deductions, from the general to the more specific by generating hypotheses (Pinter, 2011). Internal representations and thinking can be manipulated flexibly and productively, quite detached from concrete objects or situations. Adolescents classify, order, and reverse mental operations that are applied to objects and events. They have the ability to consider the future. They can also manipulate abstract ideas and consider all sorts of relationships from a number of different perspectives.

Piaget suggested that cognitive growth is driven by three underlying processes that enable the transition from one stage to the next. Children constantly compare and adjust their developing schemas (i.e. the forms in which knowledge is structured in the brain) to match what they encounter in the environment. Through **assimilation** they interpret new knowledge in terms of already possessed schemas. If the new experience does not fit their existing schemas, then they **accommodate** and modify them to fit new knowledge (De Bot, Lowie, & Verspoor, 2006). Finally, during **equilibration**, when new experience is being incorporated into the child’s existing schemas, the latter are restructured to host the new aspects of knowledge into their frame. The more experience children gain, the more sophisticated ways they develop for organising information. Piaget viewed cognitive development as a continuous construction of intermediate equilibriums within children’s successive developmental stages.
According to Piaget and *The Cognition Hypothesis* reported in Harley (2001), language needs the prior emergence and consolidation of certain cognitive, motor, and perceptual achievements to develop. Piaget believed that concepts emerge before words and logical reasoning precedes the acquisition of language.

### 3.2.1.1 Criticism on Piaget

Although influential, Piaget’s theory has been criticised for underestimating children’s mental abilities. He was accused of attaching labels to developmental stages and different ages that may not always be true for every child. Das Gupta and Richardson (1995) as well as Pinter (2011) hold that children’s performance can be influenced by a number of other factors, such as the social and learning background, with formal schooling, for instance, facilitating the development of operational thought. They also report current research which is suggestive of the fact that Piaget’s stages are going down. Children show signs of having certain cognitive abilities much earlier than Piaget suggested. This is logically explained if one considers that nowadays children experience more things and become more efficient and quick learners earlier. The stimuli they receive are numerous and come from all sorts of sources as they get exposed to TV broadcasting, radio programmes, the Internet or computer games.

Donaldson (1978) also suggested that numerous children fail in some Piagetian tasks because these are not familiar to them. If children’s experiences do not relate to the task at hand, then this makes no sense to them. This can be rather confusing since, according to Chomsky, children are innately predisposed to make sense of whatever they hear, read or do. Donaldson (ibid) and Pinter (2011) also attribute the failure of children to the misleading phrasing of the questions. If the intentions of the experimenter are not made clear, then the child does not understand what (s)he needs to do in a task. Cole and Cole (2001) believe that with the proper training, prior to the execution of the tasks, children can perform a lot better in these tasks. Also, Das Gupta and Richardson (1995: 10) cite various studies which suggest
that performance on Piagetian tasks may vary from culture to culture. On these grounds, it was suggested that his findings lack universal applicability. Finally, Piaget’s theory has also been criticised for not paying enough attention to the role the social context plays on children’s development.

3.2.2 The socio-cultural account of Vygotsky

Unlike Piaget, Vygotsky and Bruner placed a lot of emphasis on the social, cultural and historical context where the child grows and functions. Cole and Cole (2001: 36) define culture as “the designs for living that are based on the accumulated knowledge of a people, encoded in their language, and embodied in the physical artifacts, beliefs, values, customs, and activities that have been passed down from one generation to the next”. Viewed from this perspective then, socio-cultural constructivists hold that apart from the human biological predisposition (i.e. nature) and the physical environment there is a third factor which explains children’s cognitive development, that of society and culture (i.e. nurture).

Vygotsky believed that social interaction is the driving force of cognitive development. Ortega (2007), in a discussion of Vygotsky’s theory, noted that human cognition arises from the material, social, cultural, and historical context in which all human experience is embedded. Vygotsky greatly contributed to developmental psychology with his Sociocultural Theory (SCT), with which he wished to demonstrate the interplay between cognitive development and education (van Geert, 1998). He viewed education in a broad sense, one that covers not only instruction but also play, educational help, imitation, and learning. Vygotsky believed that education is essential for the individual to reach the highest levels of thinking (Das Gupta & Richardson, 1995).

Vygotsky introduced the idea of the zone of proximal development (ZPD) (where ‘proximal’ means ‘next, nearby’). Vygotsky (1978) defined this as the gap between a child’s:
… actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers. (p. 86)

Lantolf (2005: 105) defines the ZPD as “the site where future development is negotiated by the expert and the novice and where assistance is offered, appropriated, refused and withheld”. Central to the ZPD is the concept of ‘scaffolding’, a vital concept in Vygotsky’s theory, which is viewed as the amount of assistance provided by the expert to the novice at just the right time, to help the latter solve a task (Lantolf & Thorne, 2007). Assistance needs to go slightly beyond children’s current competence, to build on and extend their current abilities. It may take the form of supporting the novice with information but also of making intervention moves to simplify a task or of encouraging the novice to persevere with a task (Pinter, 2011). As the child gradually extends his/her skills to higher levels of competence, the scaffolding is slowly removed (Das Gupta & Richardson, 1995). The term was originally coined by Bruner and colleagues. Wood, Bruner and Ross (1976) referred to the transactional nature of tutoring which involves:

… a kind of ‘scaffolding’ process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts. This scaffolding consists essentially of the adult ‘controlling’ those elements of the task that are initially beyond the learner’s capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence” (p. 96)

Vygotsky believed that children initially need to be assisted by adults or more competent peers because their cognitive functioning is still limited. With assistance, they are able to move to a higher level of knowledge and performance. Once they manage to get there, this is already a first ‘victory’, indicative of what they will be able to do in the future, on their own (van Geert, 1998). In this sense, social interactions awaken children’s ZPDs. Initially,
these zones operate only in the particular interactions created but with time, as children internalise them, they become part of their independent repertoires (Brown, Metz, & Campione, 1996). Cole and Cole (2001) note that ZPDs allow children to fully participate in activities before they are 100% ready to do so alone, i.e. if left in their own resources.

For Vygotsky (1934, as cited in Wertsch, 1985: 70-1), instruction creates the ZPDs as it “awakens and rouses to life an entire set of functions which are in the stage of maturing” and lie within the ZPDs. Learning and development are interconnected, in that the former lays the path for further development while it simultaneously prepares the ground for further learning (De Bot, Lowie, & Verspoor, 2006). Learning is of a social nature and children grow intellectually only when they are in active interaction with people in their environment. Viewed from this perspective, the idea of the ZPD is inextricably linked to SLA, as L2 learners advance to higher levels of linguistic knowledge when they collaborate and interact with speakers of the L2 who are more knowledgeable than they are, such as a teacher, a more advanced learner, or even a NS of the TL.

Vygotsky believed that in the first two or three years of life, language and thought develop more or less independently to then fuse and enter into a complex relationship of interdependence (Harley, 2001). When this happens, the nature of both language and thought changes and creates a unique mode of behavior in the growing child. Language becomes the vehicle with which the child communicates his/her thoughts to others, while at the same time it is through language the child converts the social reality experienced into idiosyncratic thought. During all this time the child’s thought and language are interestingly interconnected. Thought becomes verbal and speech becomes representational thought (Cole & Cole, 2001).

When children ‘internalise’ language, this constitutes the basis of their cognitive processes. They can think in language and engage in inner, soundless speech, which allows them to direct their thinking and behavior (Bruner & Haste, 1990). Once they manage to do so
they can engage in all forms of higher mental functions, such as reasoning about time or talking about objects that are not present. Vygotsky, (as cited in Das Gupta & Richardson, 1995: 13), believed that thought is the result of language acquisition. His theory attempted to describe the transition from the inter-individual (or inter-psychological) to the intra-individual plane. He believed that children’s new psychological functions first appear in their interactions with others who “support and nurture their efforts” and at a later stage they are consolidated (Cole & Cole, 2001: 329; Wertsch, 1985). Vygotsky (1960: 118) argued that “it is not nature, but society that above all else must be considered to be the determining factor in human behavior”. For Vygotsky, language is a tool that allows learners to regulate the process of their learning. Gradually and with increasing skill they become all the more autonomous and less dependent upon others. Self-regulation is achieved through interaction with peers where children can extract the information they need to carry out a task (De Bot, Lowie, & Verspoor, 2006).

The idea that child development is a process of gradual involvement into human culture is central in Vygotsky's theory (Veraksa, 2011). Learning is a process of becoming a member of a certain community, a greater whole. It requires active participation of the individual and engagement with others in an attempt to investigate the way things can be done with words (Pavlenko & Lantolf, 2000). In this sense, the understanding of human development is the understanding of how children acquire the symbolic tools of their culture, one of them being language.

### 3.2.2.1 Piaget vs Vygotsky

Both Vygotsky and Piaget took the view that conceptual organisation is a considerable achievement in middle childhood. They both saw children as active participants in their own uni-directional cognitive development. However, they assumed opposite directions in this course. Piaget believed that development stems from the individual plane and extends to the
social one, with knowledge being constructed from the inside-outwards. Children first understand concepts, internally construct cognitive processes and then they become able to use these properly through language, and thus modify their relationship with the external world. In other words, Piaget treated the individual as ‘the locus of change’ (Rogoff, 1998: 684) and believed that language has only a supportive role to play in children’s intellectual development. Vygotsky, on the other hand, was more socially oriented (Martí, 1996; Wertsch, 1985), for he believed that language is first acquired by the child externally, in dialogues with others to be subsequently internalised and shape human thought and behaviour. He identified two different roles for language: a means of communication and a social instrument which constitutes the basis of later abstract thought. Language and thought come together around the end of the 2nd year, when children learn to use words as symbols to express their thoughts.

3.2.3 Bruner’s socio-cultural explanation for language acquisition

Bruner also provided a socio-cultural explanation for language acquisition (Harley, 2001), according to which language can only develop in the context of meaningful interactions and knowledge is constructed through education, in the way Cole and Cole (2001: 505) view education, i.e. as ‘a form of socialization’ where adults assume the role of the ‘tutor’ to ensure that children acquired specialised knowledge and skills. Bruner believed that all aspects of language lay in the social interaction between infants and their caregivers.

He acknowledged that the first seeds of later language grow in the pre-verbal exchanges between mother and child, where the mother helps the initially helpless infant work out the meaning of her utterances so that it progressively takes a more active role (Bruner, 1977). The two sides (mother and child) are heavily motivated and enter into a co-operative relationship to build some kind of mutual understanding called ‘intersubjectivity’ (Bruner, 1983: 27; Rogoff, 1998). The infant is born with some rudimentary form of social predisposition. Its first reactions are elicited by a variety of signs from the mother, usually non-linguistic in
nature: her heartbeat, the look in her eyes, her characteristic smell, the sound and rhythms of her voice, her gestures, or any other physical action that can help the child infer the intentions of her utterances. Bruner (1977: 274) argued that what the child really learns about communication before language is the way to ‘crack the linguistic code’. To give an example, through sustained eye-contact and joint attention the baby learns to follow the mother’s gaze, pay attention to what she is attending to and eventually infer the implied meaning. In this sense, Bruner (ibid: 274) held that “the mastery of procedures for joint action provides the precursor for the child’s grasp of initial grammatical forms”. The child gradually passes from pre-speech communication to ‘the use of language proper’ (ibid: 273) with the help of these pre-verbal exchanges which have the structure of conversations and share some underlying similarities with them, such as joint attention to a topic. Gradually, the mother and the child develop a variety of procedures (such as turn-taking and ways to refer to things) to operate jointly and in support of each other.

Bruner (1983) viewed the child as a social being. Making sense of the world was for Bruner a social process. In this sense, language cannot be learned in vitro but is acquired only when used for communicative purposes. The child plays and talks with others, learns through interactions with parents, family, peers and later on teachers. During these interactions, via the language of the community, the child gradually acquires a framework that helps the interpretation of new experiences, the negotiation of meaning, the evaluation of events. According to Bruner, language acquisition begins well before the child utters his/her first lexico-grammatical speech. The amazing thing is that children learn how to do all this without deviating from the cultural norms of the community. The child, with the help of his mother’s ‘cultural sensitivity’, is being trained not only in the mechanics of the language as such but also in its proper use that is acceptable within a cultural community (ibid: 125).
Bruner (1983) held that what aids LL is the existence of a Language Acquisition Support System (LASS) in the child, which frames all child-adult interactions. Bruner viewed the LASS as a vehicle for the transmission of culture. The LASS helps the child learn how to say things by making well-formed utterances that conform to the rules of grammar. In addition, it helps the child learn how to get things done with words and be communicatively effective by producing language that “is canonical, obligatory, and valued among those to whom he says it” (ibid: 120). Cole and Cole (2001: 318), cite Bruner (1982), who noted that the interaction between Chomsky’s LAD and his idea of the LASS makes possible the infant’s entrance into the linguistic community. Cole and Cole (2001) find this very crucial, as the child’s incorporation into an already existing language community guarantees its continuity and survival. Thus, Bruner (1982, 1983) viewed the LASS as the environmental complement to the LAD, the union of which gives birth to language. Bruner and Haste (1990) held that children’s development depends on their use of this cultural tool kit of the community to express their thoughts.

The LASS consists of basic scripts or formats, that caregivers provide the child with. According to Bruner (1983: 132), “a format is a routinized and repeated interaction in which an adult and child do things to and with each other”. Formats are socially patterned activities that surround routines like bathing, bedtime, having meals or game-playing. They provide a familiar sequence of events wherein children learn how to express their communicative intentions clearly, with the help of the caretaker feedback. Language acquisition, then, begins when mother and infant co-create such formats of predictable interaction that become their ‘microcosms’ or ‘microculture’ while they communicate in a shared reality. The transactions that occur in such formats constitute the input from which the child then manages to master grammar and learns how to refer, mean, and realise his/her communicative intentions (Bruner, 1983: 18).
Given that play is the culture of childhood, these formats initially take the form of games like peekaboo or hide-and-seek, story-telling, crafts and art activities (Veraksa, 2011). Bruner (1981) believed that the foundations of language lie in these early game-like formats because they give the language-acquiring child assistance in mastering the grammatical forms of their native language. Games provide the first occasion for the child’s systematic use of language with an adult. It is through them that the adult ‘tutor’ enters into a hypothetical situation, to provide a ‘scaffold’ for the child’s activities and become the model of the relevant linguistic rules. Games also provide an opportunity for distributing children’s attention over an ordered and ruled sequence of events. In this sense, they impose limits in their options. Games are the living example “of what and how one says things in what contexts” (Bruner, 1983: 127). They render familiar the semantic workspace in which utterances are to be used in the future. All this is accomplished in a playful, tension-free atmosphere, where the child is left free to innovate and exercise his/her combinatorial abilities while in interaction with others (Bruner, 1977; Ratner & Bruner, 1977). For Bruner, child play is a serious business as it helps the maturation of the child’s skills, while it permits the interaction of the child with other members of the community, especially adults who serve as the child’s models. It is the adults’ responsibility to ‘educate’ the preschoolers and pace these activities in ways that ensure the smooth passage of experiences, values, customs, via the ‘vehicle’ of language.

With experience and more systematic use, these early formats are conventionalised at a later stage to be finally transformed into speech acts, such as indicating or requesting (Bruner, 1983). These basic scripts are assembled into higher-order subroutines to constitute the basis of the more complex social interaction and discourse that the child will experience in the near future.
To conclude, Bruner and Haste (1990) believed that in these parent-child interactions meaning is jointly constructed and the child’s development is facilitated in two ways. Firstly, the adult is scaffolding the child’s efforts to make sense from experience. Scaffolding takes the form of recruiting the child’s interest in the task, reducing the number of the required steps by simplifying the task, motivating the child to maintain the pursuit of the goal, correcting the child’s early utterances, controlling frustration, offering suggestions (Bruner & Haste, 1990; Rogoff, 1998; Wood, Bruner, & Ross, 1976). Secondly, the adult is giving the child a set of grammars and ‘scripts’ to make sense of the world (Bruner & Haste, 1990). With time, children learn how to master autonomy and independence by constructing their own conceptualisations and by finding solutions to problems by themselves.

3.3 Cognitive science and the study of Memory

The principal object of inquiry of Cognitive Science is the human mind and its goal is to understand the operations of the brain (Long & Doughty, 2003). Memory development has been one of the most-studied topics in all of cognitive development as it is at the center stage of cognition (Coulson, 1995). Since the mid-1960s Cognitive Science and the research on memory development has been influenced by theoretical models that adopted an information-processing (IP) view. The IP approach employs the computer metaphor to illustrate how the human mind works. Following this metaphor, the human mind is studied both in terms of its hardware (i.e. nervous system, the brain, sensory receptors, neural connections, hardware improvements as the child matures) and its software (various mental programmes such as rules, strategies for remembering, organising, evaluating all incoming information) (Pinter, 2011). This computational logic was facilitated first by the dramatic increase in the accessibility and use of computers in numerous areas of public and private life and secondly, by the fact that new brain scanning and imaging techniques were developed which
CHAPTER 3: THEORIES OF CHILD COGNITIVE DEVELOPMENT

contributed to a greater understanding of the functioning of different areas of the cerebral cortex in the brain.

Memory is divided into short-term memory (STM) and long-term memory (LTM). The first is characterised by an extremely limited capacity\(^2\) and a fragility of storage, as any distraction activates the forgetting mechanism (Eysenck & Keane, 1995). The capacity of LTM is essentially unlimited as it can hold information over extremely long periods of time. Knowledge in LTM is organised in two different ways: chronologically and semantically. Declarative memory, i.e. the conscious remembering of facts and events, is organised in the chronological order these occurred. It is a kind of autobiographical record of one’s life experiences. Semantic or procedural memory, i.e. the automatic, unconscious memory for skilled behavior, refers to our knowledge of the world, i.e. language, rules and concepts (Schneider & Bjorklund, 1998). It is memory for facts that is organised by meaning, the near-permanent knowledge we have about the world (Gathercole, 1998). Semantic memory is characterised by a lack of recollection about the specific occasions on which the semantic knowledge was established. These are the things we just ‘know’ out of experience.

Several models have been proposed to explain how different types of information are held in semantic memory, how they are linked to each other, and how accessing one memory trace enables us to access others. Coulson (1995) holds that two are the prevalent paradigms that try to explain conceptual memory: the IP model based on semantic networks and the connectionist model based on connectionist networks.

3.3.1 The Information Processing account

The Information processing paradigm focuses on the gradual building of knowledge and the development of automaticity through practice. Information is taken via the senses and then

\(^2\) The term capacity is used with respect to the amount of information that can be held in the memory store or the length of time this information will stay alive in there (Schneider & Bjorklund, 1998).
various features are extracted through a series of memory stores. Information is stored as ‘symbols’ or meaningful units (Bancroft, 1995a).

Randall (2007: 14) discusses the multi-store model of Atkinson and Shiffrin (1968) and its three types of memory: the Sensory Register, WM or STM and LTM. These represent different stages in the serial processing of information. First, the external stimuli are taken via the senses to enter a short-lived sensory register. Essential information is extracted from the massive amount of ‘raw’ incoming stimuli, i.e. light or sounds, to form memories of a very short duration (1.5-2 seconds). The material attended to is further processed, is recoded and rehearsed in order to be kept ‘alive’, to eventually be permanently stored in LTM where it is kept until it is retrieved when needed (Eysenck & Keane, 1995: 125). A failure to remember can be due to a failure at any of these three stages (Coulson, 1995). Chase and Ericsson (1980) suggested that already available schemata in LTM act as ‘placeholders’ for newly presented information. These facilitate the faster transfer of incoming information to LTM and thus avoid congestion in STM (Langley & Simon, 1980).

In the same vein, Neves and Anderson (1980), in their discussion of skill acquisition, knowledge compilation and automatic performance of actions, described their model of Automatisation. According to this model, LTM is divided into declarative and procedural memory, both linked to the central executor. Automatisation comprises three stages: during the encoding stage, a set of facts required by the skill are committed to memory. All information is encoded declaratively, as individual episodes that form a semantic network. During proceduralisation these are turned into procedures by means of production rules, whereas in the last stage of composition, these procedures are made faster with practice until a stage of effortless production is reached. Automatisation for Anderson (1983, as cited in Randall, 2007: 133) is the conversion of declarative knowledge into procedural one. Johnson (2001) believes that one cannot live without the presence of the other. Automatisation as a
process in LL is very important because processes that become automatic demand little or no attentional resources and they can free up spare capacity for new tasks. Automatisation also helps speakers retrieve huge chunks of language from their LTM store without wasting time to consciously think about the separate constituent parts. This, according to Pinter (2011) allows speakers to develop fluency in the L2.

Randall (2007) notes that through perception, the IP system is constantly interpreting incoming information. When first captured by the senses, the information flow follows a bottom-up, stimulus-driven processing. Then, the incoming data interacts with previous experience already stored in LTM, which is a top-down or conceptually-driven processing. This interaction facilitates the interpretation of the sensory data. Although our senses are bombarded with an amazing amount of information, we become consciously aware of only a small portion of it as little actually manages to pass on beyond the sensory store. This is because we have a limited IP capacity (McLaughlin, 1990, as cited in Ellis, 2003: 390). In this sense, Schmidt’s idea (1990) of noticing is crucial for LL, as learners are not capable of attending to all the information available in the input and therefore need to prioritise where to allocate their attention to (Skehan, 1998). Randall (2007) holds that this is the responsibility of the Supervisory Attention System or the central executor, which consciously directs attention to certain aspects of the data and discards whatever seems to be irrelevant to the task at hand. For instance, to make immediate sense of incoming language stimuli the Supervisory Attention System needs to attend to data that is language-specific and to reject all other. Then, it will be able to extract the salient information from this data to create language ‘symbols’ from the raw physical input (ibid: 39). Robinson (2001) places great emphasis on the computational work undertaken by the learner with respect to attention and finds this crucial for learning, because only the input encoded in STM can be subsequently transferred to LTM.
Gass and Mackey (2007) reported that more noticing takes place at lower developmental levels whereas less noticing occurs with low WM capacities.

### 3.3.2 The Connectionist account

Quite contrary to the IP paradigm, the *connectionist account* views the human brain as a machine capable of operating at multiple levels in different areas simultaneously. According to this account, the brain has a rather complicated and sophisticated neural architecture that contains a vast number of neurons which connect to thousands of others via synapses to make up neural networks (Randall, 2007). This enables the parallel distributed processing of information, during which multiple levels of activity take place simultaneously, as activation is spread through many parts of the brain (Collins & Loftus, 1975). Because a great deal of memory is associative, one activated concept spreads its activation to all the concepts to which it is linked. Automatic learning results from the strengthening of some neural connections (in the case of highly frequent linguistic features) or the inhibition of others (in the case of less frequent items) that all build an associative network. The human neural system has the ability to form and retain such kind of networks. Infants are born with a neural system which is further developed on the basis of experience, i.e. what excites or inhibits their senses. Incoming information extends and modifies the initially available system to incorporate and host new material (Bancroft, 1995a).

Connectionist approaches work within the framework of artificial intelligence, neuro- and psycholinguistics and try to explain the workings of the brain and the architecture of conceptual memory. To this end, artificial models are built by computers that simulate the architecture and the operations of the human mind.

One of the earliest artificial models designed was that of McClelland and Rumelhart (1981, as cited in Randall, 2007: 59), which simulated the workings of the human mind to explain letter recognition and past tense acquisition. Incoming information took the form of
numbers that were easily processed by the computer. The strength of connections between the nodes was given a numerical value that represented the probability of one node co-existing with another. This process of interconnectivity reflected the probabilistic relations that might exist between linguistic features. The network was gradually ‘trained’ by the incoming linguistic data. The central idea was that after repeated presentation of the same stimulus, the network would eventually learn to produce a specific output.

Connectionists attribute greater importance to the role of the environment than to any innate knowledge of the learner. They don’t believe in the existence of a neurological module that is specifically designed for language acquisition. What is innate is the potential that all humans possess to learn in general. For them, the input is the primary source of linguistic knowledge, which is cumulatively learned and gradually built through multiple exposures to instances of particular linguistic features.

This theory has received a lot of criticism in that it makes use of controlled laboratory experiments, analyses frequency patterns to build artificial systems and computer simulations that can only replicate but not explain language behavior. Randall (2007) holds that the input from the linguistic data actually trains the network and the brain ‘learns’ a language from the input received which, according to Schmidt (2001: 4), is a rather ‘dumb’ process. Lightbown and Spada (1999) as well as Randall (2007) question the validity of such models and their power to explain actual human learning, since so far only a few, fairly simple areas in language, have been investigated and these piecemeal findings cannot be generalised to account for the totality and complexity of normal human LL. Eysenck and Keane (1995) also held that unlike the human brain, the computer has no understanding of what it is doing or why it is doing it, does not take account of non-cognitive factors (e.g. biological influences), while it lacks any social or moral dimension.
3.4 Children’s cognitive development in early childhood

During early childhood (between 2½ and 5-6) young children gradually acquire the ability to represent reality to themselves through the use of symbols, mental images, gestures, and words (Cole & Cole, 2001). This is the period during which they engage a lot in pretend and symbolic play (Das Gupta & Richardson, 1995). Preschoolers tend to think about only one salient feature or dimension of an object or event (Cole & Cole, 2001). Their reasoning is semi-logical as their thoughts are linked together rather loosely, lacking logical organisation (Miller, 1989). Although this is the case, they are able to imagine, think of the future, or recall the past. Still, their reasoning is bound by their perception of the environment and as such it is not yet fully operational. This is perhaps why Cole and Cole (2001: 338) believed that during early childhood cognitive development should be best seen as a “process of overcoming the limitations that stand in the way of true operational thought”, to use some of Piaget’s terminology.

Coulson (1995) held that the first five years of life are characterised by an almost total lack of well-structured memories. This is really vital because the development of memory is inextricably linked to the issue of learning in general. The author defined memory as the interaction between the information under consideration, the context in which this is presented, and the understanding of the individual whose memory is being examined. He noted that memory is affected by one’s understanding of the world and understanding increases with age. Also, he believed that any memory limitations attested in early childhood are due to lack of strategy use and organisation rather than limited capacity as such, as young children of this age spectrum are not able to spontaneously generalise strategies to new situations.
3.5 Children’s cognitive development in middle childhood

During middle childhood (i.e. between 6 and 12 years of age) children achieve a number of things (Cole & Cole, 2001). At a physical level, they become stronger and more agile. With regard to their emotional development, their friendships are characterised by the sharing of interests, the building of mutual understanding and the creation of trust. The nature of their play changes as they move from pretend and symbolic games to rule-governed ones. This is the period of rapidly developing metalinguistic thought. Children are able to reflect on their thoughts and language use, analyse unknown words and make inferences (Pinter, 2011). They become more logical in thinking although this is not yet abstract. They are able to categorise objects, consider multiple aspects of a problem and imagine other people’s perspectives. They are increasingly adept at turn-taking, topic maintenance, and the pragmatics of making speech-acts such as requests. They already possess a highly developed L1, and their language is increasing in vocabulary size and grammatical complexity. Their oral and written literacy is still developing, but they are exposed to a greater range of text types as they move through schooling.

Considering the biological achievements of middle childhood, Cole and Cole (2001) believe that the maturation of the brain witnessed throughout this period, in particular from 6 to 8 years of age, plays an important role in the further development of children’s thinking. Their thought processes become more systematic and can be applied across a wider variety of settings. The brain continues to grow, becomes more effective as its different parts coordinate in a better way, and develops specific kinds of functioning such as myelination and the further strengthening of the neural pathways. Myelination “provides the axon of cortical neurons with an insulating sheath of tissue that speeds transmission of nerve impulses” (ibid: 474). Both these functions, the authors hold, may underlie changes in cognitive skills. Children are now able to perform tasks on their own, although their actions are still bound by their perception
and the events they directly experience. Around their 7th or 8th birthday they become capable of performing a number of mental operations, such as forming combinations, separating, ordering, and transforming information in a logical manner. This is the time that adults begin to treat children more like grown-ups.

Cole and Cole (2001) count a number of cognitive processes and language achievements that facilitate children’s cognitive performance during middle childhood:

a) the ability to self-control and sustain attention in the execution of a task, which happens around the age of 8 (Flavell & Miller, 1998). Children are now better able to resist the temptation to abandon the task assigned, concentrate more on what they are doing and thus obtain information more efficiently. Their deeper involvement with the task itself enables them to inhibit attention to distracting events. Eysenck and Keane (1995), however, noted that anxiety may have a detrimental effect on their processing efficiency and thus lead to poorer retention, as anxious individuals tend to be more distractible than non-anxious ones,

b) the ability to form explicit plans for achieving a goal. To make a plan, children have to keep in mind what is presently happening, what they want to happen in the future, and what they need to do to get from the present to the future,

c) the ability to think about and control one’s own thought processes and engage in self-reflection,

d) the increase of their linguistic abilities. The size of their vocabulary increases from 10,000 words at the age of 6-7 years to 40,000 words at 10-11 years (Cole & Cole, 2001). As a consequence, children are better able to make logical inferences about a wide range of events. Their knowledge of categories to which objects belong is now well-founded into their active vocabularies and thus children can more easily manage classification tasks. Children between 6 and 12 years of age, use semantic relations (superordinate and subordinate ones) to group things together. They divide the world into classes of things to
decrease the amount of information they need to learn, perceive, remember or recognise.
This is done for reasons of cognitive economy (Eysenck & Keane, 1995). Superordinate

terms stand for an entire class or category of things. For example, the word *bird* is the
superordinate term, used as an umbrella term to cover the subordinate *canary*, which is a
more specific instance of this class. According to Bancroft (1995b), children categorise
new experience on the basis of shared similarity between concepts. This fundamental
ability helps them learn out of experience and interaction. They store new experiences,
compare one with another and recognise similarity and difference, even as early as infants.

During middle childhood children’s thinking becomes more organised and flexible.
While trying to solve problems, they think about alternatives and are able to reverse their
thinking. They view objects from more than one perspective and can hold two things in mind
simultaneously, to compare these with each other or with a third one. Piaget called this stage
*Concrete Operations*. Cole and Cole (2001) suggest that this kind of intellectual development
is usually the outcome of formal schooling.

The school context plays a key role in the expansion of children’s knowledge base and
the development and further enhancement of their memory strategies and certain cognitive
skills, such as logical problem-solving, memory and meta-cognition. The authors define meta-
cognition as the ability that enables one to assess the degree of difficulty of a problem in order
to choose the proper strategies that will help solve this in a flexible and effective way.

Around the age of 7 children learn to use effective strategies for remembering, such as
spontaneous rehearsal (the repetition of material in order to memorise it) or for organising
their memory, and this is partly due to formal schooling, i.e. the deliberate learning and
training efforts that take time and practice (Pinter, 2011). Before this stage, during
kindergarten, children exhibit a kind of “verbal production deficiency” as they “fail to use
naming and rehearsal as a cognitive ‘trick’ to aid their recall” (Flavell, Beach, & Chinsky,
However, Coulson (1995) reported that children fully appreciate the application of these techniques four or five years later, around adolescence. Nevertheless, once rehearsal starts to be used appropriately, children and adults only differ as to the amount of information they can rehearse. Children mentally group the material to be remembered in meaningful clusters of closely associated items. In doing so, they need to remember only one part of the cluster to gain access to all the rest. Depending on their age, children use different strategies for remembering. While younger children may use sound resemblance as they tend to remember words that may end alike, like cat and pat, children between 6 and 12 years of age use semantic relations, i.e. superordinate and subordinate terms to organise their knowledge. It seems that children of this age spectrum are better at formulating explicit strategies because they can think through alternative decisions and modify them mentally before they act.

Children seem to be able to think about their own memory processes and have some understanding of how their memory works (i.e. meta-memory). This, Cole and Cole (2001) hold, starts as early as the age of 5 and gradually increases. 8-year-olds better understand the limitations of their own memory, much more than 5-year-olds do. When it comes to memory organisation and the ways children mentally group things together to easily remember them, the differences are marked.

Vygotsky (1934, as cited in Wertsch, 1985: 184) placed a lot of emphasis on human memory and believed (1978, as cited in Wertsch, 1985: 191) that while very young, children heavily rely on their memory skills because they lack flexible and organised thinking, whereas the adolescent mind that is fully operational, makes use of memory skills to facilitate the operation of rational thought.

During middle childhood children gradually acquire an enhanced ability to store and retrieve information in a systematic way. The speed of memory processing with which children retrieve already stored information in LTM and the capacity of STM, i.e. the ability
to hold several items of information in mind at one time, both increase. So children experience an increase in knowledge about things they are trying to remember. When they need to remember something they have a richer knowledge base, already stored, to relate this to.

The memory span of their STM, i.e. the number of the randomly presented items they can repeat immediately after their presentation, also increases with age. Research findings indicate that 5-year-olds can recall about 5 digits, 7-year-olds about 6 digits, 9-year-olds about 7 ones and 11-year-olds can manage about 8 digits (Cole & Cole, 2001; Pascual-Leone, 1970). Chase and Ericsson (1980: 141) cited Miller’s findings (1956) with respect to normal people’s memory span that falls in a very narrow range (around 7 ± 2 unrelated items), that is fairly stable over a wide range of types of material. However, Baddeley (1994, as cited in Pinter, 2011: 22) held that this can be stretched to as many as 20 items, if one uses strategies to categorise the seemingly random numbers in a meaningful way. The lifespan of information also varies. Usually, it is forgotten within 1 to 2 seconds. Cognitive psychologists consider this an index of the limited capacity of STM, which severely constrains people’s ability to process information and solve problems. Coulson (1995), on the other hand, suggested that the STM span may reflect children’s current inability to use adult strategies rather than a limitation in their memory capacity. If this holds, it is quite encouraging, because it suggests that STM can be extended and significantly improved with the proper training in the use of strategies. Schneider and Bjorklund (1998) noted that only recently, psychologists came to realise that capacity is also affected by how quickly one can process and rehearse information. Coulson (1995) reported that the faster one speaks, the more information one can rehearse: 10 year-olds have 80% of the rehearsal capacity of adults and can speak about 80% as quickly as adults.
3.6 Chapter Summary

This study’s participants belong to middle childhood, in particular early middle childhood, that covers the age range of 6-8 years. This is a period in one’s life wherein major cognitive milestones are witnessed. During this period, children experience an increase of their linguistic abilities while their metalinguistic thought, i.e. the knowledge of how language works, starts to develop. They also exhibit a range of learning strategies and are capable of reflecting on their thoughts and language use, while their thinking becomes more logical, although not yet abstract. They learn to categorise objects and manage to consider multiple viewpoints and perspectives.

A number of theories have dealt with children’s cognitive development during this early period, with the most influential ones being Constructivism and Social/Cultural Constructivism. Although both schools of thought agree that LL is a cognitive operation just like any other that has to be learnt during childhood, they adopt a different perspective when viewing the very process. Piaget was one of the most prominent proponents of the former school of thought and emphasised the active role children play in their own cognitive development. He believed that knowledge is being constructed from the inside-outwards. Vygotsky, being a social constructivist, believed that social interaction drives the remarkable intellectual of children development. For Vygotsky, language is first acquired externally, in dialogues with others to be then internalised, and shape children’s thought and behaviour. Bruner on the other hand, also acknowledged the importance of social interaction but held that the first seeds of the child’s later language grow in the pre-verbal mother-child interactions.

Since the mid-1960s Cognitive Science has placed a lot of emphasis on the development of human memory, which is at the center stage of cognition, and its contribution to one’s cognitive and linguistic development. Today, the study of children’s memory development is
perhaps one of the most researched topics in all of cognitive development. Currently, cognitive research views LL from an IP perspective, which focuses on the gradual building of knowledge and the development of automaticity through practice.

As already mentioned, formal schooling can accelerate human cognition and the development of language, because it plays a key role in the expansion of children’s conceptual base and in the development and further enhancement of their memory strategies and other cognitive skills. In the following chapter we will discuss the issue of EFL within an instructed context across Europe, as one of the two primary schools of the study, the experimental school, introduces EFL from the age of 6.
CHAPTER 4
INDIVIDUAL DIFFERENCES, THE AGE FACTOR AND EARLY FOREIGN LANGUAGE LEARNING (EFLLE)

4.1 Learner differences: An introduction

Learning a FL is a complex process and for each individual this becomes a unique interplay between developmental processes, L1 constraints and transfer, instructional input or the conditions of the immediate context. In the early years, SLA addressed the universal characteristics of L2 acquisition, i.e. the common qualities across learners of different ages, in different settings, with different L1s. Nowadays, SLA research has focused on the numerous ways individuals differ, reflecting thus that a whole range of variables that are at work during this process that pertain to cognitive, affective, and social aspects of being a human (Ehrman, Leaver, & Oxford, 2003; Lightbown & Spada, 1999; Skehan, 1989). Ellis (2003) counts a number of learner interrelated differences that are generally acknowledged: learners’ beliefs about LL, their affective states (e.g. anxiety) during the L2 learning process and general factors that influence learning such as age, language aptitude, motivation, cognitive styles, or learning strategies. Some are considered to be fixed and immutable while others may display variance, depending on the social setting and the course of L2 development.

This thesis, though, focuses on the early introduction of FLL (i.e. on the first two years of primary education) and will thus concentrate on two of these IDs, namely FL aptitude (which will be discussed in depth in Chapter 5) and Age. The latter is one of the most controversial variables and has been associated with the so-called Critical Period Hypothesis (CPH). The controversy centres round whether there is a critical period for L2 acquisition and if so, when this ends.
4.2 The Critical Period Hypothesis

The question of whether there is a critical period for language development has generated a heated debate among researchers the last five decades. The CPH, based on biological evidence, sought to explain the ease with which children naturally acquire language(s) and the more effortful and less successful language acquisition that takes place at later ages.

The CPH is closely connected to the names of Penfield and Roberts (1959), who initially discussed this with reference to L1 acquisition and later extended it to SLA. Penfield and Roberts (ibid) as well as Penfield (1953, 1964, as cited in Walsh & Diller, 1981: 3), used evidence from neuro-developmental studies and predicted an optimum for L2 learning between 4-8 years of age, when the brain still retains its plasticity. Language processing initially involves both hemispheres to be slowly concentrated on the left for most people, a process known as lateralisation. Once lateralisation is completed with the start of puberty, the brain loses much of its plasticity and, consequently, its ability to take up a new language system naturally and without effort. Therefore, Penfield and Roberts (1959, as cited in Singleton, 2003: 4) suggested that the best period to begin FLL is early childhood, between the ages of 4 to 10: “for the purposes of learning languages, the human brain becomes progressively stiff and rigid after the age of nine”. So far, there is no agreement as to when exactly this period ends. Krashen (1973, as cited in De Bot, Lowie, & Verspoor, 2006: 67), for instance, held that lateralisation is completed around the age of 5.

Later on, Lenneberg (1967) formulated his own CPH account, based on neurological studies and observations of deaf children. He believed in the inherent propensity of humans to learn a language and in that language acquisition is successful and takes place naturally and effortlessly only when this innate predisposition, what Chomsky (1979) called the LAD, gets stimulated at the right time. Lenneberg argued that the critical period begins around the age of 2 and is completed with puberty. Also, he viewed children to be more effective learners than
adults due to the flexibility of their brains. He predicted eventual native-like levels of L2 proficiency in pre-puberty learners and less than native-like mastery in post-puberty learners. Beyond puberty, he held, people lose a unique chance of automatic and natural language(s) acquisition as the brain matures, at the expense of its plasticity. After this point, Lenneberg (1967) held, FLL becomes a tedious enterprise that takes a lot of conscious and labored effort.

The idea of a biological critical period for language acquisition was also supported by Johnson and Newport (1989) as well as Newport (1990) who examined morphosyntactic English structures, to conclude that an early start (before the age of 17) is a good predictor of native-like performance in the L2. While memory (but not language ability) seems to play a crucial role in the early stages of L2 learning (Harley & Hart, 1997), cognitive maturity (Johnson & Newport, 1989), verbal analytic and problem-solving abilities play a greater role in the later stages (DeKeyser, 2000). DeKeyser (ibid), one of the strongest proponents of the CPH, suggested a gradual decline of one’s proficiency that stretches from about 6 to around 17 years of age which is caused by the loss of one’s ability to learn implicitly after a certain age. Bley-Vroman (1990) also viewed that later L2 learning is subserved by different mechanisms but employed a different explanation. Children learn their mother tongue naturally because they have access to UG principles and parameters, whereas adults and post-pubertal learners need to rely on their higher verbal and problem-solving capacities to learn the L2, as UG is no longer available after that age.

4.3 A critical period vs multiple sensitive periods during L2 acquisition:

Biological and maturational accounts

In biology, a critical period refers to a limited phase in the development of an organism during which a particular activity or skill must be acquired if it is to be incorporated into the behaviour of that organism (Singleton, 2003; Singleton & Ryan, 2004). The CPH claims that there is a biologically pre-determined period, optimal for grammar acquisition. When L2
learning starts after this favourable period that ends around puberty, one cannot acquire a native-like proficiency in the L2.

In this sense, the original CPH sounds rather absolute as it implies a definite cut-off point around puberty, beyond which successful L2 acquisition is an impossible achievement. Nowadays, the conflict rests more on whether completely successful acquisition is only possible or just easier within a given time span in one’s life. Singleton and Ryan (2004), in an overview of evidence coming from neurological research, deaf studies, accounts of feral children and Down syndrome subjects, conclude that the relevant studies indicate no such cut-off point. In their discussion of the CPH, they also cite (ibid: 57) evidence from Carroll (1971) and Diller (1971) who claimed there is no end point with respect to lexical development, as research findings suggest that vocabulary increases significantly up to roughly one’s 40-50 years.

Instead of a critical and well-defined ‘window of opportunity’, Eubank and Gregg (1999), Hyltenstam and Abrahamsson (2003) as well as Long (1990) put forward the idea of a ‘sensitive period’. According to Eubank and Gregg (1999: 68), a sensitive period represents “a progressive inefficiency of the organism or a gradually declining effectiveness of the peripheral input” after a certain time. In the same vein, Hyltenstam and Abrahamsson (2003) as well as Long (1990) argue that language sensitivity does not disappear at a fixed point but fades away over a longer period of time that possibly covers later childhood, puberty and adolescence, as lateralisation is a gradual process that may carry on over many years (Seliger, 1978).

Research that suggests the existence of multiple sensitive periods for the different components of grammar and lexis is abundant. There is still much disagreement as to the end point of each although most, but not all, researchers agree that phonology is the one component that is affected first. Olson and Samuels (1982) suggested an adult superiority
over children’s performance with respect to FL pronunciation, while Bongaerts, Planken and Schils (1995) found that a late start, around the end of the CPH, does not always block the possibility of a native-like pronunciation. This, they held, is especially true when the languages in question hold a close typological relation, as is the case of English and Dutch, and when the learners are highly motivated, receive massive input from NSs and intensive training in the perception and production of the British speech sounds.

Long (2007) indicates four sensitive periods with respect to SLA: one for categorical perception which stretches from birth to 6-9 months, a second for phonology that cannot exceed the 6 years of age, a third for lexical or collocational abilities that may extend up to 4-7 years, and finally the morphosyntax period which may reach the age of 17.

Oyama (1976) reported that learners who start as children achieve a native-like pronunciation, unlike those who start as adolescents, as this ability gets gradually lost between 4 and 7 years of age and suggested (1982a) an early introduction of FLs into the 1st grade as she believed that superior performance in phonology is the result of both inborn propensity and early training. This does not imply that L2 acquisition is an all-or-nothing phenomenon after a certain age, since evidence shows that very often adults do manage and learn to speak new languages very well (Oyama, 1982b). Oyama (1982a: 33) viewed the sensitive period as ‘a finely graded sequence of periods’ that involves different types of motivation, cognitive and emotional learner involvement: “Each period presumably involves both a special sensitivity to certain aspects of input, perhaps highly abstract, and a special propensity to process data in a certain way”. Scovel (1988) believed that the period within which learners can master the L2 phonology to pass for natives may even reach the age of 12.

Fathman (1982) also adhered to the idea that the ability to learn certain aspects of an L2 is age-related and that there are optimal times for learning different aspects of an L2. Younger learners (6- to 10-year-olds) are better at discriminating, interpreting, or imitating sounds
while older L2 learners (between 11 and 15) score higher in morphology and syntax as they are better able to make generalisations or memorise patterns.

Patkowski (1982) allowed for the possibility of an adult acquiring a FL, but not to the extent of passing for native. Nativelike acquisition of L2 syntax can only be achieved if learning begins before the age of 15. By that time the learner has already attained a high level of cognitive development to be able to make use of the genetically based language acquisition system. However, she raised another interesting issue. She believed that formal schooling instruction helps learners better develop their learning strategies and support their own learning and thus go through the learning stages even faster.

Flege, Yeni-Komshian and Liu (1999) as well as Walsh and Diller (1981) believed that the phonetic/phonological aspect of language ‘closes’ earlier than any other. Therefore, Walsh and Diller (ibid: 14) held the best time to master pronunciation is before puberty, during the earliest stages of neural development, as the later developmental stages that follow puberty are more pertinent to “the integration of higher order linguistic processes”, like semantic processing, and word-object relationships. In their effort to explain the general superiority of older students over younger ones, they provided a neurological account of studies which suggest that different brain cell types serve phonology and morphosyntax. Phonology, which is a ‘lower-order process’, is subserved by pyramidal cells, while Morphology and Syntax, both ‘higher-order domains’, are subserved by stellate cells. Because the former follow a maturational cycle that closes earlier than that of the latter, the authors held that pronunciation is first affected and a foreign accent is difficult to overcome after childhood.

Stevick (1976) approached the issue of near native-like pronunciation and performance somewhat differently. He employed a psychological explanation as he viewed that native-like accent is very difficult to acquire after adolescence, because pronunciation runs “deeper into the center of the student’s personality than any other aspect of language” (ibid: 64). He spoke
of the egocentricity of adolescents, who believe the rest of the world is only thinking about them and is preoccupied with their appearance and behaviour. This egocentrism may lead to increased self-consciousness, feelings of vulnerability, a lowered self-image and finally to a high affective filter, which in turn, lowers their overall ability to successfully acquire an L2.

However, there are voices that strongly disagree with the idea of a successful FL performance being evaluated and measured against the NS standards. Preston (1981) put forward the idea of the ‘competent or successful bilingual’ as opposed to the successful and efficient L2 learner that faithfully and blindly follows the NS-model. L2 learners strive to attain native-like standards which to his view, is impracticable, unachievable and also unnecessary as very often there is not enough time for such an achievement. Cook (1995), on the other hand, viewed that accent is the least important aspect of L2 proficiency therefore older students who fail to acquire a native-like accent do not lose much. He also believed it is both wrong and unfair to judge the competence of a person who knows more than one language (‘multicompetent person’) with that of a monolingual speaker who serves as the standard of perfection, as these are two qualitatively different states of mind: “Failure is defined with reference to monolingual native speakers’ and L2 learners’ performance is measured against that of native speakers” (ibid: 54). Taking a similar line with Preston, Cook suggested the term ‘effective bilingual’ or ‘multicompetent’ person (ibid: 55).

The question, however, is not whether there is one critical period or many sensitive ones but rather whether the attested child-adult differences do exist because of them. Singleton and Ryan (2004) find ‘the younger the better’ version of the CPH to be outdated and rather simplistic as current research indicates that things are much more complex than that. Instead, they put forward ‘the younger the better in the long run’ compromise view which is based on three arguments. First, young children’s superiority can be attributed to the fact they have more time to spend than adults and thus are more exposed to the L2. Second, it is very
difficult, if not impossible, to determine the boundaries of a critical period. Cook (1995: 52), too, believed “it is hard to find an overall critical period effect with a particular cut-off age” since age is ‘a factor from birth to death’. Third, despite the general difficulty of most adults to achieve native-likeness, some do manage to do so and thus provide counter-evidence to the CPH.

Nowadays, a more dynamic view of language acquisition has been adopted (De Bot, Lowie, & Verspoor, 2006; Singleton, 2001), according to which, age per se is not the decisive factor. Rather, the strong interaction between individual factors such as motivation, learning strategies, aptitude, gender, personality factors that constantly change during the L2 learning process as well as other variables that relate to a supportive scholastic environment, good quality teaching, authentic input, well-designed and carefully piloted language policies, etc., determine eventual success in the FLL process.

4.4 The Age Factor viewed from a non-biological perspective

A wide variety of views have been recorded in the literature on the age factor, that range from the position that children are in all respects more efficient as well as effective second language learners (SLLs) than adults to the exact contrary view, i.e. that adolescents and adults outperform children in all respects during the L2 learning process.

Rosansky (1975, as cited in Ellis, 1985: 108) spoke in favour of children and argued that cognitive development is to account for the greater ease with which these learn languages. The young child sees only similarities, lacks flexible thinking, and is self-centered. Young children automatically and implicitly acquire a language. Also, children have not yet developed social attitudes towards the use of one particular language as compared to another. Thus, they are cognitively more receptive to the new language. On the other hand, adults cannot learn an L2 automatically and naturally. The onset of abstract thinking around the age of 12 (Piaget’s Formal Operations), means they are predisposed to learn explicitly, recognise
differences and similarities, think flexibly, and become increasingly de-centered. They possess a strong sense of meta-awareness of how the language works and form linguistic generalisations. They are also likely to hold strong social attitudes towards the use of their native language against that of the TL. All these factors, according to Rosansky, may block natural language acquisition in adults.

Ellis (2003) attributes the general lack of consensus to the very different research methods and experimental designs these studies employ: some are longitudinal, others cross-sectional. Some examine naturalistic exposure, others instructed learning in SL or FL contexts, respectively. Krashen, Scarcella and Long (1982) believed that this contradiction could be resolved if one distinguishes rate of acquisition from eventual attainment.

4.4.1 Rate of acquisition

As far as developmental rate or rate of acquisition is concerned, research findings support the idea that older students are better and more efficient learners than children (Krashen, Long, & Scarcella, 1979). Nevertheless, these studies have been predominantly conducted in natural settings where the level of proficiency of immigrants in the TL has been examined on the basis of their age of arrival in the TL community (Muñoz, 2010). In such situations of naturalistic exposure, when the variables of time and exposure are held constant, older learners proceed faster than children through the early stages of language development, especially in the fields of syntax and morphology, due to their more advanced cognitive skills. Krashen (1982b) believed that the onset of Piaget’s Formal Operations around the age of 12 essentially enables the child to better understand and construct a theory of a language and its grammar. This is the age of abstract thinking and conceptualisation of linguistic generalisations, both cognitive processes that underlie language development. Genessee (1977) held that:
... the adolescent’s more mature cognitive system, with its capacity to abstract, classify, and generalize, may be better suited for the complex task of second language learning than the unconscious, automatic kind of learning which is thought to be characteristic of young children. (p. 148)

Krashen (1981a, 1985) also attributed the initially quicker progress of older acquirers to their advanced conversational skills that allow them to elicit more comprehensible input which meets their needs as well as to their active participation in face-to-face interactions where a lot of negotiation for meaning takes place.

Oxford (1989) held that adults are superior to children because they use a variety of different and more sophisticated learning strategies, such as metacognitive, affective, social, memory, cognitive, and compensatory ones. Adults use a lot of their metacognitive strategies, such as paying attention, self-evaluating, self-monitoring, that ultimately help them manage their own learning process by means of realising the conditions under which they learn best (see O’Malley & Chamot, 1990 for an extensive list of strategies, as cited in Ellis, 2003: 537). Children, on the other hand, have a poorer repertoire of strategies, basically memory and cognitive ones, like repetition or directed physical response (Pinter, 2011).

Advanced cognitive maturity is the one factor that explains the outperformance of older students over younger ones in two projects: the Basque country project (Cenoz, 2003; García Lecumberri & Gallardo, 2003; García Mayo, 2003) and the Barcelona Age Factor (BAF) Project (Muñoz, 2003). When exposure time was held constant, older beginners significantly outperformed younger ones on a variety of written as well as oral tests. Older students also displayed better perception and production skills (Fullana, 2006). The subjects were bilingual in Basque-Spanish and Catalan-Spanish respectively, with English being their 3rd language.

Snow and Hoefnagel-Höhle (1982) too, attributed the superiority of older learners over younger ones in vocabulary, syntax, morphology, even pronunciation in the short-term (the
first year of learning) to their cognitive maturity. Also, they held that the short-term phonological superiority of their informants provides strong counter-evidence against the CPH.

Ervin-Tripp (1974) believed that older learners display a more rapid L2 grammatical development in the early stages because they have a more efficient memory and a greater knowledge of the world. They learn word combinations faster and map the new words more quickly as they already possess a more developed and well-established semantic system in their L1.

### 4.4.2 Eventual attainment

*Eventual success* seems to favour young children (Krashen, Scarcella, & Long, 1982). The authors held that even though young children progress rather slowly in the beginning, they eventually surpass older students and achieve native-like levels of proficiency, provided they are given sufficient time and exposure to the L2. The number of years of exposure to the L2 and the starting age affect the ultimate level of success. The longer the exposure to the L2, the more native-like the L2 proficiency becomes. On the other hand, older children and adults, despite their initial advantage and superior cognitive abilities, very rarely reach such high standards in the long run. As already mentioned, Singleton and Ryan (2004: 137) put forward ‘the earlier the better in the long run’ view, on the grounds of abundant good supportive evidence and the absence of any counter evidence.

Krashen (1982b) also acknowledged that younger children do better in the long run because of their low affective filter that facilitates a native-like mastery of the L2. Nevertheless, he finds it is not impossible for an adult with a low filter to attain a native-like performance in the L2.

Scarcella and Higa (1982) attributed the eventual outperformance of children to the simpler input they receive, which facilitates L2 acquisition. They claimed that adult English
NSs provide larger quantities of simple input, a more supportive atmosphere, and constantly check to see that the child attends to and understands the input received. To their mind, the older L2 acquirers are better in the short-term, because they possess more sophisticated conversational strategies.

Oyama (1982a) suggested an early introduction of FLs into the 1st grade as she believed that superior performance in phonology is the result of both inborn propensity and early training. Oyama took into account non-maturational factors that contribute to children’s superior L2 learning, such as ample practice in the TL, heightened motivation, better imitating skills, although children’s weak point, the author acknowledges, rests on their short memory spans. With respect to vocabulary acquisition and an accurate foreign pronunciation, Yamada et al. (1980) suggested that younger children learn foreign words faster due to their better rote-memory skills and better motor ability. The first involves the association of words with the corresponding pictures (or meaning), while the second involves the accurate pronunciation of unfamiliar sounds.

Wong Fillmore (1976, as cited in Philp, Mackey, & Oliver, 2008: 11) and Pinter (2011) argue that, although age is inevitably important, young children’s eventual success may be explained by social, environmental and individual factors. It seems that younger learners typically enjoy special support such as the advantage of a nurturing and optimal environment, rich exposure to easy and simplified input, good educational opportunities as well as ample opportunities to practise and thus experiment with the language, high motivation, some explicit instruction and cooperative peers. All these factors facilitate successful language acquisition at a young age. Bialystok (1997), though, argued that if late starters also enjoy such favourable conditions, they may also achieve perfect mastery of an L2.

Asher and Garcia (1982) as well as Asher and Price (1982), found the idea of children’s superiority an illusory one. They acknowledged that children may have a pre-pubertal
biological predisposition to achieve native-like mastery of pronunciation, but believed that biology alone cannot determine near-native language acquisition. They took into account non-biological factors, such as the conditions of one’s learning, and adhered to the idea that physical action facilitates learning. Children’s superiority was explained on the grounds of their learning the new language in play-like, full of action and physical engagement situations while adults engage in non-play and static activities.

McLaughlin (1984) spoke of the child superiority which might be caused by psychological and social factors rather than biological ones. According to the author, children: a) are more exposed to the FL, b) have fewer inhibitions and are thus less embarrassed when making mistakes, c) speak more and receive more feedback, d) are more motivated, e) are placed in more situations than adults are, where they are forced to speak the FL. Finally, he held, there is more incentive in the playground or school as compared to the job context or a friendly meeting. All these advantages, he believed, should be utilised by means of an early introduction to a second language.

Schumann (1978) viewed things from a social-psychological perspective, with his *Acculturation Model*. The central claim was that success in L2 acquisition depends on the amount of exposure and the quality of contact the learner has with the TL and culture. He spoke of ‘cultural stress’ and ‘language shock’ and acknowledged that children normally have less difficulty than adults in adapting themselves to a new culture and feel less threatened by the sounds of the new language and concluded that this social and psychological maturation may be far more important than the neurological maturation supported by the CPH. Following the same line of thought, Preston (1989) believed that younger learners might be more successful than older ones, possibly because they do not feel their identity being threatened by the TL’s societal norms. Children are not subject to peer pressure as they have not yet formed stereotypes of their own identities.
To sum up, the nature of evidence is contradictory. However, two are the main findings about age differences in SLL: first, children learn better in terms of ultimate attainment, mainly because of their implicit learning mechanism. The long-term advantage of younger starters usually comes from studies of naturalistic SLA, i.e. contexts of full immersion in the language community. Provided children are given ample time and massive exposure to the TL, they may outperform adults (Munoz, 2006). Second, adults learn faster because their capacities for explicit learning let them take short cuts. In a traditional school context, however, where time is limited and learning is highly structured, DeKeyser (2003) argues that adults and older children learn more in the same amount of time.

The FL context, though, is a ‘low input level context’, quite different from naturalistic contexts as the amount of exposure to the TL is limited and children do not receive enough contact hours to ‘catch up’ in the long run (Pinter, 2011). The author continues by saying that, unlike full-immersion SLA contexts, there is very limited access to the TL outside the classroom in the FL context, with the exception of the Internet which has become a reality in everybody’s lives. Nowadays, it is easier for FL learners to have contact with native-speaking children in virtual environments or with a huge variety of authentic materials to access outside the classroom.

Krashen, Long and Scarcella (1982) as well as Singleton and Ryan (2004) recognise the importance of crucial factors such as exposure time and learning conditions - naturalistic exposure vs. instructed school context – in their capacity to differentiate levels of language proficiency and account for the discrepancy found in the research results. This is reasonable, as five years of exposure to a new language in a SL environment involves much more exposure to the TL than five years of formal schooling of the FL. In the former case, the TL is used both inside and outside the school across a wide range of activities with a variety of other speakers of the language. In the latter case, the language in question is simply one
school subject among many others, treated for only a few hours per week and delivered usually by non-native speakers (NNSs). Thus, Singleton and Ryan (ibid) believe that only rather long-term studies will be able to demonstrate the eventual benefits, if any, of early L2 learning in a formal instructional environment.

Although this is the case, one cannot deny that if one starts early, one gets the opportunity of more instructional time, is more exposed to the language, gets more practice and has more opportunities for interaction. Hatch (1983) supported the idea of EFLL: while still young, children have a low affective filter and high motivation. Why not then start young, when conditions are optimal to build the foundations upon which more sophisticated skills will be laid later on?

Long (2007) added another dimension to this discussion. He spoke of the enduring benefit of EFLL and of the positive effect that one or more additional languages, beyond the L1, have on later LL capacity. The early exposure to two or more languages as opposed to only one occurring before the end of one or more sensitive periods, he held, might trigger some kind of neurophysiological development that leaves early L2 learners with a lasting advantage that persists in adulthood.

Nevertheless, Singleton and Ryan (2004) warned that an early start in a FL does not automatically guarantee ‘eventual’ success. The matter is rather complicated as there are a number of other factors that come into play and contribute towards a successful FL outcome. Young FL learners, for instance, should be given the right learning conditions and more hours per week should be dedicated to the teaching of the FL. In addition, the FL classes need to host only a small number of children, so that everybody gets enough chances to participate, talk, ask questions, etc. All these and many others issues will be extensively discussed in the sections that follow, which focus on EFLL.
Chapter 4: The Age Factor and EFLL

4.5 EFLL: An introduction

The present thesis examines the impact of EFLL in an instructed context in Thessaloniki, Greece. The age range examined is early middle childhood, in particular the first two years of primary school. As Greece is one of the member states of the European Union (EU), the study will first need to be placed in a wider educational setting.

During the past decades, the major reason for learning a FL was to communicate with others, get a better job, or keep up with the pace of technology, business, and commerce. In today’s world, people are more mobile than ever before, in the search of a job that will ensure the well-being or, in extreme cases, the mere survival of the family. Therefore, it makes sense for children coming from such migrant families to have an early contact with the language and culture of the host country. In this sense, one could argue that EFLL is a byproduct of the temporary family mobility which, in turn, is the outcome of the globalisation of professional activity. For the countries of the EU, for instance, the residents have opportunities to live and work in another country. This requires both competence in the TL and adaptation to the TL community, culture and way of living. Gürsoy (2011) holds that EFLL and the consequent early contact with another culture can be facilitative as they both bring children to a better understanding and appreciation of differences, a faster development of multicultural understanding and friendship and, finally, to a smoother integration in the TL community. Ján Figel’, the European Commissioner for Education, Training, Culture and Multilingualism, at the presentation of the new Eurydice report that bridges data from 32 countries regarding FLL (www.eurydice.org), said back in 2005: “In an enlarged and multilingual Europe, learning foreign languages from a very young age allows us to discover other cultures and better prepare for occupational mobility” (as cited in Edelenbos, Johnstone, & Kubanek, 2006: 13).

EFLL is a live issue as research on age effects provided a strong impetus for early teaching. Presently, teaching FLs to young learners is gaining popularity all over the world,
with the age for learning a FL being reduced in many countries, driven by ‘the younger the better in the long run’ consensus view of Singleton (1995). This is a compromise view which supports that the sooner a child is exposed to the L2, the better the results in the long run while exposure to the FL is substantial.

Recently, the learning of modern FLs and levels of attainment in them have moved to the foreground in Europe with the strong support of the Council of Europe (2001) and the Commission of the European Communities (2003) (see Csapó & Nikolov, 2009). The European Commission, in particular the Division for Promoting Language Learning and Linguistic Diversity, with the Action Plan of 2004-2006, has expressed its intention to extend, consolidate and develop the early learning of one or more foreign or additional languages in each of the EU member states provided that “member states should move towards ensuring that foreign language learning at primary school and kindergarten is effective…” (as cited in Tragant, 2010).

4.6 EFL learning and teaching in the European context

The current European context strongly promotes an earlier start and the establishment and further development of a plurilingual citizenry. Since 2008, the EU sponsors a longitudinal research project called ELLiE: Early Language Learning in Europe, which is supported by the European Commission (EC) (Lifelong Learning Programme, Project n. 135632-LLP-2007-UK-KA1SCR). This has been set up in response to the rapid expansion of provision for EFLL which recently occurred in Europe and many other parts of the world.

The transnational project in question investigates the early introduction of FLs across seven European countries (England, Poland, Italy, Spain, Sweden, The Netherlands, and Croatia) where FL teaching begins from the age of 6/7. Its aim is to find out what can realistically be achieved with respect to EFLL in European state schools, where the FL is one of the many school subjects and only a limited amount of time is devoted to the FL.
CHAPTER 4: THE AGE FACTOR AND EFLL

instruction. The project follows the progress of around 1,200 children (from 6/7 years to 10/11 years) and is concerned with issues that pertain to: a) policy implementation, b) factors that contribute to an effective early LL, such as learners’ characteristics, attitudes and motivation, c) the learners’ FL progress over the years. The significance of the teacher’s role and the impact of digital media on EFLL success are also being investigated (Enever, 2010). It is hoped that this longitudinal research will provide sufficient data to strategically inform future policies and practices.

The first two interim reports of 2008 (children aged 7/8; 1\textsuperscript{st} year of study) and 2009 (children aged 9/10; 2\textsuperscript{nd} year of study) suggest that:

a) teacher qualifications and teacher development reveal a mixed pattern (Krikhaar & Lopriore, 2010), with FL teachers varying from generalists teaching the full curriculum to FL specialists,

b) there is substantial need for teacher’s further training, as limited pre- and in-service FL training is currently available,

c) the amount of class time allocated to FL teaching varies across counties (from 30 to 90 minutes per week), which has an immediate impact on the learners’ linguistic progress in their 2\textsuperscript{nd} year of study (Krikhaar & Lopriore, 2010),

d) school principals, teachers and parents are mainly positively oriented (Tragant, 2010),

e) children are initially highly motivated to learn a FL. By their 2\textsuperscript{nd} year of study some of the enthusiasm diminishes as the FL becomes ‘just another lesson’ in the school curriculum,

f) motivational levels also vary considerably as children mature, with boys generally exhibiting less enthusiasm for FLL,

g) children of this early age are already aware of their FL abilities and their progress, as these are compared to those of their peers (Mihaljevic Djigunovic & Lopriore, 2010),
h) a positive environment, access to a variety of materials and media, and active involvement in language activities advantageously contribute to children’s progress and motivation (Krikhaar & Lopriore, 2010),

i) the initial strong emphasis on oral development (songs, rhymes, games) in the 1st year has a positive effect on children’s vocabulary production and pronunciation. In terms of lexical knowledge, 7/8-year-olds produce more nouns than verbs and gradually develop the ability to produce meaningful chunks of language, such as greetings,

j) in their 2nd year learners increasingly move beyond individual words to produce more complex phrases. They exhibit a growing interest in knowledge about language, literacy and enjoy opportunities to build their literacy skills in the FL,

k) children show clear preferences with respect to the FL environment and classroom activities. They show an initial preference for game-like activities and gradually become interested in the layout of a traditional classroom, where desks are arranged in rows facing the front, as this facilitates a more effective FLL (Mihaljevic Djigunovic & Krevelj, 2009; Mihaljevic Djigunovic & Lopriore, 2010).

Table 1 below depicts the current situation in Europe, in relation to the compulsory starting age for FLL. The information was taken from Blondin et al. (1998), the ELLiE Policy Summary (October 2010), and Enever (2009):
Table 1: Compulsory start age for Foreign Language Learning: The European context

<table>
<thead>
<tr>
<th>5 years</th>
<th>6 years</th>
<th>7 years</th>
<th>8 years</th>
<th>9 years</th>
<th>10 years</th>
<th>11 years</th>
</tr>
</thead>
</table>
4.7 The benefits of bilingualism

The term ‘bilingual’ is used in a variety of ways by many different people. According to The Early Child Bilingualism project (2012), the term is used in a broad sense to refer to children who are being brought up with two languages either from birth or who are exposed to an additional language at a later stage, when they go to school or move to a different country.

In the first half of the 20th century and until the 70s, the learning of an additional language from an early age was seen as a curse rather than a blessing, as with the case of child SLA or Bilingualism. The idea of bilingualism was followed by an ‘emotional paranoia’ (Bialystok, 2001: 169) in the fear of a potential negative impact this might have on the linguistic and cognitive development of the child (for a review see Hakuta, 1986; Hakuta, Ferdman, & Diaz, 1987). Toukomaa and Skutnabb-Kangas (1977, as cited in Ekstrand, 1982: 137) believed that EFLL could be harmful and should be postponed until the firm establishment of the L1, i.e. around the age of 12 to 13. As already mentioned, Peal and Lambert (1962) were the first to show that bilinguals are not cognitively handicapped individuals. On the contrary, they enjoy a cognitive advantage, as research findings suggested that in a wide range of intelligence, non-verbal tests they often outperformed monolinguals.

Today, there is abundant research which suggests that bilingualism is an asset rather than a drawback. Bialystok (2001: 180, 2006) speaks of the ‘formidable advantage’ of bilinguals over monolinguals in cognitive processing. García Mayo (2003), García Lecumberri and Gallardo (2003) as well as Muñoz (2006) report findings from studies where English was more easily learnt as an L3 and developed in harmony with the other two languages (Catalan-Spanish and Basque-Spanish), while no L1 loss was attested. Likewise, Lasagabaster Herrarte (1998) reports a significant advantage shared by children attending a total immersion programme in the same context, namely the most intensive programme of all, namely Model D programme, which falls very close to balanced bilingualism. These
programmes were driven by the socio-political changes witnessed some decades ago which favoured the resurrection and survival of the Basque language. In the same vein, Nikolov and Mihaljevic Djigunovic (2011) report that, apart from the linguistic and socio-affective gains learners experience through the exposure to the new language, their L1 development is also positively affected as children develop and boost their metacognition and their learning strategies which, in turn, positively interact with their L1. Similar findings came from The Early Bird Project (Goorhuis-Brouwer & De Bot, 2010) where the learning of L2 English by L1 Dutch or non-Dutch students at the age of 4 went smoothly, as did L1 development. Interestingly so, lower language ability children were the ones who benefited the most from such an early start. Taylor and Lafayette (2010) also report the beneficial nature of EFLL programmes which was transferred to general school achievement (math, social studies and science tests) and the learners’ L1.

According to Bialystok (2006), Cenoz (2005), Johnstone (2009a, 2009b) and McLaughlin (1984), early bilingualism (i.e. the early mastery of more than one language) is closely linked to a number of linguistic and cognitive gains as it:

a) enables children to communicate in two (usually official) languages,

b) boosts the communicative sensitivity of bilingual children as they become expert at adjusting their language to meet the language practices of their interlocutors,

c) enhances children’s linguistic, social, cognitive and metacognitive development. As regards the latter, Mihaljevic Djigunovic and Krevelj (2009) report that by Grade 3 many children develop the ability to think about the processes of LL and teaching. They gradually adopt an analytic attitude towards language and cognitive tasks and are sensitised to the formal aspects of language,

d) helps children acquire some insight into the arbitrariness and complexity of language which is then generalised to other cognitive tasks,
e) further promotes the interdependence and ‘collaboration’ between the languages acquired, as the linguistic abilities acquired in one language can be transferred to the other,
f) improves access to literacy if the two writing systems are close,
g) develops a greater executive function, which is at the centre of intelligent thought. This is especially so in problem-solving, cognitive ability tests that require attention and inhibition control. This can be explained on the grounds that bilinguals constantly control attention between two active and competing language systems to achieve fluent communication in the language that is required every time. The bilingual experience is a source of practice that boosts those control processes and makes them available for other tasks (Martin-Rhee & Bialystok, 2008),
h) develops positive attitudes in children towards the learning of FLs in general.

The reverse picture emerges when bilinguals’ verbal abilities (vocabulary tests) or speed of responding are being tested (Bialystok, 1988, 2006; Bialystok & Feng, 2009; McLaughlin, 1984). McLaughlin (ibid) held that although the bilingual competence in both languages may subtly differ from the monolingual one, this is not detectable unless a very detailed linguistic analysis is conducted. In any case, the fact remains that bilinguals have two languages at their possession, which is an asset on its own (Cook, 2001).

A closer look at the two processes shows a number of commonly shared advantages. However, if an EFLL programme is to be effective and perhaps enjoy some of the bilingual advantage, certain minimal conditions need to be satisfied. The section that follows provides an extensive discussion of this issue.

4.8 EFLL and childhood bilingualism in a natural setting

It is important to distinguish between child SLA in a bilingual context and EFLL in an instructed context as these are different processes that offer different learning opportunities and lead to different outcomes (Muñoz, 2010). In the first case, the regular and continued
exposure to more than one language may begin from birth or any time between birth and the age of 2-4 years - the exact time frames are still debated (Meisel, 2008; Unsworth & Blom, 2010). Thus, two language systems develop simultaneously in the child and exposure to the TL(s) is done with or without formal instruction. In the case of SL formal schooling the setting followed is that of total immersion wherein equal amount of time is allocated to both languages in the curriculum. Finally, the TL is also the language spoken by the community.

Peal and Lambert (1962) were the first to discuss the bilingual advantage in a wide range of intelligence tests and aspects of school achievement. Albert and Obler (1978) held that bilinguals are cognitively more flexible, possess greater linguistic sensitivity, have better developed auditory language skills and a wider range of processing strategies than monolinguals. Since then, research findings show that even though bilinguals tend to have a smaller vocabulary in each language than monolinguals in their L1, they do understand better linguistic structure as they possess a better developed metalinguistic awareness and thus are better at learning a new language (for a review see Bialystok, 2001; Lasagabaster, 2001). It seems that bilinguals are also better at problem-solving than monolinguals because of their more advanced executive functions (Lasagabaster Herrarte, 1998). In particular, they are better at allocating attention to specific aspects of a display while they inhibit attention to salient but irrelevant and misleading aspects to the task at hand (Bialystok, 2001, 2006; Bialystok & Martin, 2004). Interestingly, Oberhofer (2008, as cited in Goorhuis-Brouwer & De Bot, 2010: 291) suggests that the advantages of bilinguals with respect to their advanced metalinguistic skills, may also be shared by children who start FLL later (during kindergarten) and have considerably less input in the L2. The fact that bilingualism was found to positively affect children’s cognition, marked the beginning of the immersion tradition in Canada and the US, that developed programmes that fully or partially immerse children into a SL or FL.
In the case of EFLL in an instructed context, the basic properties of L1 grammar are in place before regular FL exposure begins (Pinter, 2011). This, according to Guasti (2002), happens around the age of 4. Exposure to the FL is only limited to the school context which has immediate consequences on the quality and quantity of the input received. Usually very few hours per week are dedicated to the teaching of the FL, limited to 2-4 sessions of approximately 45 minutes each per week. It is also the case that FL learners have no communicative need to use the FL outside the classroom as this is not the language of the community. In addition, FL classes are typically conducted by non-native teachers, whose oral fluency may be limited (Muñoz, 2010). Due to all these facts, younger children in the FL context need longer time than younger beginners in the SL context (Nikolov & Mihaljevic Djigunovic, 2006). Muñoz (2010) estimates that 10 years of access to L2 input in a naturalistic language setting exceeds the 50,000 hours. Thus, it would take more than 200 years for a FL learner to receive this same amount of hours, in contexts where the FL would be taught for 4 hours per week. Then, by definition alone, it is clear that EFLL programmes cannot possibly aim at native or native-like proficiency in the FL but at the development of favourable attitudes in learners and in LL and the creation of successful users of the FL in later life (Nikolov & Mihaljevic Djigunovic, 2006).

The following sections discuss the advantages and disadvantages of both processes.

4.9 The advantages of an earlier start and the benefits of EFLL

EFLL programmes have spread dramatically throughout Europe since the 90s and are increasing at a very high rate globally (Edelenbos, Johnstone, & Kubanek, 2006). According to the Eyridice report (2012), English is by far the FL that is mostly taught at primary level in nearly all EU countries. In 2009-10, 73% of students in primary EU education were learning English while governments all over the world are introducing English language programmes as a compulsory subject earlier in the curriculum, during the primary or even preschool years.
Enever and Moon (2010) report the introduction of English as a compulsory school subject from Grade 5 in Japan (since 2011) and Grade 4 in Iceland (since 2007). Interestingly, this early introduction is generally followed by the parents’ enthusiastic and supportive attitudes towards their children learning a FL earlier than before, as this is associated with the future potential of a country to engage in international business, and thus a better future for its citizens.

As more and more learners start learning FLs at an earlier age in different educational contexts and under varying conditions, EFLL has become a significant research field in its own right and one of the key areas in Applied Linguistics and Language Pedagogy (Johnstone, 2009a). The interested reader should read Nikolov and Mihaljevic Djigunovic (2006, 2011). The early introduction of FLs in the present European context seems to be a fruitful enterprise, as issues like assessment measures and design of proper teaching materials also come into play (Hasselgreen, 2005; Jones & Saville, 2009). The unprecedented increase in the number of very young or young learners and their teachers, has also given rise to a plethora of empirical studies devoted to the field of EFLL, where the most frequently researched FL is English (Cenoz, 2005; Enever, Moon, & Raman, 2009; Johnstone, 2009a; Muñoz, 2006; Nikolov, 2000, 2009a, 2009b, 2009c, 2009d; Nikolov & Curtain, 2000).

Despite the enormous variety of primary-level modern FL programmes around the world (for a review see Nikolov & Curtain, 2000), the objectives of EFLL programmes are a combination of the following aims (Cenoz, 2005; Nikolov & Curtain, 2000; Nikolov & Mihaljevic Djigunovic, 2006; Pinter, 2011). More specifically, EFLL programmes aim to and in fact have been found to:

---

1 Within the European context, the pre-school children between the ages of 3 to 6 are referred to as very young learners while the school-aged children between the ages of 7 to 11 as young learners (Nikolov & Mihaljevic Djigunovic, 2011: 96).
a) “help develop the multilingual potential of every child by activating the language acquisition mechanisms that young children still possess” (Edelenbos, Johnstone, & Kubanek, 2006: 10),

b) prepare the ground for communicating in a FL by developing basic communication abilities, what Cummins (1981, 1983) called Basic Interpersonal Communications Skills (BICS). These focus on oral fluency and facilitate a socio-linguistically appropriate communication. Once BICS are established, Cognitive Academic Language Proficiency (CALP) will also develop. The latter is required for academic activities and includes linguistic knowledge and literacy. Mihaljevic Djigunovic (2010: 308) holds that “It is possible to suppose that longer L2 learning enables higher transfer of cognitive/academic language knowledge” between the L1 and the L2,

c) foster motivation for life-long LL by making initial LL experiences ‘fun’. This can have a positive effect on variables such as aptitude, motivation, learners’ attitudes, etc. that later on ensure good language proficiency (Curtain, 2000; Enever, 2009; Farkasova & Biskupicova, 2000; Harris & O’Leary, 2009; Mihaljevic Djigunovic, 2009; Mihaljevic Djigunovic & Vilke, 2000),

d) generate in children positive attitudes towards other cultures, values, and speakers of different languages and cultivate in them a respect for other ways of thinking and acting. This is of vital importance in today’s globalised world, as tolerance and favourable attitudes will thus be established in tomorrow’s citizens of Europe (Csizer & Kormos, 2009),

e) enhance children’s cognitive development, their aptitude, especially in the area of metalinguistic awareness and sensitivity to sound (Johnstone, 2009a). If learning an additional language raises the metalinguistic awareness in learners, this will in turn help them to be more aware of their L1 skills as well LL process in general (Edelenbos,
Johnstone, & Kubanek, 2006; Nikolov & Mihaljevic Djigunovic, 2006). Mihaljevic Djigunovic (2010) raises the issue of bi-directionality by reporting the outperformance of early beginners (before the age of 10) on all L1 and L2 tests over later beginners. The author holds that a longer L2 exposure “may lead to language use behaviours (at the levels of language reception and production) that can be easily transferable” (ibid: 313). It is because of the longer L2 experience that early beginners reach the necessary competence levels in the two languages earlier to allow transfer of these behaviours from one language to the other.

As already mentioned, there is much variation attested within the EU with respect to EFLL programs. The differences pertain to the exact starting age, the amount of time devoted to the FL teaching per week, the type of curriculum used, the specialisation or not of the FL teachers, the amount of continuing professional teacher training and support (Johnstone, 2009a).

4.10 The minimal conditions that make an EFLL programme effective

It is because of this variation attested that a growing body of researchers voice the following word of caution: an earlier start and a longer exposure to the L2 do not automatically guarantee L2 success (Blondin et al., 1998; DeKeyser, 2012; Edelenbos, Johnstone, & Kubanek, 2006; Johnstone, 2009a; McLaughlin, 1993; Nikolov & Curtain, 2000; Nikolov & Mihaljevic Djigunovic, 2006; TESOL, 2009). What an earlier start simply offers is a longer overall period of learning and the potential to influence children’s personal development when they are still at a highly developmental stage. On its own, an early start is unlikely to make a substantial difference in the children’s learning progress.

Johnstone (2009a) holds that an early start gives learners the opportunity to integrate and further develop different sorts of learning experience, from the natural to the more analytical way of learning. On the one hand, starting early is advantageous as younger
children are more motivated, less likely to be ‘language anxious’, acquire the FL sound system with a relative ease, while their cognitive, linguistic, emotional, social, and intercultural awareness and skills are still in the process of formation. Through a prolonged period of FL exposure, younger learners are allowed an overall longer time for FLL. On the other hand, older beginners have an advanced and more sophisticated conceptualisation of the world and only need to learn the unknown new word and match this with an existing concept. They are more experienced in handling conversations and negotiating meaning and have already developed a wide range of learning strategies (i.e., rehearsal, note-taking, revising, etc.). Johnstone (ibid) summarises this in what follows:

Overall, an advantage of an early start is that in principle at least it allows young beginners to exploit such advantages as they possess, but in addition, as they become older, to make use of the advantages that older learners possess. So, over time, both sets of advantages are available to those making the early start, whereas only the second set of advantages is available to those beginning later. To my mind, this is a compelling argument in favour of ELL, provided that the key conditions for its success can be put in place. (p. 34)

Thus, the issue of EFLL is multidimensional and the success and effectiveness of such a programme heavily depends on finding the right balance among a variety of factors (Csapó & Nikolov, 2009; Hasselgreen, 2005). First, realistic objectives and expectations need to be clearly set (McLaughlin, 1993). This is crucial as an initial sense of failure may have damaging consequences for learners’ L2 school career in the long run. As already mentioned, the ultimate goal in EFLL is not native-likeness but, rather, the development of children who are able to use language in a functional way. This presupposes a good listening attitude and a sense of the functions of the spoken and the written language. Van den Branden, Van den Nulft and Verhallen (2009: 13) hold: “Whatever starting point one chooses, one must end up
Chapter 4: The Age Factor and EFLL

with a child who is able to perform with linguistic adequacy in situations that are relevant and important to him/her .” EFLL goes beyond the development of NS ‘clones’ and is instead interested in the development of ‘competent’ (Preston, 1981), ‘effective bilinguals or multicompetent’ users of the FL (Cook, 1995: 55), who develop favourable attitudes towards different languages and LL in general. As children are not expected to achieve native L2 level in school, “achievement targets tend to be modest and different levels may be required in the four skills” (Nikolov & Mihaljevic Djigunovic, 2006: 241).

Also, learner-centered approaches need to be adopted for the initial intrinsic motivation of young learners that is based on enjoyment, interest, and curiosity which will further develop to some kind of self-awareness. In this way, children will become motivated and will take pleasure out of the realisation they are gradually becoming successful learners of an additional language (Johnstone, 2009a).

In addition, the language input needs to be child-friendly, one that takes into account the learners’ age, their abilities and communicative needs. The teaching material should be designed in a way to highlight the learners’ strengths. Activities need to be captivating, motivating and cognitively challenging, always tuned to their age and level (Enever & Moon, 2010; Hasselgreen, 2005; Nikolov & Mihaljevic Djigunovic, 2006; Wong Kwok Shing, 2006). While the widespread view is that EFLL should initially concentrate on listening and speaking, the example of the Croatian Project of EFLL in 1991 (where children started at 6/7), reported by Johnstone (2009a, 2009b), suggests that early literacy practices (from the age of 7/8) can actually accelerate the learning process, as children are helped in thinking analytically, strategically, and in monitoring their own learning (Pinter, 2011). Also, natural content-based conditions should be taken into account. Mihaljevic Djigunovic and Vilke (2000) report immediate benefits for younger children when the quality of input and the tests themselves are more naturalistic. Younger beginners are significantly better at pronunciation,
orthography and vocabulary than older ones and develop a number of learning and communication strategies.

The benefit of an early start can only be maintained with the systematic continuity of FL teaching from one educational cycle to another (Edelenbos, Johnstone, & Kubanek, 2006). To this end, secondary school teachers need to build on EFL teaching in effective and motivating ways. In this way, experiences gained over the primary school years will be known, accepted and further developed in the secondary school years to come. Unfortunately so, continuity is not always ensured and programs often fail to integrate what children already know and are good at (Nikolov, 2000; Nikolov & Mihaljevic Djigunovic, 2006).

The teacher, being a role model, plays a very important role in students’ success and their attitudes towards school (Heining-Boynton & Haitema, 2007; Nikolov & Mihaljevic Djigunovic, 2006). According to Sahinkarakas (2011: 884), “Teachers shape children just as bakers shape dough”. During the early stages children are forming their learning experiences and are highly dependent on their teacher who provides exposure to the language and opportunities for learning through classroom activities. Thus, teachers have a duty and a responsibility to develop the language skills of their pupils to the maximum. Because they represent powerful resources of influence (Pinter, 2011), teachers need to be highly-skilled and appropriately trained, with pre- and in-service training courses in teaching EFL to very young learners (Enever & Moon, 2010). Continued specialised training will strengthen their teaching competency and increase their motivation in implementing the Early Language Learning (ELL) policy (Edelenbos, Johnstone, & Kubanek, 2006; Johnstone, 2009a). This is important as they are responsible for motivating pupils over longer periods of LL (Cameron, 2003). Nikolov and Mihaljevic Djigunovic (2011: 112) hold that “Preparing enough teachers who are motivated and able to work efficiently with YLs in good quality programs is the way to move forward”. Through in-service training teachers also ameliorate their English oral
proficiency. This is very important if one thinks that the majority of FL teachers are NNSs and also that children tend to reproduce the accent of their teacher with ‘deadly accuracy’ (Cameron, 2003: 111).

The EFLL context provides limited exposure to learners. Thus, the lowering of the starting age of learning needs to be accompanied by an increase of the level of intensity of instruction and the total amount of time invested on the teaching of FLs. In this way, children will receive enough contact hours for a ‘catching up’ to be completed. Singleton and Ryan (2004) compared the amount of exposure in naturalistic and formal LL contexts and calculated that it would take approximately 18 years, if at all, for EFL learners to catch up with older beginners. Thus, if more time is not allotted to FL teaching, the early start advantage will yield very modest proficiency outcomes (Blondin et al., 1998; Johnstone, 2009b; Larson-Hall, 2008; Nikolov & Mihaljevic Djigunovic, 2006). Johnstone (2009a) views that maximising the key variables of time and intensity through CLIL and immersion programmes would lead to more impressive linguistic and content outcomes. This is so because in these programmes learners need to learn not only the additional (foreign) language but also important material of another subject through the medium of the FL. Interestingly enough, Johnstone (ibid) concludes, many children cope well with this additional challenge. The author holds that immersion programs produce fluent and confident L2 users with impressive levels of listening comprehension, without this happening at the expense of the subject-knowledge acquired through the immersion language. In a similar vein, Mihaljevic Djigunovic (2009) and Nikolov and Mihaljevic Djigunovic (2011: 112) hold that “the integration of content and language at an early age may help learners progress” as it makes learning immediately more meaningful and purposeful for the child.

In addition, the learning of the FL should be conducted in small size classes to make sure that all students have ample opportunities to participate in classroom activities,
experiment and further develop their ILs (Edelenbos, Johnstone, & Kubanek, 2006; Nikolov & Curtain, 2000).

Finally, the wider community such as school principals, parents and the rest of the teachers all need to support the earlier start of FLL. Most importantly, though, the Ministries of Education need to cater for an adequate supply of resources (adequate audio-visual equipment, proper teaching materials, etc.) so that teachers can do their work well and in a productive manner.

When these basic conditions are met, early beginners have the potential to outperform later ones (Mihaljevic Djigunovic & Vilke, 2000). It should also be stressed that if the decision for an earlier start is taken without the necessary planning and piloting phase to prepare the ground for the integration of modern languages on a larger scale, there is the danger of children making virtually no progress (Enever & Moon, 2010; Johnstone, 2009a).

If these conditions are not satisfied, the EFLL enterprise ‘may lose momentum’ (Johnstone, 2009a: 33). The initial heightened motivation of early learners (Nikolov, 2000) and their eagerness to learn new things and words will slowly fade away and they may become bored or disillusioned after a few years of L2 learning if conditions are not the ones previously described. As a consequence, learners will develop negative attitudes and the L2 will become ‘just another school subject’ (Cenoz, 2003; Heining-Boynton & Haitema, 2007; Muñoz, 2003; Nikolov & Mihaljevic Djigunovic, 2011; TESOL, 2009). This goes contrary to the EFLL goals originally set and should be avoided by all means. It is therefore crucial to maintain and develop further both the learners’ proficiency and willingness to learn. Mihaljevic Djigunovic and Krevelj (2009) also suggested that direct contact with NSs or other L2-speaking people may contribute to the maintenance of the learners’ initial positive attitudes towards L2 learning.
4.11 EFLL in an instructed context

The advantages of EFLL have not attracted the same degree of attention when compared to the investigation conducted so far relative to the influence of age on L2 acquisition. The educational decisions taken concerning the optimum time for introducing FLL in schools have been influenced by findings obtained in naturalistic LL settings which indicate the advantages of an early start. If the value of EFLL is to be proved, then the SLA field will be in need of longitudinal studies which will follow up the effects of such programmes. To my knowledge, such kind of studies is only rare, if not scarce. Still, some attempts have been made in the past that measured the linguistic outcome of such an initiative. We will mention only a few.

In the 60s, when French was introduced in primary schools in England and Wales, the Burstall report evaluated the whole project (Burstall et al., 1974, as cited in Pinter, 2011: 56 and Larson-Hall, 2008: 39-40). Some interesting findings emerged for the younger learners, in relation to listening and speaking skills and overall positive attitudes were reported towards the learning of French. However, the initial advantage of younger learners disappeared in the course of five years. Thus, the programme was negatively evaluated, the early introduction of FLs was not considered a worthwhile activity and the programme’s funding was cut. However, as Muñoz (2010) and Larson-Hall (2008) argue, one of the drawbacks of this project was that groups of learners with different initial age of learning and proficiency level were mixed up in the same classroom.

In 1995 the BAF project was put into action (Muñoz, 2006) which explored, cross-sectionally, the age effects on FLL with respect to different language abilities. Children started to learn English as an L3 (the other two languages being Catalan-Spanish or Basque-Spanish) for 2.5 hours on average per week at two different ages: the age of 8 and 11. They were tested after 200, 416 and 726 hours of English instruction. The battery of tests included a cloze test, a dictation, a grammar test, a listening comprehension test and a written
CHAPTER 4: THE AGE FACTOR AND EFLL

composition. Also, the learners underwent a series of oral tests that included an oral interview, a picture-elicited narrative, a minimal pair discrimination test and a word imitation test, while learners also performed in a pair-wise manner a role-play and a map task. Overall, late beginners were superior to younger ones, they were faster (higher scores achieved after the first 200 hours of teaching) and more efficient, due to their superior cognitive development that allowed them to take greater advantage of explicit teaching processes in the classroom (Muñoz, 2006). On the other hand, a long-term superiority of early beginners was not confirmed. At most, younger learners were more motivated than older ones (Cenoz, 2003) and showed the highest rate of learning in the last third of the followed period (i.e. between 416 and 726 hours of instruction) in less cognitively-demanding tasks, such as the listening comprehension test or the fluency measurements (Muñoz, 2006). This observation led Muñoz (2010) to predict that younger learners would benefit from an early exposure only with significant exposure to the FL that would be intensively distributed along the years of FL schooling, during which learners should have ample opportunities to actively participate in a variety of L2 social contexts. Only under such favourable conditions would their implicit learning mechanisms be stimulated faster and FLL proceed in the same way as L1 learning.

García Lecumberri and Gallardo (2003) as well as García Mayo (2003) found that older beginners outperformed younger ones on measures of phonetic perception and production, fluency, accuracy and complexity. However, their initial superiority was attributed to factors other than age itself, such as instructional practices, the teachers’ fluency and accuracy, the learners’ motivation, their cognitive maturity and their explicit learning mechanisms that were at work (Lasagabaster & Doiz, 2003; Muñoz, 2006; Pinter, 2011). The findings from the BAF project, although at first sight suggest an overall superiority of the later beginners, they simultaneously do imply that if certain conditions are carefully controlled, younger beginners
can actually make use of their inherent implicit LL mechanism. This idea is not far from what has been discussed earlier in this thesis.

The Croatian project of EFLL (Mihaljevic Djigunovic & Vilke, 2000) followed the participants’ FL (English, French, German, and Italian) development over the course of eight years, starting from Grade 1. It set out to find the optimal starting age at which the FL should be introduced in the primary school curriculum. The studies carried out during the eight years showed an overall faster development of 6/7-year-olds over the older beginners (10-year-olds), with respect to pronunciation, orthography, vocabulary, reading, learning and communication strategies, and implicit learning (for a detailed discussion see Nikolov & Mihaljevic Djigunovic, 2006).

In the same vein, Larson-Hall (2008) examines the long-run linguistic consequences of minimal input on young Japanese learners of EFL of different ages (9 vs 12-13 years of age), who had been taught English for less than 4 hours per week. The research findings on the grammaticality judgement task and the phonemic task suggested that even in such low-input situations there can be perceivable age effects on linguistic measures that relate to phonology and morphosyntax, which, however, emerge only after a substantial amount of input has been gained.

4.12 Chapter Summary

SLA research findings demonstrate that FLL is a complicated process which accommodates a number of IDs (cognitive, affective, social, etc.) that determine the eventual outcome.

For almost three decades (from the late 50s until the late 80s), early SLA research, based on biological and maturational accounts, witnessed a fierce debate of whether there is one critical period or multiple sensitive periods, as regards L2 learning. The evidence was abundant but contradictory, acknowledging in some cases the overall superiority of children and in others, that of adults. The findings suggested that young children display superior
Chapter 4: The Age Factor and EFLL

pronunciation skills while older L2 learners exhibit more advanced morpho-syntactic ones. This was explained on the grounds of a shorter maturational cycle of Phonology as opposed to those of Morphology and Syntax. Also, the short-term superiority of adult FL learners was attributed to their more efficient cognitive processing. Nevertheless, all these studies were predominantly conducted in natural settings and not in instructed ones.

Since the 90s, the interest of SLA research has focused on EFLL within a school context, with the age for learning a FL being reduced in many countries to the 6 years of age. This tendency was driven by Singleton’s (1995) ‘the younger the better in the long run’ view, according to which the sooner a child is exposed to the L2, the better the long-term results, due to the substantial and massive exposure to the FL that is expected to take place in the mean time.

As of 2001, the EU strongly promotes and supports an early start (even from the kindergarten years) to establish and further develop a plurilingual citizenry. Projects such as ELLiE have been launched, seeking to find out the realistic achievements of EFLL in European state schools. The findings demonstrate an overall positive attitude of teachers and parents, an initial heightened motivation of the learners, and a mixed pattern regarding teacher qualifications and the time devoted to FL teaching across Europe.

Projects like the ones that were conducted in Croatia, The Early Bird Project or the BAF one, demonstrate that there is no L1 loss when a FL (English) is learnt as an L2 or an L3. In fact, FLL is in harmony with L1 development, which seems to be positively affected by the EFLL experience, as it helps children develop and boost their metacognition and learning strategies which, in turn, help them become more aware of their L1 skills as well of the LL process in general. Nevertheless, if the value of EFLL is to be proved, more longitudinal studies must be conducted to follow up and record the positive effects of such EFLL programmes.
The present thesis wishes to explore the cognitive impact of EFLL. To the best of my knowledge, and from what has been reported so far, no study has ever examined this aspect of EFLL. If this thesis proves that EFLL exercises a positive cognitive impact on young learners’ cognitive skills (i.e. on WM, PSTM, and FL aptitude), this would be the study’s significant contribution to the discussion of EFLL, as cognitive resources determine one’s performance in general LL (see Chapter 6).
CHAPTER 5
FOREIGN LANGUAGE APTITUDE

5.1 Definition of the term

Carroll (1981), in his pioneering work on aptitude, viewed this as speed of learning a FL in some sort of formal context, which, according to Sawyer and Ranta (2001), can be either a language course or a program of self-study. He employed the general term *Foreign Language Aptitude* (FL aptitude hereafter) to refer to the learners’ initial state of readiness and capacity for learning a FL, and ‘probable degree of facility in doing so’ (p. 86), given that they have ample opportunities during learning. Thus, individuals may differ as to their capability of learning FLs, in the sense of rate or ease of learning “under optimal conditions of motivation, opportunity to learn, and quality of instruction” (Carroll, 1973, as cited in Dörnyei, 2005: 43). He considered FL aptitude to be an innate but not language-specific capacity, a set of cognitive qualities, largely independent of personality traits like motivation, attitudes, or general intelligence. According to Carroll, this ability is fairly constant and ‘relatively enduring’ (1985: 84-5), ‘hard to modify in any significant way’ (p. 86), which may improve through special instruction or training. He saw no relationship between FL aptitude and proficiency, at least not in the beginning of a language programme (1985). Similarly, Gardner and Lambert (1972, as cited in Carroll, 1981: 86), believed that this is dissociated from motivation.

Sladen (1981) provided genetic evidence for the biological basis of IDs in FL aptitude that came mainly from dyslexic populations, where the males outnumbered the females. She spoke of a cognitive disorder that is genetically influenced and can be manifested in one’s difficulty in learning to read despite conventional instruction, adequate intelligence, and sociocultural opportunities. Thus, dyslexic children have an extended period of learning compared with the normal population and face difficulties with reading and spelling. This is
so because they have a poor recall of symbols and a poor memory for sequences, due to a weak or incomplete lateral dominance. Although in normal populations speech is interpreted dominantly in the left hemisphere and spatial and temporal sequencing in the right one, this may not be the case with dyslexics.

FL aptitude has been viewed from a number of perspectives and defined in a number of ways. The central claim in the relevant research is that aptitude is a specific talent for learning FLs which exhibits considerable variation among learners. The following are only some of the most prevalent views concerning the construct:

a) It is as an innate predisposition, a special propensity for learning a SL/FL, the ‘picking up’ of languages (Abrahamsson & Hyltenstam, 2008; Ellis, 1985, 2003: 494).

b) It is conceived as a relatively stable talent for learning a FL that differs considerably between individuals (Dörneyi & Skehan, 2003). McLaughlin (1990, as cited in Harley & Hart, 1997: 381) did not share this view as he believed this to be a dynamic construct, amenable to training as novices can become experts with experience, i.e. by learning appropriate strategies. Neufeld (1978: 16) followed the same line of thought, viewing FL aptitude as some kind of ‘language learning readiness’, an ‘inclination’ to learn an L2. He claimed it was not something fixed and depended largely on one’s learning experiences prior to starting any FL study. Cook (2001: 124) defines aptitude as ‘the ability to learn from teaching’, as he notes, this term has almost invariably been used in connection with students in classrooms.

c) FL aptitude may vary significantly among individuals, with respect to rate of acquisition and ultimate attainment (Sawyer & Ranta, 2001). Children, despite their origin and environment, normally learn their native language (L1) unconsciously at a highly predictable rate. However, Neufeld (1978) held that the disparity attested in school-age children learning an L2 is due to their psychological diversity and the social context
wherein they acquire the new language that has nothing to do with FL aptitude. Kiss and Nikolov (2005) also believe that LL aptitude is responsible for a considerable portion of the variance in L2 achievement, with individuals displaying different relative strengths and weaknesses on the various component abilities.

d) According to Ellis (2003), FL aptitude is not a prerequisite for L2 acquisition but a capacity that enhances both the rate and the ease of learning. In fact, he considers it to be the best single predictor of SLA achievement in the long run. On the other hand, Dörnyei (2005, p. 31) cites Ehrman and Oxford (1995) who found that aptitude was the ID variable that most strongly correlated with L2 proficiency. Krashen (1985) defined aptitude as the speed of conscious learning and an overall ability to acquire SLs. He thought this was very important in formal situations such as classrooms, whereas in informal real-world situations attitude was the decisive factor. DeKeyser (2000), in his discussion of the CPH, concludes that language aptitude is not a necessary condition for the child learner, but one that puts the child at an advantageous position for attaining a (near) native-like L2 level of proficiency. Bialystok (1997) dismissed the connection between aptitude and L2 near-nativeness and considered this to be entirely irrelevant and further claimed that the second occurs when circumstances (social, attitudinal, input related, educational, etc.) are favourable.

e) Skehan (1990, as cited in Sawyer & Ranta, 2001: 331), in an investigation of early childhood (15 months to 60 months) concluded that:

… aptitude is a product of two separate groups of influences: one which reflects an innate capacity for learning that is a “residue” of L1 development, and the other which reflects the development of the ability to abstract from experience to handle decontextualised material, which is influenced by family background factors such as parental literacy.
f) FL aptitude is a multi-component notion and consists of several independent cognitive abilities that come into play in LL (Larsen-Freeman & Long, 1991). Gardner and McIntyre (1992, as cited in Alexiou, 2005: 216) viewed aptitude as a kind of cognitive ‘sponge’: new skills or knowledge are naturally absorbed and associate to existing ones.

g) Ellis (1985) reported that aptitude was not an easy term to define because it was not clear what cognitive processes are subsumed under its label. He also believed that FL aptitude did not influence the route but rather the success and the rate of development in SLA, particularly in the conditions of Krashen’s ‘learning’ and Cummins’ (1979) Cognitive/Academic Language Proficiency (CALP).

h) Johnson (2001) believes that aptitude is quite distinct from achievement. An aptitude test has a strong predictive element to it. It looks at how well one will do, as opposed to achievement and proficiency tests that measure how well one has done.

i) Robinson (2005) views FL aptitude from an interactionist perspective that focuses on the Information Processing (IP) demands of L2 acquisition. He defines this as the strengths of individual learners in the cognitive abilities the IP paradigm draws on during FL performance in various contexts (i.e. naturalistic, instructional) and at different stages. He suggests that aptitude should be conceptualised as a complex and dynamic construct where cognitive resources and primary abilities combine into high-order abilities that are directly involved in various learning tasks. These high-order abilities are then grouped into aptitude complexes that play different roles under different learning conditions (incidental, implicit and explicit learning).

### 5.2 FL aptitude research: A brief historical account

Carroll (1981: 87) reported that intimations of a concept of FL aptitude appeared as early as 1575, in Huarte’s *Exámen de Ingenios (Examination of the Aptitudes for the Sciences)*. However, it was not until the 20th century that the first attempts were made to devise FL
aptitude tests. He cited that in the first half of the century the first aptitude test batteries were developed, like the *Luria-Orleans Modern Language Prognosis Test* (Luria & Orleans, 1928), the *Foreign Language Prognosis Test* (Symonds, 1930), the *Iowa Foreign Language Aptitude Test* (Stoddard & Van der Beke, 1925). However, as Skehan (1989) reported, they all had their drawbacks as the correlations with achievement scores were not impressive, while the tests relied heavily on a grammar-translation methodology.

Before and during World War II, there was a shift of emphasis from grammar-translation methods that provided explicit knowledge through rule explanation and guaranteed the simultaneous learning through all four skills, to the audio-lingual ones which focused on inductive rule learning through listening and extensive oral practice. These tests followed the paper-and-pencil methods of test administration, common in the 50s through the 70s.

The effort to devise new measures (see Reed & Stansfield, 2004), conducted for the US Army by Dorcus, Mount and Jones (1952, as cited in Carroll, 1981: 88) proved to be quite unsuccessful. In 1957, the Army developed *The Army Language Aptitude Test* (ALAT), a 30-minute, artificial language test. Soon, this was adapted and became *The Defense Language Aptitude Test* (DLAT), used for screening candidates for the Defense Language Institute (DLI).

Towards the end of the 50s, Carroll (1981, 1990) realised the need for an aptitude measure that would predict success - starting ‘from scratch’. This would be very useful in cases like intensive language courses conducted in schools, colleges and governmental or military organisations, to reduce the number of failures in language training and thus minimise the costs of training programs. Together with Sapon (1959), they devised *The Modern Language Aptitude Test* (MLAT). The battery was initially named *Psi-Lamda* FL battery (‘psi-lamda’ standing for psycholinguistic). The MLAT received wide acceptance in the Foreign Service Institute at the US, Australia (Hanna, 1968, as cited in Carroll, 1981: 91)
and the Canadian Public Service Commission (Wesche, 1981). It was originally designed for ages 14 and above.

According to Carroll’s (1981) model, FL aptitude is a multi-componential ability with at least four basic capacities involved (Dörneyi & Skehan, 2003):

1. **Phonemic Coding Ability**: Singleton and Ryan (2004: 216) cite the words of Carroll (1977) who defined this as “… the ability to learn L2 sounds or words, to identify them as distinctive, and then to store them in memory so that they can later be recalled accurately on an appropriate occasion”. This ability is important because “individuals lower in this ability will have trouble not only in remembering phonetic material, words, forms, etc., but also in mimicking speech sounds” (Carroll, 1965, as cited in Johnson, 2001: 124).

2. **Grammatical sensitivity** is the capacity to recognise the grammatical functions words fulfill in sentences. The principal memory involved is LTM, according to Carroll’s classification (1981: 111-2), which he believed was highly speculative, as all the aptitude components might be highly complex from an IP viewpoint.

3. **Inductive language learning ability** is the ability to infer rules, patterns from previously unknown language materials to create new sentences. Again, according to Carroll (p. 109-10), the principal memory involved is ‘intermediate-term memory’ (ITM).

4. **Associative memory** is the capacity to form links in memory. Wesche (1981) noted that low scores on Parts 1 and 4 may indicate problems with attentiveness and concentration.

Carroll (1981) also reported various versions of the MLAT, devised to be taken by the blind, by learners whose L1 was not English, such as Italian, French, Japanese, Spanish, German, Thai, and lately Hungarian (Sawyer & Ranta, 2001).

Closely related to the MLAT is the *Pimsleur's Language Aptitude Battery* (PLAB, 1966). This was developed to be administered to junior high-school children, aged 13 to 19 (Skehan, 1989). Pimsleur claimed that a) there is a relationship between language aptitude
and IQ, b) language aptitude includes motivation, verbal intelligence and auditory ability (as cited in Wesche, 1981: 129). He believed that his battery had the potential to identify students’ strong points and areas of difficulty, so that instruction could better meet individual needs. Skehan (1989, 2002) reports that the test emphasised inductive LL capacities and the auditory factor at the expense of memory.

In 1967, Carroll and Sapon developed a version of the MLAT (MLAT-Elementary) for children in grades 3 to 6 (i.e. for 8-11 years of age). This consisted of 4 sub-tests: a) hidden words, b) matching words, d) finding rhymes, d) number learning (as cited in Kiss & Nikolov, 2005: 105).

In 1975, Green developed The York Language Aptitude Test, which tested inductive LL and memory for school-age children in the UK (as cited in Skehan, 2002: 74).

The Defense Language Aptitude Battery (DLAB), devised by Petersen and Al-Haik (1976, as cited in Carroll, 1981: 94-5), was designed to detect in what language or language family a candidate might do best. The DLAB involved auditory and visual materials, and the learning rules of an artificial language. Very soon they reported that the test failed to permit any differential prediction of success in different languages or types of language courses. Kiss and Nikolov (2005) as well as Skehan (2002) report that the battery was designed to discriminate better at the top end of the ability range, i.e. high aptitude students.

In 1990 Parry and Child developed The VORD, a test of an artificial language that resembled Turkish, which focused primarily on grammatical analysis. They suggested the test might be useful for indicating which channel of communication (the eye or the ear) makes an essential difference in FLL. The test sought to expand the concept of FL aptitude by including a variety of new predictors such as measures of general ability, learning styles and strategies, attitudinal and motivational characteristics. When they piloted it, they admitted that more research would be needed, as the test proved to be less successful in predicting LL success.
than the MLAT (Ellis, 2003; Skehan, 2002; Sparks & Ganschow, 2001). Sawyer and Ranta (2001), report that neither the DLAB, nor the VORD demonstrated superiority over the MLAT.

Another attempt was taken by Grigorenko, Sternberg and Ehrman (2000) with *The Canal-F battery (Cognitive Ability for Novelty in Acquisition of Language (Foreign).* It was for use in government context only. Their claim was that the central ability required for FL acquisition is the ability to cope with novelty and ambiguity (as cited in Wen, 2012: 234). They held that, contrary to previous conceptualisations of FL aptitude as being an undifferentiated construct largely fixed at birth, aptitude can be developed throughout a person’s life like any other form of expertise (as cited in Kiss & Nikolov, 2005: 107).

More recent are *The Language Aptitude Test* (LAT) of Meara, Milton and Lorenzo-Dus (2003), which is a computerised set of aptitude test for adults, and *The Llama Language Aptitude Test* (LLAMA) (Meara, 2005), a more refined version of the LAT.

In 2005, Kiss and Nikolov developed the *MENYET* for 12-year-old Hungarian learners of English. The authors hold that cognitive factors such as aptitude, memory and attention can be hypothesised to contribute to young learners’ achievements. The battery has four sub-tests: a) hidden sounds, b) language analysis, c) words in sentences, d) vocabulary learning. They report that language aptitude explained 20% of the variation in English language performance, with motivation contributing another 8%.

The most recent aptitude test which is suitable for the age range examined in this thesis is *The Young Learners’ Aptitude Test* (YLAT) of Alexiou (2005), designed for 5- to 9-year-old learners. Its orientation is primarily cognitive in nature, as children as young as these do not yet possess sophisticated language skills. Her results suggest that:

a) language aptitude can actually be tested as early as the age of 5 in different cognitive skills,
b) there is a strong relation between cognitive development and FLL,
c) phonological skills play an important role to FL achievement,
d) most cognitive skills improve with age: 5-year-olds outperform 6-year-olds in memory and analytic tasks but do less well than older children in the same skills. 7-year-olds have an advantage in memory and phonological skills. 8-year-olds do better in analytic tasks. Children are not as memory-dependent (memorisers) as previously believed (Harley & Hart, 1997). Their intellectual and cognitive development is much quicker than was assumed, as very young learners have primary analytic skills at their disposal from the age of 5. Children as young as 5 or 6 are not ‘tabula rasa’ as previously thought,
e) girls on the whole do better than boys, both in cognitive development and language scores. Boys are slightly better in phonetic recall and in more analytic tasks (spatial ability and inductive learning ability). These differences, however, did not reach statistical significance. Vecchi, Phillips and Cornoldi (2001, as cited in Alexiou, 2005: 7) found boys to be better in visuo-spatial tasks and girls in verbal memory tasks,
f) language aptitude is not fixed and unchangeable. On the contrary, it seems to be quite plastic, at least at this early age,
g) paired associative memory impacts strongly on language achievement,
h) memory, analytic skills and phonetic ability determine language achievement.

5.3 Carroll reconsiders the concept of FL aptitude

By 1990, Carroll realised that since the initial development of the MLAT, a much heavier pedagogic emphasis was placed on the building of learners’ communicative ability and success in using the language. He acknowledged that in the 30 years that had passed since the test was originally designed, considerable research had taken place in cognitive psychology within the IP framework and observed that further research in the construct of language aptitude “would require a more refined analysis of foreign language learning tasks in terms of the different cognitive abilities they call upon” (ibid: 26).
CHAPTER 5: FOREIGN LANGUAGE APTITUDE

Carroll (1990) had realised that different language learning (LL) tasks may require different sets, or combinations of abilities to be successfully performed. Also, he was aware of the fact that the MLAT was more effective at predicting success not in the early language learning stages but at the more advanced ones. His acknowledgement indirectly suggested a revision of his original view of aptitude, which was stable and not dynamic in nature. Robinson (2005), in line with Carroll’s later view of FL aptitude, believes that the contributing abilities of FL aptitude constantly reconfigure as learners reach higher levels and shift to different contexts of practice and exposure.

5.4 Criticism on the construct of FL aptitude

With all its shortcomings, the MLAT became a model that has guided almost all ensuing aptitude research. Nevertheless, it has been criticised for being used as a diagnostic tool to predict success in instructed learning only. As a consequence, FL aptitude research focused heavily on FLL in formal settings (Krashen, 1981a) and very rarely on SLA in naturalistic settings (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Harley & Hart, 2002; Robinson, 1995a). Skehan (1989) criticised the MLAT for failing to discriminate adequately at higher levels. Also, he believed that aptitude tests were not “transparent measures of the underlying constructs” (ibid: 44). Their scope was limited as they had only some predictive value but no explanatory function. Sparks and Ganschow (2001) hold that if aptitude research is to be effective and with positive learning outcomes, it should seek the development of new norms for younger age levels and should also incorporate phonological tasks such as pseudo-word ones.

With all this skepticism expressed, it is of no surprise why the concept of FL aptitude research gradually faded to become a marginal activity in the wider area of FLL by the 70s. In the meantime, i.e. from the late 60s (for a review see Sparks & Ganschow, 2001), a productive and fruitful research on individual differences and affective variables was being
conducted, as they were all considered important factors that influence the rate and success of FLL. In particular, it was thought that individual characteristics such as intelligence, attitude, motivation, LL strategies, cognitive styles or personality traits might influence learners’ perceptions and impressions of the LL context (Clément & Gardner, 2001; Dörnyei, 1998; for a review see Gardner, 1990; O’Malley & Chamot, 1990; Oxford, 1989; Oxford & Anderson, 1995, as cited in Dörneyi & Skehan, 2003: 605; Skehan, 1989, 1998).

During this time, Skehan (2002: 69) reports the existence of a rather long period of “remarkably little theorizing” and “relatively little empirical work” on FL aptitude research. Several were the reasons for its unpopularity:

a) The aptitude construct was considered to be of an undemocratic nature with respect to learners: if the fixed and immutable interpretation of aptitude was to be accepted, then students with low aptitude scores could never overcome this ‘handicap’ (Skehan, ibid).

b) It was strongly linked to methodologies prevalent at the time of Carroll’s research, which are now out of date conceptually, having little or nothing to do with the current communicative classroom (Dörneyi & Skehan, 2003; Skehan, 2002). Instructional methods changed radically since the 50s and 60s. The times of the structuralist view of language, the behaviourist learning theories, and the audio-lingual teaching methods had long passed. In the 70s the first communicatively oriented classrooms emerged. It was then questionable whether the traditional FL aptitude tests would really be effective in predicting learner development that was taking place in contexts where oral interaction and communicative activities dominated (Kiss & Nikolov, 2005; Robinson, 2001; Skehan, 1998).

c) Krashen’s assessment of FL aptitude (1981a) was really damaging. He held that it does not relate to unconscious acquisition but is rather linked to explicit, conscious learning and to non-communicative activities that are teacher-led and explicitly rule-focused. Dörneyi and Skehan (2003: 594) considers this to be ‘the kiss of death’ for aptitude as it strongly
associates it with activities that were condemned by communicative classrooms. Krashen suggested that aptitude is an important predictor of success only in the case of learning whereas in acquisition, attitudinal factors such as a low affective filter, will better predict the outcome of implicit acquisition.

d) During the same period great emphasis was placed on the design of language teaching materials that would have the widest possible acceptance by the international market. Thus, the narrow orientation of publishers was bound to underestimate the important role of learners’ individual differences because, commercially speaking, that was an unattractive option (Skehan, 2002).

e) There was an unclear relationship attested, a kind of mismatch between the four putative factors and the five subtests of the MLAT (Sawyer & Ranta, 2001). In some cases, the sub-tests involved a number of cognitive operations and not just one, as is the case with Carroll’s Spelling Clues (sub-test 3). The test examines the learners’ Phonetic Coding Ability but clearly L1 vocabulary knowledge also plays an important role.

Sparks and Ganschow (1991) gave language aptitude another dimension and suggested an alternative explanation to account for FL failure. They believed that the primary source of difficulty in learning to read and write in one’s L1 was the lack of phonological awareness and advocated a reconsideration of FL aptitude in the light of their Linguistic Coding Differences Hypothesis (LCDH). This assumes that FLL is enhanced or limited by the degree to which learners have control over their L1 skills (mainly phonological, orthographic, semantic, and syntactic). A deficiency in one or more of these skills is likely to affect the students’ ability to learn a FL.

5.5 A reconceptualisation of FL aptitude: The post-Carrollian era

Since the time Carroll and Sapon (1959) developed the MLAT, the field of SLA has grown enormously, and the understanding of the psycholinguistic processes implicated in acquisition
and learning have deepened considerably. Chan, Skehan, and Gong (2011) report that conceptions of L1 acquisition and SLA were then vastly different from current ones. Also, conceptions of memory did not include the separation between WM and LTM. Advances in cognitive psychology allowed for a more accurate representation of the various mental skills and aptitudes that make up the composite LL ability (Dörnyei, 2005). Dörnyei (ibid: 42) thinks that “language aptitude research has recovered completely and currently it is one of the most promising areas of SLA research” as it draws increasingly on cognitive psychology, psycholinguistics, and neurolinguistics and cognitive abilities such as WM.

According to Chan, Skehan and Gong (2011), the reawakening of interest in FL aptitude also triggered a re-analysis of the concept. The latest attempt is to reconceptualise it via the construct of WM (Robinson, 2005). Chan, Skehan and Gong (2011) propose that WM measures need to be integrated into a broader battery of aptitude sub-tests. Sawyer and Ranta (2001: 340) suggest that WM capacity “may be the key to elaborating the concept of language aptitude itself and to clarifying its relationship to the SLA process”. They base their alternative FL aptitude model on three assumptions: a) WM plays an important role in various SLA stages and cognitive processes, b) there are variations in WMC in individual L2 learners which can be measured by WM span tasks, c) different WM components have been found to correlate with different aspects of L2 performance (vocabulary, grammar acquisition) and the development of specific L2 skills (listening, reading, speaking, writing, and interpreting) (Wen & Skehan, 2011). Thus, Sawyer and Ranta (2001) conclude that:

Assuming that noticing is crucial to learning, and attention is required for noticing, and attention at any moment is limited by WM capacity, then there must logically be a close relationship between amount of learning and size of WM. It is also likely that WM serves as an arena in which the effects of other components of aptitude are integrated.

(p. 342)
Dörnyei (2005: 34) also finds it questionable whether it is still useful to use “the umbrella-term of ‘language aptitude’ with all this recent research into specific cognitive skills and capacities that relate to learning, such as WM or ‘phonological coding/decoding’.

Aptitude research so far has examined one of its components in greater depth, namely memory and, in particular, associative memory, which, however, is only “a limited aspect of memory” (Skehan, 1989: 30). Skehan believed the time has come to consider several other aspects of the functioning of memory, the size of WM being one of them. This is so because of the general consensus that individual differences do exist because of differences in WMC: some people may have a more effective WM than others. To understand the importance of WM, one needs to think of the loads of incoming information (e.g. language) as passing through a gate, a temporary, limited-capacity WM store (Baddeley & Hitch, 1974) before entering (if it eventually does) the permanent store of knowledge (long-term memory). It is clear, then, that greater WM size might be an advantage with respect to FLL. The construct of WM is rather complicated, since speed of processing, executive control, and workspace size (and structure) all combine to influence the level of its effectiveness for any one individual (Chan, Skehan, & Gong, 2011). The Executive Functions (EFs) of WM and Response Inhibition will be discussed in detail in Chapter 6.

For now it suffices to say that one of the components of WM, the phonological store, (Baddeley & Hitch, 1974) is the cognitive capacity to store information in real time. This was not measured directly by the MLAT (Sparks & Ganschow, 2001; Wen, 2012). It is clear though, that Baddeley and Hitch’s concept of the phonological store and Carroll’s phonemic coding ability share a lot (if not everything) in common. Carroll (1965, as cited in Chan, Skehan, & Gong, 2011: 60) was very close to the construct of WM when he said that it was not sound discrimination *per se* that contributed to aptitude – but the capacity to analyse sound so that this could be retained. He had realised from early on that there was a memory
link up to the auditory component of aptitude. Irrespective of how one chooses to name the construct, Chan, Skehan and Gong (ibid: 67) find this to be “perhaps most discriminating as the basis for aptitude testing” and hold the view that Carroll’s original conception of phonemic coding ability needs to be further explored and extended.

According to Kyllonen and Lajoie (2003), current FL aptitude research covers three major categories: a) the development of new aptitude tests, b) Carroll’s componential view of FL aptitude with the ultimate goal of further refining it, c) aptitude-treatment interactions with the design of learner profiles (e.g. the work of Robinson). Matching individual differences in abilities to the IP demands of different L2 tasks is an area that is now being systematically researched (see Wen & Skehan, 2011 for a detailed table).

Robinson’s work (2001, 2002a, 2005) builds on the work of Snow (1994) on the magnitude of individual differences in ability and personality, what he called aptitude complexes. Snow adopted a very broad definition of aptitude that would host any personal characteristic that could affect one’s learning (from cognitive characteristics such as reasoning, verbal ability, and spatial ability to affective ones like attitudes, talents or even physical characteristics). Snow also believed that instruction ought to be adapted to accommodate the aptitude(s) of individual learners. Lohman summarises this in what follows: “But which propensities function as aptitudes (or inaptitudes) depend ultimately on the transactions that occur on a moment by moment basis as the person perceives, acts, reacts, and transforms both by situation and self” (cited as personal communication in Kyllonen & Lajoie, 2003: 79).

Robinson’s work on language aptitude-treatment interaction (as cited in Dörnyei, 2005: 59-60) is significant because he makes the first attempt to describe concrete sets of cognitive skills that can be associated with some basic learning tasks, and then to identify specific aptitude complexes to match these cognitive processing demands.
Robinson’s framework (2002a) of aptitude complexes attempts to explain child-adult differences by exploring the ways individual differences interact with learning conditions. He agrees with Bley-Vroman (1990) that adults rely heavily on general problem-solving activities and exhibit much greater variation in levels of attainment. His framework is based on four hypotheses: a) the Aptitude Complex Hypothesis that supports that the IP demands of tasks draw differentially on aptitude complexes, b) the Ability Differentiation Hypothesis which holds that some learners have very differentiated strengths in abilities that contribute to aptitude complexes, c) the Fundamental Difference Hypothesis of Bley-Vroman (ibid) which claims that UG is no longer operative in adult SLA, in contrast to L1 acquisition which is UG-driven, d) the Fundamental Similarity Hypothesis, according to which adult learning under any condition (implicit vs explicit learning) is fundamentally similar. It is the result of the interaction between a pattern of cognitive abilities and the processing demands of the task. Hypotheses (a) and (d) together may help explain variation in adult L2 learning outcomes.

Robinson (2001, 2005) suggests that sets of basic cognitive resources (such as phonological WMC, processing speed or pattern recognition) have their effects on what he calls higher-order aptitude complexes like noticing the gap for instance, which, if joint together, facilitate processing and learning under different instructional contexts. Thus, learners may be classified on the grounds of their different multiple aptitudes. He strongly believes that assessing IDs in relation to aptitude complexes may be used diagnostically to match learners with an instructional task or treatment. He acknowledges the particular importance of these cognitive abilities to processing and learning from input, especially during the early stages of SLA (Robinson, 2005). He thinks that because early and later stages of LL draw on different abilities, or combinations of abilities, larger aptitude batteries are needed to inform selection. He claims that noticing is a necessary precursor to language development and SLA. Therefore, further aptitude research needs to identify IDs in abilities
that promote this cognitive ability (ibid). Schmidt (1990, 2001) was the first to discuss the importance of noticing the gap and believed that the process of SLA is largely driven by what learners pay attention to and notice in TL input.

In a similar vein, Skehan (2002) proposes an alternative approach to FL aptitude, driven by current SLA processing stages. Based on Carroll’s original model, Skehan (1998) suggested a three-factor model that combines processing stages with potential FL aptitude components. His model, as is clear from the table below, involves three abilities: a) the ability to notice what is in the input (thus, Carroll’s phonemic coding ability is related to input processing), b) central processing (thus, Carroll’s language analytic ability, i.e. grammatical sensitivity and inductive LL, are related to central processing), c) the ability to retrieve chunks from memory to support fluent speech production (thus, Carroll’s associative memory is related to output production). For a detailed discussion of Skehan’s approach, see Dörnyei and Skehan (2003) as well as Skehan (2002).

Table 2 summarises Skehan’s (ibid) proposal of SLA stages and aptitude constructs. What appears in italics in the aptitude column constitutes a future research agenda, as these are the areas that have not been explicitly addressed yet, by existing aptitude tests. The information was taken from Dörnyei (2005: 62):

Table 2: Skehan’s proposal of SLA stages and aptitude constructs

<table>
<thead>
<tr>
<th>SLA Stage</th>
<th>Corresponding Aptitude Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input processing strategies, such as segmentation</td>
<td><em>Attentional control</em></td>
</tr>
<tr>
<td></td>
<td><em>Working memory</em></td>
</tr>
<tr>
<td>Noticing</td>
<td>Phonetic coding ability</td>
</tr>
<tr>
<td></td>
<td><em>Working memory</em></td>
</tr>
<tr>
<td>Pattern identification</td>
<td>Phonetic coding ability</td>
</tr>
<tr>
<td></td>
<td><em>Working memory</em></td>
</tr>
<tr>
<td></td>
<td>Grammatical sensitivity</td>
</tr>
<tr>
<td></td>
<td>Inductive LL ability</td>
</tr>
</tbody>
</table>
Pattern restructuring and manipulation | Grammatical sensitivity
---|---
| Inductive LL ability
Pattern control | Automatisation
| Integrative memory
Pattern integration | Chunking
| Retrieval memory

### 5.6 FL aptitude and SLA achievement

Aptitude tests have been found to correlate with general as well as FL achievement. In general, the total score of the MLAT yielded multiple correlations of .40 to .60 (Carroll, 1981) between aptitude measures and measures of achievement. Sáfár and Kormos (2008: 116) cite Grigorenko, Sternberg and Ehrman (2000), who report correlation coefficients between various language aptitude tests and tests of FL proficiency that range between 0.23 and 0.73.

Eisenstein (1980) provided a more loose definition of childhood bilingualism, to include these individuals who attain considerable mastery in the two languages, but also those who have been exposed to more than one language before the age of 10-15 in a formal context. Thus, the author associated childhood bilingualism with enhanced FL aptitude and better FLL in adulthood.

FL aptitude tests have served both SLA research and SL pedagogy well: aptitude scores were often correlated quite highly with instructed LL success in a variety of institutional contexts (Dörneyi & Skehan, 2003; Robinson, 2001; Sawyer & Ranta, 2001; for a review see Skehan, 2002), and have been useful in predicting some areas of learning difficulty in the early stages of SLA (Grigorenko, 2002; for a review see Sparks & Ganschow, 2001). Additionally, versions of the MLAT (DeKeyser, 2000) and the PLAB (Harley & Hart, 1997, 2002) have been used to show that learners with a post-pubertal L2 learning onset depend more on general analytical abilities than pre-pubertal child L2 learners, who rely more on
memory. Harley and Hart (1997) concluded that aptitude has little power in predicting competence in the L2 for the younger learner. The exact opposite holds for older students, for whom accuracy and academic language proficiency is a more focal point of instruction than it is for the younger students. In the case of older students, aptitude correlated highly with most proficiency measures.

In terms of ELT pedagogy, FL aptitude research suggests that aptitude profiles, when used appropriately, in other words by matching the learners’ IDs with specific instruction methods, may enable the students to learn more effectively (Wesche, 1981) and also allow teachers to identify and manage L2 learning problems with much success (Sparks & Ganschow, 2001). Skehan (1998), Robinson (2005) and Wesche (1981) all agree that the future applications of current aptitude research have a great potential for serving as diagnostic tools. Robinson (2001), in particular, holds that if the purpose of intelligence and aptitude measures is not solely to predict but also to indicate the ways cognitive performance can be improved, as Pellegrino and Glaser (1979) have rightly argued, then this goal can only be achieved if IDs are interpreted in terms of cognitive processes that enhance or impede this. Robinson (2001: 378) concludes that “conditions of education might then be implemented that adapt to these individual characteristics”.

In addition, Cook (2001) thinks that through the early administration of FL aptitude measures, it will be possible to a) select students who are likely to succeed in the classroom, b) stream students into different classes, based on their aptitude levels, c) provide different teaching for different types of aptitudes with different instructional methods and final exams, or even d) excuse students with low aptitude from compulsory FL requirements. Reed and Stansfield (2004) however, wonder whether the MLAT can be used to early detect and identify a Foreign Language Learning Disability (FLLD) and raise questions that relate the test to ethical concerns.
CHAPTER 5: FOREIGN LANGUAGE APTITUDE

From all the above, it is clear that FL aptitude research has very clear implications for those engaged in a variety of important practical activities related to the teaching practice: from politicians and policymakers to school district administrators and language teachers.

5.7 Chapter Summary

The concept of FL aptitude has been the concern both of the SL/FL teacher and of the theoretical linguist. Nevertheless, a deeper understanding of this cognitive construct and of the faculty of language requires the interdisciplinary collaboration of linguists, psychologists, biologists, neurologists, philosophers and language teachers.

The construct of FL aptitude has developed considerably in the last 15 years, from being viewed as a unitary and stable trait (Skehan, 1998), to a more dynamic (Robinson, 2005) multi-componential property with “sets of malleable abilities” that interact with other internal “learner attributes and attitudes” such as motivation and learning styles (Larsen-Freeman, 2001, as cited in Wen, 2012: 234). In terms of ELT pedagogy, FL aptitude research suggests that profile-based information, when used appropriately, may allow material designers and policy makers to design “more effective interventionist techniques” like “cyclical syllabuses, or core plus satellite approaches” (Skehan, 1998: 199). These, in turn, mean “the tailoring of courses” (Skehan, 1989: 39) to help students learn more effectively (Wesche, 1981). In this sense, it may provide a framework for the study of learner strengths and weaknesses and thus of learner types. This can, perhaps, explain why FL aptitude research managed to ‘survive’, rejuvenate and still be in the heart of current SLA research.
CHAPTER 6
EXECUTIVE FUNCTIONS: WORKING MEMORY, RESPONSE INHIBITION AND THEIR CONTRIBUTION TO FL SUCCESS

6.1 Introduction
One of the basic questions in SLA/FLL research is what accounts for the learners’ differential success. The individual factors that influence this process have been widely researched in the past 30 years (for a review see Dörnyei, 2005). As already mentioned earlier in this thesis, L2 learners vary enormously in rate of learning and ultimate achievement levels. Several factors contribute to such individual variation, including: a) the instructional context (Wesche, 1981), b) the degree and type of their motivation (Dörnyei, 1998, 2005; Dörneyi & Skehan, 2003), c) the age at which instruction begins (Johnson & Newport, 1989; Oyama, 1976), d) the learners’ personality profile (Gardner, 1990). FL aptitude has also been viewed as one of the most important cognitive determinants of FL success (Carroll, 1977, 1981, 1985; Ehrman & Oxford, 1995). Recently, the focus of interest has shifted to the contribution of executive functions (EFs) such as WM, Inhibition and Shifting (for a review see Sawyer & Ranta, 2001) to FLL success across a number of settings, including the instructional one.

The following sections describe in detail the nature and practical applications of the two most relevant (to this age group) aspects of executive functioning under investigation in this thesis, namely WM and Inhibition.

6.2 Executive Functions (EFs)
This thesis focuses on the cognitive development and changes witnessed during the early school years. WM and Inhibition have been selected for further investigation because they are the two most relevant EFs for this age group, as they closely relate to children’s academic progress in the earliest stages of their schooling experience. A rapidly growing literature has
demonstrated their crucial role in many fields of cognitive development such as that of linguistic reasoning.

The umbrella term *EF* refers to general-purpose, higher-order cognitive control mechanisms which are interrelated and regulate behavior towards the accomplishment of a future goal (Huizinga, 2006; McAuley & White, 2011). Miyake et al. (2000: 49) define EFs as “general-purpose control mechanisms that modulate the operation of various cognitive sub-processes and thereby regulate the dynamics of human cognition”. The term refers to the many skills required to prepare and execute complex behaviours, such as attentional control, the organisation, monitoring and regulation of an activity, the inhibition or resistance to the interference of irrelevant stimuli, the cognitive flexibility to switch between tasks, the evaluation of the results of an activity, etc. (Huizinga, 2006; Klenberg, Korkman, & Lahtinen, 2001; Smith & Jonides, 1999). Carlson (2011) considers EFs to be especially important in new and/or demanding situations that require a rapid and flexible adjustment of behavior to the changing demands of the environment, while Carlson, Mandell and Williams (2004) suggest that EFs can predict the development of children’s Theory of Mind (ToM) from the age of 2-3. However, WM, Inhibition and Shifting are the three most frequently researched executive processes with respect to adults (Miyake et al., 2000) and children (Lehto et al., 2003; Monette, Bigras, & Guay, 2011) because of their close relevance to L1 acquisition and FL comprehension (Andersson, 2010).

Response Inhibition is a non-verbal EF which has been found to rapidly improve between the ages of 6-8 (Luciana & Nelson, 1998) while WM (what Miyake et al., 2000 call *Updating*) tends to be more strongly associated with verbal tasks and school achievement (St. Clair-Thompson & Gathercole, 2006). According to Miyake et al. (2000: 57), the “Updating function requires monitoring and coding incoming information for relevance to the task at hand and then appropriately revising the items held in working memory by replacing old, no
 longer relevant information with newer, more relevant information”. Thus, it actively manipulates relevant information in WM rather than passively stores information. Additionally, the authors speculated that WM, Shifting and Inhibition all involve some sort of inhibitory processes to operate properly.

Miyake and Friedman (1998: 360) proposed that WM for language should be equated with FL aptitude because it captures “the essence of the three important components of the language aptitude suggested by Skehan (2002): the language analytic capacity, the memory ability and the phonetic coding ability”. The idea has been overwhelmingly supported by Robinson (2002b) as well as by Sawyer and Ranta (2001), although Robinson (2001), following Schmidt’s (2001) theory of noticing and attention, emphasises the key role that attended processing plays in establishing new language and memory skills. This is extremely important in FLL as, according to Randall (2007: 127), FL learners are at a ‘severe disadvantage’ when compared to L1 acquirers in that FLL is a very explicit and effortful process. For language routines to become automatic a conscious, directed and targeted repetition of key language patterns is needed, which takes place within WM.

Research findings suggest that EFs emerge as a unitary domain-general process during the preschool years (Beveridge, Jarrold, & Pettit, 2002; Wiebe et al., 2011) to become fractionated into distinct yet interrelated processes around the age of 6 (Bjorklund & Harnishfeger, 1990; Dempster, 1981; Gathercole, Pickering, Ambridge, & Wearing, 2004; Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006). Neuro-imaging studies lend further support to the multi-faceted nature of EF, as different regions within the prefrontal cortex have been found to sub-serve different EFs (Huizinga, Dolan, & van der Molen, 2006) with the phonological loop being served by a neural circuit in the left hemisphere (in the Broca’s area) and the visuo-spatial sketchpad being associated with the right hemisphere and the premotor cortex (Gathercole, 1999; Smith & Jonides, 1999).
WM and Inhibition have been the focus of many studies that involve children. A general finding is that they both improve with age, although the precise rate at which they develop and fully mature varies across studies (for a review see McAuley & White, 2011). The interaction between these two EFs, affective factors and emotional states has also been investigated. It seems that high-anxious individuals are less efficient in controlling their inhibition mechanism and more susceptible to distraction, while positive emotions are usually associated with cognitive flexibility and the broadening of attention (Derakshan & Eysenck, 2010).

The capacity to suppress a prepotent response (i.e. inhibitory control) is evident from the age of 3 and reaches adult level of performance in late childhood, around the age of 12 (Klenberg, Korkman, & Lahti-Nuutti, 2001), while response accuracy (i.e. the ability to perform correct inhibits consistently) continues to improve into young adulthood (Bedard et al., 2002; Durston et al., 2002; Luna et al., 2004; van den Wildenberg & van der Molen, 2004). WM and Shifting, on the other hand, attain mature levels between 11-15 years of age (Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006). WM capacity, in particular, gradually develops from early childhood into adolescence (from around 4 to around 15) (Alloway, 2006; Beveridge, Jarrold, & Pettit, 2002; Brocki & Bohlin, 2010; Luciana & Nelson, 1998).

WM and Inhibition are inherent to the cognitive control of behaviour. Even though each has a distinct role, they often work in collaboration and modulate each other’s performance (Bjorklund & Harnishfeger, 1990). The efficient functioning of WM depends on inhibitory processes that limit the access to information which is irrelevant to the task. Given the limited capacity of WM, it is clear that an inhibitory dysfunction can be associated with difficulties attending to and processing new information (Joormann & Gotlib, 2010; Luna et al., 2004) that, in their turn, will cause learning difficulties.
McAuley and White (2011) suggest that the individual variation witnessed in WM and Response Inhibition performance across development directly relates to the speed with which individuals process incoming information. As children grow older and become more experienced, their processing speed increases, making it easier for them to maintain and manipulate information at a faster rate, and interpret contextual cues as to whether a certain behavior is relevant and appropriate to the accomplishment of a goal. Bjorklund and Harnishfeger (1990: 60) attributed the IDs attested in the speed of IP to the myelination of the nerves: “Myelin is the fatty sheath that surrounds nerves and facilitates the transmission of nerve impulses”. Myelination of the sensory and motor areas of the brain matures by the age of 2, while that of the integrative systems and higher-order cognitive processes is completed by the teen years.

Special emphasis has been placed on attention and inhibition because they are the two most elementary executive processes of all that emerge first (Luciana & Nelson, 1998), to be followed by EFs of verbal fluency and complex planning which reach their maturational peak around adolescence (Klenberg, Korkman, & Lahti-Nuuttila, 2001; Smith & Jonides, 1999). Reck and Hund (2011) claim that sustained attention (i.e. the ability to maintain focus continuously on specific stimuli) is a good predictor of inhibitory control during early childhood (between 3-6 years of age). Young children in the early school years are still stimulus-bound and impulsive and are thus likely to perform poorly on tasks that require the flexible switching between task demands and the inhibition of irrelevant responses (Klenberg, Korkman, & Lahti-Nuuttila, 2001). As children grow older and pass from the early preschool years to their middle teenage years, they become increasingly more field independent and manage to resist distractions (Dempster, 1991).
6.3 Defining Working Memory (WM)

In our everyday life activities, we often need to hold in mind pieces of information such as telephone numbers or PIN codes for short periods of time during some other parallel activity. This flexible capacity to store and process information is called working memory, and is extremely important for our effective cognitive functioning. To hold information in memory is an effortful, attention-demanding and highly prone to failure ability, due to memory overload or age. Children’s WM capacities undergo a remarkable increase (a two- to three-fold expansion in memory) between 4-14 years of age (Gathercole, 1999). This rapid increase in the functional capacity of memory during childhood makes the subject a crucial and fruitful area of research for cognitive psychology, developmental psychology as well as SLA research.

Despite much heated debate over the exact nature of WM or the extent to which its capacity limitations arise from processing skills or other resources, such as controlled attention (for a review see Gathercole, 1999), the most widely accepted and influential model of WM and its development was proposed by Baddeley and Hitch (1974) to be later revised by Baddeley (1986). Previous theories of memory systems focused mainly on its storage function, whereas this model views WM as a ‘workspace’ of limited capacity which can be divided between storage and control processing demands. Thus, a trade-off between the amount of attention directed to processing and the amount of material that can be held in WM buffers, when either exceeds the limits is a natural expectation (Baddeley, 2007; Walter, 2007).

The model views memory as a ‘flexible mental workspace’ (Gathercole & Alloway, 2008: 16; Gathercole & Pickering, 2000a; Gathercole et al., 1992) where incoming information is stored, processed and manipulated by means of executive control processes. In this view, WM plays a far greater role in cognitive activities such as FL comprehension,
reasoning and learning than was previously assumed (Baddeley, 2003; Gathercole et al., 1992).

The central executive, ‘a chief executive officer’ (Kimberg, D'Esposito, & Farah, 1997: 185) is in charge of the whole operation as it monitors the activities of the lower level systems. It is analogous to ‘the central processor or computing space M’ of Pascual-Leone (1970: 307). It is a modality-free or domain-general system, limited in capacity (Baddeley, 2003; Lehto, 1996) which performs a range of high-level regulatory EFs, such as directing attention, planning actions, solving problems, reasoning logically, or engaging in mental arithmetic (Baddeley, 1986; Kimberg, D'Esposito, & Farah, 1997). It integrates information from different WM subsystems and LTM, allocates cognitive resources, and generally organises the operations of WM. It supervises the workings of the two ‘slave’ systems, the phonological loop and the visuospatial sketchpad (Baddeley, 1996), which are domain-specific and responsible for the retention and manipulation of very specialised incoming material.

The phonological loop encodes, maintains and manipulates all speech-based material. Its primary purpose is to store unfamiliar sound patterns while more stable and detailed phonetic representations are being constructed in LTM (Baddeley, Gathercole, & Papagno, 1998). It consists of a passive short-term phonological store that holds material in a phonological code. This decays within 1.5-2 seconds but can be refreshed by subvocal rehearsal. The amount of material that can be rehearsed before its phonological trace fades depends on speech rate, which produces the so-called word-length effect: a smaller number of long words or a larger number of short words can be articulated within this time period. Consequently, memory span is a joint function of the rate of decay in the phonological store and the rate of articulatory rehearsal (Papagno & Vallar, 1992; Service, 1992). The loop is also responsible for registering visual information within the phonological store, provided this is silently articulated (Baddeley, 2003; Sáfár & Kormos, 2008). Developmentally speaking,
the phonological store is in place around the age of 3, but it is not before their 7th birthday that children engage in conscious and consistent rehearsal, a fact that also increases their PSTM capacity (Gathercole & Pickering, 2000a, 2000b; Gathercole, Pickering, Ambridge, & Wearing, 2004). Hence, it is clear that the phonological loop plays a vital role in the learning of new words, FL pronunciation, oral production and FLL in general.

The other slave system, the visuo-spatial sketchpad, deals with visual images and stores incoming material in terms of their visual and spatial features. Children younger than 7 heavily rely on this buffer to support recall of the physical forms of such stimuli, whereas older children also make use of the phonological loop to recode the visual input into a phonological form through subvocal rehearsal (Gathercole, Pickering, Ambridge, & Wearing, 2004).

The episodic buffer that has recently been added to the model (Baddeley, 1986, 2000) is a temporary system with limited capacity. Dörnyei (2005) explains that it is called a buffer because it acts as an intermediary between the sub-components of WM and LTM that all use different codes. Also, it is called episodic because it allows all kinds of sensory information (visual and verbal) to be combined into unitary and coherent episodes to be then linked to multidimensional representations in LTM (Alloway et al. 2004; Andersson, 2010; Nevo & Breznitz, 2011).

The capacity of all three systems is limited by the amount of information that can be maintained before this is lost and by the amount of time incoming information is available for processing. Figure 1 illustrates the multi-component model of WM in its current form:
Figure 1: The current multi-component model of WM (Baddeley, 2000, 2003)

The capacity of the phonological loop or PSTM is typically measured by the forward digit span and word or non-word span tasks, i.e. by the number of unrelated digits or (non)words an individual can recall, that constitute lexically familiar or unfamiliar verbal material respectively (Gathercole et al., 1992). On the other hand, the capacity of the central executive or complex WM (CWM) is jointly determined by storage and processing components (Gathercole, 1999). It is typically measured by span tasks, such as the backwards digit span (Kormos & Sáfár, 2008; St. Clair-Thompson, 2010) as well as reading or listening span tasks, in which both immediate recall of memory traces and simultaneous computation of some sort are involved (e.g. judging the veracity of a sentence and recalling a word) (Daneman & Carpenter, 1980).

To sum up, CWM combines both storage and processing functions. As already mentioned, because there are resource limitations to these two cognitive functions, during the handling of complex cognitive tasks a trade-off relationship is very probable to occur: if the resource devoted to ongoing processing is great, the other dedicated to storing its products will be less (Daneman & Carpenter, 1980). An alternative view has been suggested by Case, Kurland and Goldberg (1982), who argue that the total processing space (i.e. the limited M(ental)-capacity of Pascual-Leone, 1970) remains stable in one’s life span. What gradually increases is the efficiency and automaticity of the operating activities. Younger children allocate most of their total processing space to the operational function, leaving little space for
storage, which eventually leads to poor performance. As they grow older and with experience, they become more skilled at processing information because their cognitive functioning improves and their EFs and strategies become more efficient and automatic. This results in the release of more of the total processing space for storage.

WM has been identified as a key component of language aptitude (see Chapter 6) and has been found to play a central role in L1, child and adult L2 development, classroom learning, in particular vocabulary learning, as we shall see in the section that follows.

6.3.1 WM and its involvement in the LL process and FL success

The learning and use of a SL/FL draws on a range of cognitive processes, one of which is WM, the system that controls and regulates behavior in real time. WM, the ability to maintain and manipulate information online, is essential for the voluntary control of behaviour based on internal plans (Baddeley, 1986). In this sense, its development is believed to underlie the emergence of complex mental abilities (Dempster, 1981).

According to Juffs & Harrington (2011: 138) “Working memory refers to the mental processes responsible for the temporary storage and manipulation of information in the course of on-going processing”, while higher-order cognitive tasks such as comprehension, learning and reasoning are taking place. Therefore, it becomes clear that any differences in WMC will have immediate consequences on children’s cognition, their L1 acquisition and L2 learning.

Over the past years WM has been considered to critically influence SLA. This is based on a number of assumptions. First, unlike L1 acquisition which is UG-driven (Bley-Vroman, 1990; Chomsky, 1976, 2005, 2011) SLA is constrained by general learning mechanisms (such as FL aptitude and WM) (Miyake & Friedman, 1998; Sawyer & Ranta, 2001). Second, unlike the automatic and implicit processing in L1 acquisition, SLA is characterised by explicit and controlled processing, which is more effortful and demanding in terms of cognitive processing (Randall, 2007). Third, the components of WM, together with other aptitude
components are implicated in the different stages of SLA (input, central and output processing) and in the cognitive processes that take place within these stages (e.g. noticing, pattern recognition, automaticity) (Skehan, 2002). For instance, a large and efficient WM makes the noticing of important aspects of the language input more likely to occur (Robinson, 2005) which further facilitates the allocation of L2 learners’ attention on form. In other words, once learners manage to optimise their phonological memory processing speed and capacity, this immediately sets free their attentional resources (that would otherwise be tied up in the superficial processing of incoming material) to perform a more in-depth processing of syntactic patterns and semantic content (Hummel & French, 2010).

In this sense, an efficient WM constitutes a very important issue in today’s communicative FL classrooms. Children learning a FL these days are required to make sense of large amounts of aural data, especially in the earliest years of FL schooling (Mackey et al., 2002; Sáfár & Kormos, 2008). This places heavy processing demands on their PSTM and may put low WM learners at a disadvantage. Hummel and French (2010) propose some ways that will reduce the cognitive load placed on PSTM such as the design of communicative activities that make use of increased amounts of visual aids, the occasional use of reading aloud stories that promote rehearsal, the regular use of online resources. These pedagogical tools and strategies can enhance the FLL process in the totality of learners, but most importantly, in those learners that have a less efficient or impaired PSTM.

WM is directly implicated in a wide range of language processes, with PSTM and executive control functions being particularly important elements in L1 and/or FLL and use. Because SLA and FL pedagogy seek to address the question of why some learners struggle to learn a SL/FL within an instructed context, researchers have been interested in finding out whether WM limitations explain differences in success. The assumption is that a higher WM will lead to more successful learning in a variety of domains, without disregarding the
contribution of other factors such as attitudes, motivation etc. to be equally powerful explanatory variables in FLL.

It is well documented that the contribution of WM, the phonological loop in particular, is most critical in the early stages of vocabulary acquisition and the internalisation of new linguistic material (Mackey et al., 2002), be this L1 or L2 material. Therefore, deficits in this component will impair the learning of new words. More specifically, PSTM is closely associated with central aspects of LL such as:

a) L1 vocabulary development in children (Adams & Gathercole, 1996; Baddeley, 1986; Gathercole & Baddeley, 1993). The association is strong until the age of 5 (Gathercole & Adams, 1994; Gathercole et al., 1992) while the existing L1 or FL vocabulary knowledge, already established in the learner’s lexicon becomes the critical factor in the later stages of vocabulary learning. The relation between PSTM and LTM knowledge is reciprocal: the phonological loop supports the learning of new phonological patterns while the already stored knowledge of the phonological structure of the language supplements the phonological loop (Masoura & Gathercole, 1999).


c) adult L2 vocabulary learning (Cheung, 1996; Papagno, Valentine, & Baddeley, 1991). Kormos and Sáfár (2008) claim that PSTM plays a far greater role in less proficient L2 learners while its power gradually diminishes as the lexicon of the TL is being built up,

d) general school achievement (Alloway & Alloway, 2010; Gathercole, Lamont, & Alloway, 2006; Gathercole, Brown, & Pickering, 2003; Gathercole & Pickering, 2000a, 2000b;
Gathercole, Pickering, Ambridge, & Wearing, 2004; Gathercole, Pickering, Knight, & Stegmann, 2004),
e) reading abilities and comprehension (Daneman & Carpenter, 1980; Μαριδάκη-Κασσωτάκη, 1998; Maridaki-Kassotaki, 2002; Nevo & Breznitz, 2011; Walter, 2007),
f) (adult) writing skills (Kellogg, 2001, 2004; Kormos & Sáfár, 2008),
g) math skills (Monette, Bigras, & Guay, 2011) or English and maths (Gathercole & Pickering, 2000b; St. Clair-Thompson & Gathercole, 2006),
h) listening comprehension performance and speech production (for a review see Alloway, 2006 and Wen & Skehan, 2011) and the development of expressive language in preschoolers (Adams & Gathercole, 1996).

An impaired or insufficient WM (usually associated with poor PSTM skills) is closely linked to problems with EFs and inattentive behavior in normal children (Gathercole et al., 2008), learning deficits, poor school achievement with respect to reading and writing tasks. Children often have difficulties to follow instructions and seriations (what comes 1st, 2nd, etc.), face problems with their attention span, lose track in complex tasks, fail to cope with the simultaneous processing and storage demands of some tasks (Gathercole, 1999; Gathercole & Baddeley, 1993).

What is being worked on in WM may potentially become part of LTM. On these grounds, WM has been viewed as a gateway to LTM (Wen & Skehan, 2011). Gathercole and Alloway (2008) summarise the key role WM plays in supporting learning within an instructed context and view WM as ‘a bottleneck for learning’:

… the acquisition of knowledge and skills in complex domains such as literacy and mathematics requires the gradual accumulation of knowledge over multiple learning episodes, many of which will take place in the structural learning environment of the classroom. Learning is thus an incremental process that builds upon the knowledge of
structures and understanding that have already been acquired: any factor that disturbs this acquisition will have deleterious consequences for the rate of learning. (p. 23)

In a similar vein, Alloway (2006: 134) holds that because WMC is limited, an excess of either storage or processing demands during an on-going cognitive activity will lead to a ‘catastrophic loss’ of information from this temporary memory system. It is clear that the implications of research on WM for classroom practice are great and include intervention programmes designed to improve the learning outcomes of children with poor WM function.

A number of researchers (Alloway, 2006; Archibald & Gathercole, 2006; Gathercole & Alloway, 2005, 2008; Gathercole, Pickering, Knight, & Stegmann, 2004; Holmes, Gathercole, & Dunning, 2009; Maridaki-Kassotaki, 2002) support the idea that the phonological loop can be improved through early intervention and adaptive training. In this way, students will be helped to make maximal use of their PSTM capacity before academic difficulties are well entrenched (around Grade 3) (Roberts et al., 2011). To this end, a number of very discrete steps have been proposed that all reduce the demands classroom activities impose on learners’ WM: a) the use of simple and brief instructions that are frequently repeated, b) the use of simple and short sentences that contain a frequent vocabulary, c) the use of separate steps in tasks, d) the use of visual aids that are regularly practised, d) the development of self-help strategies in learners, such as asking for forgotten information, using memory aids, continuing with complex tasks even if some steps are not completed.

As already mentioned, WM has been implicated as a core element in FL aptitude (Chan, Skehan, & Gong, 2011; Dörnyei, 2005; Dörneyi & Skehan, 2003; Kormos & Sáfár, 2008; Miyake & Friedman, 1998; Randall, 2007; Robinson, 2005; Sawyer & Ranta, 2001). The above-mentioned research findings suggest that an efficient PSTM provides an advantage in the case of EFLL and can be a good predictor of L2 attainment. Therefore, it may be useful to reconsider the measure of FL aptitude that addresses very young learners, and design
additional tasks that tap children’s PSTM. This, according to Harley and Hart (1997), would add predictive value to the test as a whole.

6.4 Defining Response or Motor Inhibition

Inhibition has been defined by Dempster (1991) as resistance to interference, whereas inhibition control as the deliberate, intended and controlled suppression of responses when the context changes and these are no longer appropriate (Logan & Cowan, 1984; Miyake et al., 2000). The voluntary suppression of responses (also referred to as inhibitory control), “the ability to filter out distractors and retain a response set, is crucial for choosing a course of action based on a cognitive plan over alternative task-irrelevant behaviours” (Luna et al., 2004: 1357). The inhibitory process also involves speed of information processing, quality or quantity of information represented, other executive processes such as strategy selection, etc.

An efficiently-operating inhibition mechanism keeps information that is irrelevant to the demands of the task from intruding into WM, leaving thus more mental space available for the processing of task-relevant information only. This is very important given the limited storage capacity of WM, which is set free to process at a deeper level the filtered information.

Improvements in response inhibition may decrease WM demands as only relevant information will be processed, while faster processing may allow information to be more rapidly and effectively encoded in WM (Luna et al., 2004). Also, WM may support the preparation to inhibit a response (ibid). This is very important as IDs on the efficiency of the mechanism have direct consequences for knowledge acquisition (Dempster, 1992) and general academic learning (St. Clair-Thompson & Gathercole, 2006). In this sense, response or motor inhibition, one of the many components of behavioural inhibition (the others being inhibition of prepotent responses or the stopping of an ongoing response), fundamentally contributes to the efficient functioning of all other EFs (Brocki & Bohlin, 2010; Klenberg, Korkman, & Lahti-Nuuttila, 2001).
The control of inhibition is a major determinant of cognitive development, as effective performance on many tasks as well as in real-life situations (e.g. driving) requires this ability (Bjorklund & Harnishfeger, 1990). Dempster (1991: 157) acknowledged the importance of inhibition by suggesting it should be viewed as “a logically necessary component of intelligent behavior”. The capacity of inhibition is also an important source of IDs, closely associated with the frontal cortex of the brain. The frontal lobes are extremely fragile, exhibit marked individual differences in size and weight and are the seat where correlated information from all sources is integrated and enters into ongoing activity (Dempster, 1992). They play a superior role in human intelligence because they enable people to engage in higher cognitive functions such as goal-formulation, planning, effective performance, problem-solving, etc. (Smith & Jonides, 1999).

As already mentioned, younger children are sensitive to the interference caused by task-irrelevant information (Band et al., 2000). However, as they grow older and become more cognitively mature, they respond faster, are less susceptible to distraction and as their inhibitory mechanism becomes better developed, their selective attention also improves (Bjorklund & Harnishfeger, 1990).

Different aspects of inhibition mature at different ages. Durston et al. (2002) provide a neural-based explanation of the development of inhibitory control, which is still in the process of maturation across the ages of 6-10 while, according to Brocki and Bohlin (2010), a first stage of maturation of this mechanism is early childhood (from 6-8 years of age). Nonselective inhibition, measured by stop-all tasks where all responses need to be stopped indiscriminately, matures by the age of 5 (Band, van der Molen, Overtoom, & Verbaten, 2000). Selective inhibition, on the other hand, requires discrimination between responses where only one response needs to be inhibited (Logan, Schachar, & Tannock, 1997). This is a more cognitively demanding process since it requires the inhibition of some responses only.
Response inhibition is commonly measured by the Stop-signal task which requires participants to withdraw from a prepotent tendency to respond. It has been used in the normal population, young adults and children, as well as in clinical groups such as ADHD children (Band, van der Molen, & Logan, 2003). The primary index of performance in stop-signal tasks is called the SSRT (stop-signal reaction time) and indicates the speed of the inhibition process (see Chapter 7 for a detailed discussion of the SSRT). This increases throughout childhood, reaching a peak in early adulthood (around 16-17) to then gradually decrease throughout adulthood (Band et al., 2000; for a review see Bedard et al.; Dempster, 1992), while brain maturational processes such as myelination is held responsible for cognitive development and the enhancement of neuronal transmission during this period (Luna et al., 2004).

Martin-Rhee and Bialystok (2008) report significant differences between bilingual and monolingual children. The former are better able to resolve perceptual conflict and respond on the basis of a non-salient cue by ignoring a misleading but perceptually salient property (interference suppression). To give an example, in the Simon task two stimulus cues compete for the child’s attention. The child is asked to press the red button if a red square appears on the computer monitor and the blue button if a blue square appears on the monitor. The left and right shift keys on the keyboard are labelled with a red sticker and a blue sticker respectively. In congruent trials the coloured square appears on the right side of the appropriate shift key: the red square appears on the left side and the blue square appears on the right side. In incongruent trials the coloured square appears on the side opposite to the appropriate shift key. To efficiently resolve the conflict between the two cues, the child needs to ignore the most salient but irrelevant stimulus feature in favour of the less salient but correct option. In this task, bilingual children outperform monolingual ones, while they show no advantage on tasks that require response inhibition, i.e. when they need to exercise control over competing
responses, as is the case of the stop-signal task that will be described in detail in the following chapter. In this task, the two groups perform similarly. Very briefly, the stop-signal task requires children to override a habitual response to a stimulus (green arrow=go process) and replace it with a contrary response (red arrow=stop process) but also to perceive task instructions, comprehend task demands, and attend to the experimental stimuli (McAuley & White, 2011). The need to respond immediately upon seeing the stimulus increases the demands of the task as the perceptual competition between the two cues is enhanced, whereas the imposed delay on the stop-signal at a number of trials makes the task simpler because it offers the child some time to resolve the competition, resist the immediate association and respond in a more controlled manner (Martin-Rhee & Bialystok, 2008). Unlike perceptual inhibition, response inhibition is concerned less with attentional control and more with the execution of motor responses to familiar stimuli. In this sense, Martin-Rhee and Bialystok (ibid) conclude that being a bilingual does not entail withholding or replacing habitual responses.

Given the fact that inhibition tasks are rather multifaceted as already discussed, multiple indicators of inhibition fail to converge on a unitary construct (Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006; McAuley & White, 2011).

6.5 Chapter Summary

Over the past 30 years the learners’ differential success in FLL has been accounted with respect to their IDs (e.g. motivation, starting age, FL aptitude, etc.), while only recently the academic interest shifted to the contribution of EFs, which are general-purpose, higher-order cognitive control mechanisms that regulate behavior towards the accomplishment of a future goal. Of the three most frequently discussed EFs (WM, Inhibition and Shifting), this thesis concentrates on the first two, namely WM and Inhibition, as these are the two executive
processes that emerge early in life and thus relate directly to the age group (6- to 8-year-olds) examined in this thesis.

The study of WM has influenced SLA thought lately. This is so because, unlike the automatic and implicit processing in L1 acquisition, SLA is characterised by explicit, conscious and controlled processing, which is more effortful and cognitively demanding as an activity. In addition, WM has been identified as a key component of language aptitude, i.e. one’s readiness and capacity to learn a FL. The ability to maintain and manipulate auditory information online, has been found to play a central role in L1, child and adult L2 vocabulary development and general classroom learning. Viewed from this perspective, it is important that children have an efficient WM nowadays. In today’s communicative FL classrooms they are required, even from the earliest years of FL schooling, to make sense of large amounts of aural data, which places heavy processing demands on their PSTM.

To conclude, WM and inhibition are inherent to the cognitive control of behaviour. WM depends on inhibitory process for its efficient functioning, as these limit the intrusion of irrelevant information into WM’s mental workspace. Given the limited capacity of WM, it is clear that an inhibitory dys- or malfunction will have immediate consequences on the proper processing of new information and on eventual knowledge acquisition.
PART II: RESEARCH METHODOLOGY AND THE FINDINGS OF THE STUDY
7.1 The study and its goal

The present study is longitudinal in nature as it took place during the academic years of 2010-11 and 2011-12. It seeks to investigate the cognitive impact of the early introduction of English as a FL on Greek learners in an experimental, primary public school, where exposure begins at the age of 6 (Grade 1) and is intensive (5 hours per week). In an attempt to explain individual differences in the acquisition of L2 English vocabulary, it aims to explore whether the early exposure to EFL positively affects young learners’ cognitive functioning. To my knowledge, this has never before been examined as EFLL studies so far have primarily focused on the linguistic outcome of EFLL.

7.2 A brief account of the theoretical framework of the study

FL aptitude research on very young learners (Alexiou, 2005, 2009) suggests children’s cognitive (non-verbal) skills, i.e. memory (e.g. associative memory, recoding ability) and analytic ones (e.g. inductive reasoning, visual perception or classification ability), are in place at the age of 6/7 and relate closely to FLL success. Of the age-group of 5- to 9-year-old children, 7-year-olds exhibit stronger memory skills than 9-year-olds while at the age of 8 they are more analytic than at the ages of 7 or 9. Finally, 7- and 9-year-olds score high in FL language tests (receptive and productive) (Alexiou, 2005). Although preliminary, promising were the results on phonological discrimination tasks (repetition, discrimination and recall of non-words) as they yielded significant correlations with FL achievement. The author also reports an overall outperformance of girls on a number of cognitive tasks without this reaching statistical significance. The present study will further investigate this last issue, with respect to EFLL, to see whether this has a different impact on the performance of boys or girls in a variety of cognitive tasks that tap different aptitude components.
We will also examine whether Carroll’s (1981) phonetic component or the more recent construct of PSTM is the sole predictor of FL comprehension and production. WM has recently been identified as a key component of FL aptitude (Chan, Skehan, & Gong, 2011; Dörnyei, 2005; Robinson, 2005; Sawyer & Ranta, 2001; Sparks & Ganschow, 2001; Wen & Skehan, 2011) and the current prevalent view is that the time has come for a drastic reconsideration of the traditional aptitude construct Carroll originally designed, which Carroll (1990) himself also realised was necessary. Sawyer and Ranta (2001: 342) conclude that “WM serves as an arena in which the effects of other components of aptitude are integrated”. WM plays a central role in L1 development, child and adult L2 development, scholastic achievement in general and vocabulary learning in particular. Any future aptitude measure will need to take into account and make use of an extensive body of findings coming from cognitive psychology research. The incorporation of new sub-tests that examine learners’ ability to temporarily store incoming verbal stimuli for further phonological processing is of vital importance as today’s FL classes are communicatively oriented and rely heavily, if not solely, esp. during the earliest years of FL schooling, on the development of oracy in young learners with the implementation of a variety of oral activities that make FLL fun.

WM, as this was conceived by Baddeley and Hitch (1974), critically influences the acquisition of the L1, the L2 or the FL. Both empirical research and theoretical findings so far suggest that the phonological loop (PSTM) plays a key role in the long-term retention of new L1 or L2 vocabulary (Gathercole & Alloway, 2008; Gathercole et al., 1992; Masoura & Gathercole, 1999; Μασούρα, Gathercole, & Μπαμπλέκου, 2004).

Nevertheless, unlike the automatic and implicit processing in L1 acquisition, FLL requires a more effortful and controlled processing which is overall more demanding in nature (Randall, 2007). The components of WM, such as PSTM, along with other aptitude components are directly implicated in the different stages of SLA/FLL (input, central and
output processing) and a wide range of cognitive processes that take place within these stages (e.g. noticing, pattern recognition, automaticity) (Skehan, 2002). It is clear that a large and efficient WM makes the noticing of important aspects of the verbal input more likely to occur (Robinson, 2005). This, in turn, facilitates the allocation of L2 learners’ attention on form, which is an issue of extreme importance in FLL. In the light of new empirical research and theoretical findings (Baddeley & Hitch, 1974, Baddeley, 2000) and the L2 learning theories that follow the IP paradigm, we need to further examine the role of CWM, i.e. of the attentional controller, in EFL.

Finally, research findings indicate the beneficial effect of bilingualism on children’s cognitive skills. Bilingual children (as young as 4 years old) tend to score higher than monolingual ones in non-verbal tests due to a greater executive control and better developed inhibition mechanism (Bialystok & Feng, 2009; Yang & Lust, 2005). On these grounds, it would be interesting to explore whether an intensive EFLL programme within a school context has a positive effect on young learners’ executive functioning and, in particular, on their response inhibition control. Thus, apart from WM, we will also investigate another most frequently discussed EF relative to this age group, namely Inhibition which reflects the ability to keep information which is irrelevant to the task from intruding into WM.

7.3 The research questions

Having said that, it would be wise to pose the Research Questions of this study along with the hypotheses formulated, on the basis of the literature review conducted earlier in this thesis (Chapters 2-6) and the brief rationale just given in this chapter.

**Research Question One: Does the early introduction of EFL enhance any of the aptitude skills associated with young learners** (Alexiou, 2005)? Children at this age (6- to 8-year-olds) are in a transitional period with respect to their cognitive development, moving from the pre-operational to the concrete operational stage of Piaget. During this period, their
memory and analytic abilities increase considerably. Alexiou (ibid) found that the skills that are mostly affected by FLL are inductive reasoning, paired associative memory and visual perception. Based on this previous finding, it is hypothesised that these will be the skills that will be positively affected by EFLL. The learning of an additional FL, esp. at this early stage, is based on the constant repetition of linguistic material through various oral activities to ensure that learners start gradually forming associations in their mind (Webb, 2007a, 2007b).

By the age children enter primary school, they have an already established conceptual knowledge base and basic vocabulary in their L1. What they first need to do during FLL is to establish new links between pre-existing concepts and the new foreign words. The schooling experience also plays a facilitative role in the expansion of children’s knowledge base and the further development of their memory strategies and logical problem-solving skills (Cole & Cole, 2001). In other words, what is suggested here is that these will further be facilitated by the nurturing environment provided by the experimental school and the optimal learning conditions children enjoy there, the intensity of the program followed, the advanced qualifications of the English teachers, the teaching methodology applied and the input flood learners receive (see Chapter 4 for the current European situation regarding EFLL). On these grounds, it is predicted that the experimental group will experience a cognitive benefit from EFLL and a boost of at least some of their cognitive abilities mentioned above.

If this proves to be the case, this would be a first major finding. If indeed some cognitive skills are enhanced because of EFLL, then a cognitive interaction between the two languages that goes both ways is possible. If the learning of an additional language has a beneficial effect on an individual’s cognitive functioning, this will be to the benefit of both language acquisition processes, i.e. of the L1 which is still in progress and of the FL. The reciprocal nature of the relationship between the two languages is very interesting and often documented in the literature (Taylor & Lafayette, 2010). According to Kecskes and Papp
(2003), the transfer of cognitive skills from one language processing to the other is possible because of a Common Underlying Conceptual Base (CUCB) that is formed when the L2 proficiency reaches a necessary threshold (after some years of FL schooling). If our hypothesis is confirmed and these children have a positive cognitive benefit from EFLL, then this will possibly benefit their L1 processing and acquisition which, like a boomerang, will further determine FLL success.

**Research Question Two:** Is PSTM, at this early stage of FLL, the sole critical factor of FL performance? It is well-documented that L1 and L2 vocabulary acquisition can be predicted by PSTM. Nevertheless, FL learners are at a disadvantage when compared to L1 acquirers (Randall, 2007) because they have to undergo a very explicit and effortful process before their FL routines become automatic. FL learners need to engage in conscious, directed and targeted repetition of key language patterns. These cognitive activities all take place within WM, i.e. the temporary storage and processing of incoming information before this enters LTM.

Schmidt’s theory of noticing and attention (2001: 29) emphasises the key role attended processing plays in the establishment of “new or modified knowledge, memory, skills, and routines”. With the huge amount of stimuli that are of different types and which come from many different modalities, it is now known that only a very small portion manages to enter our WM for further processing. In this sense, CWM, the attentional controller of WM, is expected to play a key role in this early process of FLL. Nevertheless, participants are expected to be better in PSTM than CWM tasks in Grade 1 for three reasons: a) they are still in the process of building up their L1 vocabulary and thus are still very much dependent upon their phonological store to do so, b) although previous research indicates a close relation of the phonological loop and L1 or L2 acquisition up to the age of 5, Service (1992) argued this can even stretch up to the age of 11-12, c) the schooling experience will have trained learners
CHAPTER 7: THE PRESENT STUDY

enough by Grade 2 to consciously attend to certain aspects of a task (i.e. CWM) to get the desired results. Taking all these factors into account, it is hypothesised that the central executive will take over in Grade 2.

**Research Question Three:** *Can the early introduction of the FL, implemented in an intensive manner, be associated with a firmer control of one’s response inhibition mechanism?* As discussed earlier in this thesis, the two EFs, WM and Inhibition control, are considered indispensable factors to (F)LL (see Chapter 6). WM is responsible for the simultaneous storage and processing of incoming information. An efficient use of WM processes must involve interactions between a complex of component processes that involve accurate sensory perception, discrimination of relevant environmental stimuli, their encoding and maintenance in STM, efficient memory retrieval and the final “integration of mnemonic representations with relatively sophisticated behavioural responses that will guide behaviour towards desired goal” (Luciana & Nelson, 1998: 274). An effective Inhibition mechanism, on the other hand, will allow only information which is relevant to the task to enter WM for further processing.

Inhibition comes in many varieties, depending on the paradigm used. In this study we tested the informants’ ability to inhibit motor responses, i.e. response inhibition, by means of the stop-signal task. CWM is also involved in this task as learners need to keep the task instructions in mind (green=go process, red=stop process) and use this information to guide their immediate behavior (the pressing of the button or the withholding of a response).

On the one hand, research findings demonstrate that children younger than 8 (Flavell & Miller, 1998) are more impulsive than older ones and get easily distracted and thus struggle to self-control and sustain attention during a task. On the other hand, evidence suggests that bilingual children develop a greater executive function and exhibit a remarkable cognitive flexibility, especially with problem-solving tasks that require attention and inhibition control.
CHAPTER 7: THE PRESENT STUDY

This has been explained on the grounds of their constantly controlling attention between their two active and competing language systems, suppressing the one to communicate fluently in the other that is required. In this sense, the bilingual experience is a source of constant practice that boosts those control processes and makes them available for other tasks (Martin-Rhee & Bialystok, 2008). However, as the two authors argue, bilingual children retain this advantage only with tasks that require perceptual inhibition where they need to exercise control of attention over competing cues and suppress interference. With response inhibition tasks, as is the case of the stop-signal one, the imposed delays on the stop signal eliminate the bilingual advantage of learners (who perform the same as monolingual children) as the additional time given allows for the resolution of the conflict.

By making this comparison, I do not of course wish to equate the two processes (the bilingual with the FL experience). Still, because of the intensive FL programme followed in the experimental school, the input flood learners receive, the many opportunities they have to play, practise and experiment with the new language, it was viewed interesting to explore whether the early introduction of the FL would facilitate the emergence of a firmer control of the experimental group’s response inhibition mechanism, even though on the grounds of what has been said so far, the prediction was that the two groups would perform the same.

Research Question Four: Is there a gender effect associated with EFLL? As already mentioned, Alexiou’s findings (2005) suggest an overall better but non-significant performance of girls over boys on a variety of cognitive abilities that relate to FL aptitude. In addition, the EFLL literature indicates that young girls are initially more enthusiastic and motivated than boys and hold a more positive attitude toward FLL in general. The last research question will investigate whether a gender effect will emerge in Grade 2 (after the two-year FL intervention) with respect to certain skills: PSTM, CWM, Inhibition control, L1 verbal intelligence, and FL aptitude. It is predicted that no such effect will be found, as it is
perhaps too early in the FLL process for any differences between the two genders to come about. If differences are to be found, these are expected to emerge in favour of the experimental group as a whole with respect to their PSTM ability, CWM and possibly Inhibition control due to the FL intervention. Although a gender effect is not an often reported factor that explains differences among individuals in the early stages of LL, it would still be interesting to explore whether such a possibility is viable in this case.

The above-mentioned research questions were addressed through examining the database of the results compiled in the two academic years the study lasted. Details are given in the following sections.

7.4 The two schools

The research was carried out in Greece, within an instructed context and the primary schools under investigation were both located in Evosmos, an area in the western part of Thessaloniki, where families are of a similar socio-economic status, i.e. low to average. The two schools of the study are the following: a) the 2nd primary school, in which English is introduced in the curriculum in Grade 3, when children are 8-8.5 years old, and b) the 3rd Model Experimental School which introduces English as a FL (EFL) from Grade 1 with five 45-minute lessons per week. From Grade 3 to Grade 6 English is taught for eight 45-minute lessons per week. Throughout the study English was taught by four teachers who collaborated very closely. The school’s curriculum is communicatively-oriented and teaching follows the lines of Asher’s (1982) Total Physical Response (TPR) method and Krashen and Terrell’s Natural Approach (1983). Exposure to the L2 involves receptive and productive language, as “both qualify equally as examples of practice” (Bialystok, 1981: 25). Without knowledge of vocabulary there can be neither comprehension nor production. The major idea and one of the aims of the experimental school under investigation is to build up a basic vocabulary in young learners (i.e. words that are salient to their life experiences), in a ‘fun’ and pleasant way. This, after all,
will serve as a solid foundation for the grammar teaching that is to come in the following years.

7.5 Recruitment of participants

All participants were native speakers of Greek. As already mentioned, to control socioeconomic factors, all participants were recruited from their 1st grade from a similar low to middle-class neighbourhood in Evosmos. The exclusionary criterion applied upon recruitment was only one: bilingual and trilingual children or children with previous contact with English would be excluded, as an additional language might influence their English language proficiency. The informants came from three Grade 1 and Grade 2 classes respectively from both schools, which make 6 classes in total.

More specifically, 49 children (23 boys and 26 girls) were recruited from the 3rd experimental group. The mean ages and standard deviations for Grades 1 and 2 respectively were the following. For Grade 1: Mean 6 years, 4 months, SD 3 months; range: 5 years, 10 months to 6 years, 9 months. For Grade 2: Mean 7 years, 8 months, SD 3 months; range: 7 years, 2 months to 8 years, 2 months.

Likewise, 49 children were recruited from the control group: 23 boys and 26 girls. The mean ages and standard deviations for Grades 1 and 2 respectively were the following. For Grade 1: Mean 6 years, 7 months / SD 3 months; range: 6 years, 2 months to 7 years, 1 month. For Grade 2: Mean 7 years, 7 months / SD 3 months; range: 7 years, 2 months to 8 years, 1 month. The difference between the mean ages of the experimental and the control group during Grade 1 was because the experimental school was visited first for the 1st wave of data collection. The control group was visited immediately after Christmas vacation. This is so because permission was only then granted by the Pedagogical Institute of Athens. This procedure is common ground but rather time-consuming and needs to be followed by every
prospective researcher, interested in gaining access and collecting data either from the staff or the students of the primary or secondary sector of education.

Due to the informants’ very young age, informed consent was obtained from the parents. The monolingual students were located by means of a questionnaire and a letter distributed at the very outset of Grade 1 (via the school principals and 1st grade teachers) to the students’ parents/guardians. The detailed letter explained the purpose of the research and requested permission for the child’s participation. It emphasised that participation was by no means obligatory and that all informants would be seen by the researcher in hours that fall outside the school’s ‘core’ program (i.e. Greek language, English language, Mathematics): Gym, Music, Flexible Zone, Art classes, or during the whole-day school hours (from 14:00-16:00). It was mentioned that all data would be codified to ensure confidentiality (these two conditions were set by the Pedagogical Institute). Thus, the children whose parents/guardians consented to their participation attended the testing sessions. The children received a sweet as a small compensation for their participation in the research.

21 and 18 students were excluded through the questionnaires from the experimental and the control school respectively, either because they did not meet the criterion of monolingualism or because their parents did not wish their children’s participation in the research.

7.6 The testing procedure
The informants were seen in two testing sessions by the researcher: in the beginning of Grade 1, before the FL intervention on the experimental group, and towards the end of Grade 2, i.e. after their two-year exposure to EFL. At the beginning of each testing session, task instructions were given to participants in Greek. With the exception of the English Vocabulary test, the verbal tests were all administered in Greek. Participants were tested individually in a quiet room, free of distractions, in the school’s premises. Because of the
plethora of tests and their variable duration, all participants were seen in four different occasions each year, separated by at least one to three days. Each interview lasted approximately 20 minutes.

All participants completed the same tasks, most of which are standardised, while others were either previously tested in large populations or are well-established in developmental literature. There was one exception only, that of the two non-word repetition tests that were re-administered in the beginning of Grade 2 (October-November 2011), as in the first testing session the words were not recorded but presented auditorily, both the Greek-sounding and the English-sounding test. During the first presentation of my PhD course, I was advised by Dr. Nikolaidis to use NSs for both languages for the two timed recordings of the nonword tests. The second and final testing session for all the testing tools was conducted between February and May 2012 in both schools. All tests were previously piloted and taken by five young children of that same age.

The testing schedule was held constant for each child as far as possible, although some ad hoc reordering was occasionally necessary to prevent boredom and to sustain the young learners’ interest and attention. The order in which the tests were presented was the following: the informants were first given the non-verbal tests and then the verbal ones. The same battery of tests was administered in both Grades to both groups, with the exception of the English vocabulary test which was only taken by the experimental group, just before the end of Grade 2. The English test examined the informants’ receptive and productive skills. It was devised by the researcher and was based on the material covered during the first two years of the FL intervention. As already mentioned before, the control school had no contact with the English language, for it first introduces EFL in Grade 3.

All in all, primarily the cognitive but also some aspects of the linguistic development of the participants were investigated both within and across groups in the course of the two years
of the study. The author carried out all testing sessions. The testing materials, all child-friendly, are described and explained in the section that follows.

### 7.7 The tools used and the skills tested

All the individual tests used, verbal and non-verbal, are well established in developmental literature and relate to the theoretical framework that underpins this thesis with respect to FL aptitude (Alexiou, 2005) and its relation to WM (Baddeley & Hitch, 1974), to Inhibition control (Logan & Cowan, 1984) and L1 verbal IQ.

The non-verbal tests of the battery are the following: a) Young Learners’ Aptitude Test (YLAT) (Alexiou, 2005) indexes young learners’ FL aptitude and b) the Stop-Signal test that taps Response Inhibition. The verbal tests tap participants’ PSTM (Digit recall_Forward, Nonword repetition_Greek/Nonce_Gr., Nonword repetition_English/Nonce_Eng.), their L1 verbal intelligence (DVIQ I: Vocabulary and Metalinguistic concepts sub-tests, DVIQ II Metalinguistic concepts sub-test), and finally the second of the two EFs studied in this thesis, i.e. the central executive (the Listening and Recall task and the Digit recall_Backward). The procedure for every individual test is described in detail below. First are described the non-verbal cognitive tasks (7.7.1) and then the verbal ones (7.7.2):

#### 7.7.1 Non-verbal tasks: Measures of cognitive processing

#### 7.7.1.1 The Young Learners’ Aptitude Test (YLAT)

Nonverbal ability is a cognitive skill linked with children’s capacities to acquire knowledge and skills in the early school years. The cognitive skills of the informants were tested by the computerised version of the YLAT (Alexiou, 2005). This tests two types of skills in young learners between the ages of 5 and 9, namely analytic skills (Classification, Spot the Differences, Jigsaw, Story Sequences) and memory skills (Short-term Rote Memory, Paired Associates pictorial game, Semantic Integration). The execution of the tasks that follow presupposes in some, simple storage and in others a simultaneous further processing of the
non-verbal material presented. In this sense, the test is a measure of non-verbal WM. The sub-tests are more like games than formal testing tasks. The whole test is run in Greek, as due to their small age, learners are not yet literate in English. The overall score is 136 points.

More specifically, the test is comprised of the following sub-tasks:

- **Classification** (or artificial language) game: The activity tests participants’ inductive reasoning/learning ability where learners need to discover and apply new rules in the same way foreign grammatical rules work. “Word learning is a paradigm case of inductive learning” (Bloom, 2002: 4). It possibly affects organising in the mind parts of speech and grammar inference rules as well as thematic concepts of words (Alexiou, 2009). Six colours represent six different groups. It is first explained to the child that red stands for all animals, blue for all flowers, yellow for food, green for drinks, purple for clothes, and orange for means of transport. A small software demo precedes the actual task to make sure the child establishes the connection between the colour and the group it represents. When the picture card appears on the screen the child has to drag this to the appropriate colour, blue or red in the beginning. The test gets more difficult as progressively more colours are shown on the monitor. Score: 48 points.

- **Spot the Differences**: The visual perception of participants is tested twice. Two seemingly identical pictures appear on the screen each time. The child is asked to observe them closely and identify six differences in the first slide and twelve in the second. These relate to colour, number of objects in the picture, placement of objects in the picture, etc. The sub-test identifies the learners’ ability to recognise the presence, absence or a change in information. (Alexiou, 2009). Score: 18 points.

- **Jigsaw game**: Sensitivity to image perception and spatial ability is gradually tested through three series of unfinished puzzles (from the very easy to the more difficult). In the first slide the child sees an unfinished puzzle and the three missing pieces. These are
left out of the puzzle, dismantled on different corners of the slide. The child is asked to carefully observe and identify which piece fits which void in the unfinished puzzle. In the second slide three pieces are again missing, but this time the picture is slightly more difficult as one extra piece, the distractor, is given along with the three correct pieces. Four pieces in total, from which the child has to choose only three. In the third slide there are three missing pieces and two distractors. The total number of pieces for the child to choose is five. This is an ‘analysis by synthesis’ task, where the child is tested whether (s)he is able to “break up the visual field and keep part of it separate”, essential to language use (Alexiou, 2009: 53). Score: 9 points. However, it should be noted that this task was left outside the overall score, as all children managed to get all nine points and thus the task made no difference to the overall scoring of the test. Thus, the overall score for YLAT changed to 127.

- **Story Sequence**: The activity tests children’s reasoning ability that resembles cloze tests or gap filling tasks in EFL. Their inductive ability is tested with the help of situational clues. The test requires both perceptual and conceptual skills. The child is shown two slides. In the first, four jumbled pictures appear on the monitor. The child has to put the pictures in the correct order to make some sense of the story illustrated in them. There is no immediate rule; the child has to imagine logical rules to create the story and see the whole picture from the parts (Alexiou, 2009). The same procedure is followed in the second slide, which contains six disordered pictures that, once put in the correct order, make up a different story. Total score: 10 points.

- **Short-term Rote Memory** (or Kim’s game): In the first slide eight unrelated objects are shown for 30 seconds. When they disappear, the child is asked to recall as many of them as possible. In the second slide the objects are twelve (all different from the ones shown earlier). Again, the same procedure is followed. Score: 20 points.
Paired Associates pictorial game: This task is about associative (rote) memory, in particular the ability to retain sign pairs, a necessary ability in the retention of FL vocabulary. This is tested twice in this game. In the first slide, six different picture cards are shown, coupled with a set of six various shape-like figures (e.g. a turtle makes a perfect match with a line ending in a triangle). It is explained to the child that (s)he should look at them carefully in order to be able to recall them in the slide that will follow, where the set of figures is given while the pictures are all jumbled. The child is asked to provide all six perfect matches the way these were initially shown. The same procedure is followed in the second slide, where eight sets of matches are shown on the screen, six of which were included in the first slide and only two sets of picture-figure matches are new to the child. Score: 14 points.

Semantic Integration: This game tests participants’ recoding ability (shapes with numbers), which serves to increase storage capacity. There are two series of slides. In the first, a learning list of four shapes appears on the screen for 30 seconds (square, hexagon, diamond, X). The child is asked to look at them carefully and name the shapes. Then follows a new slide with a recognition list of six shapes (triangle, parallelogram, X, circle, hexagon, and a star) where the child is asked to recognise the new entries in the learning list (triangle, parallelogram, circle, star). In the third slide, the child has to recognise the shapes that appeared in the first learning list and finally, in the last slide, the child has to recall the shapes that were in the learning list but are now omitted (square, diamond). In the second series of slides the learning list comprises six shapes and the recognition list eight. The recording ability measured here, the ability to recognise the presence or absence of significant information might be associated with the capacity to learn language features such as word endings. (Alexiou, 2009). Score: 17 points.
7.7.1.2 The stop-signal task

The stop-signal task is a two-choice reaction task to visual stimuli, designed to assess Inhibition, one of the components of the EF. More specifically, it tests the control over competing motor responses. The stop-signal paradigm provides a way to test everyday requirements such as the abortion or the replacement of an action. It involves two tasks that run concurrently, a go-task (=the primary one) and a stop-task (=the secondary one). When the latter occurs it tells the subjects to suppress response to the former.

The task follows the stop-signal paradigm proposed by Logan and Cowan (1984) that resembles a horse race. According to the horse-race model, successful inhibition and failure to inhibit can be viewed as the result of an independent race between two horses (=sets of processes) for the first finishing time. The first set refers to the execution of a response (the go horse or response to the arrow stimulus, i.e. the green arrow). The second set refers to its withholding (the stop horse or the inhibition of response when the stop-signal, i.e. red arrow, is presented). The basic assumption of the model is that the two horses do not affect each other’s speed. In other words, there should be no interference between the two tasks as the distribution of the Go-RTs (reaction times on trials when no signal is presented; the primary measure of response execution) matches that of the SSRT (those when a stop-signal is presented; the primary measure of selective inhibitory control). In other words, Go-RTs and SSRTs are the independent variables and the relative finishing times of the horses are decisive for the outcome of the race (Logan & Burkell, 1986; Logan & Cowan, 1984).

Whether or not participants inhibit their response depends on this race between the two processes. The stop process wins the race if it finishes before the go process. When this happens response to the red arrow is correctly withheld. Otherwise, if the go process (=reaction to the arrow stimulus) finishes first, it is the winner of the race as the participant responds to the stimulus by pressing the button, escaping from inhibitory control as if no stop-
signal had ever been presented. The probability of stopping is subject to experimental control. The relative timing of the stop process is manipulated by delaying the presentation of the stop-signal (SOA=stimulus-onset asynchrony) vis-à-vis that of the reaction signal (Band et al., 2000). In this way, participants cannot tell in advance whether they will inhibit the response to the first stimulus or not. Thus, inhibitory control depends on the latency of the response to the go-signal (Go-RT) and on that to the stop-signal (SSRT). Inhibiting when given a stop-signal is evidence of good impulse control, while failing to inhibit indicates poor impulse control (Logan, Schachar, & Tannock, 1997). Poor inhibitory control may result from responding either too quickly to the go-signal or too slowly to the stop-signal. This is because fast responses to the go-signal would be executed before the participant could respond to the stop-signal, and slow responses to the stop-signal would allow normally speeded responses to the go-signal to escape from inhibition.

Task administration was computerised and presented on a Toshiba Satellite C660 laptop (Intel Core i5-2410M Processor, 15.6” Toshiba TruBrite HD TFT High Brightness display). Care was taken to ensure that participants understood the instructions. The experimental session began with a demonstration of the task (50 practice trials) which gave participants an opportunity to become familiar with the test while they simultaneously practised responding with the keyboard buttons. This first session was considered practice and data from it were not analysed. Participants were instructed to respond to the primary stimulus arrow (=green) as quickly and as accurately as possible while trying to inhibit their responses to the red arrow, whenever this was presented. However, they were instructed not to wait for the stop signal as these occurred randomly. They had to respond to a left or right pointing arrow by a left (‘z’ on the computer keyboard) or right button (‘?’ on the computer keyboard) press. The task lasted about 10 minutes.
Data for each participant were collapsed across two sessions and entered into the analyses described in Chapter 8. Participants received two blocks of 100 experimental trials each. The stop-signal occurred occasionally on 25% of the go-task trials, i.e. 50 trials in total, and involved the change of the arrow colour from green to red which signaled that participants should refrain from responding to the go-process on that trial. This low proportion of stop-signals is used by the paradigm to avoid the activation of unwanted strategies by participants, such as delaying the response to the go-signal to increase the likelihood of successful inhibits (Band, van der Molen, & Logan, 2003). The order in which trials were presented was randomised for each individual. They were told to inhibit their response if they could but not to worry and continue with the rest of the test if they were not able to inhibit in some trials. The time interval between the response and the next arrow onset on the subsequent trial was 1000 ms. Once started, the programme ran continuously for the first 100 trials. Participants had some time to rest in-between the two sessions.

When a stop-signal was presented, this happened unpredictably and with variable delays (SOAs) that ranged from 0 to 900 ms after the onset of the arrow for that trial. The SOAs were selected automatically by the built-in algorithm to ensure that 50% of the time participants would inhibit their responses to the stop-signal (the idea was first introduced by Osman, Kornblum & Meyer, 1990). The manipulation of SOAs is a necessary design feature because failing to trigger inhibition processes means failing to engage participants. This, in turn, would make the paradigm unreliable because it would let the primary task win the race on a subset of signal trials as if stop processes were not executed at all (Band, van der Molen, & Logan, 2003). Therefore, the SOA represents the amount of handicapping necessary to force a ‘tie’ finish between response execution and response inhibition. At that delay the two processes are assumed to finish at the same time, on average. SOAs were set dynamically by the in-built algorithm that tracked the individual’s response rate, i.e. their performance on
The main dependent variables are the mean (SSRT) and the mean Go-RT, reflecting the latencies of the internal response to the stop signal and to the go-signal respectively (Logan, Schachar, & Tannock, 1997).

Scoring method

a) In the go-trials, i.e. when the green arrow was presented and there was no stop-signal presented, responses were codified as follows:

- Condition 1 (correct response): participants correctly responded to the green arrow, by pressing correctly the left or the right button: GLL (Green: left presented-left pressed), GRR (Green: right presented-right pressed)
- Condition 2 (incorrect response): participants correctly responded to the green arrow, but failed to correctly press direction button: GLR (Green: left presented-right pressed), GRL (Green: right presented-left pressed)
- Condition 3 (incorrect response): participants failed to respond to the green arrow.

b) In the stop-signal trials, i.e. when the red arrow appeared, responses were codified as follows:

- Condition 1 (correct inhibition): participants correctly inhibited their response by refraining from pressing any of the two buttons
- Condition 2 (incorrect inhibition): participants failed to inhibit their response, by pressing either of the two buttons.
The data on the distribution on the Go-RTs, SSRTs, and their corresponding error rates, along with the statistical analyses, appear in Chapter 8 that analyses the participants’ data.

7.7.2 Verbal tasks and WM assessments

The capacity of the phonological loop or PSTM was tested via the forward digit span test (Wechsler, 1991) that requires a storage capacity as well as via two nonword repetition measures, the Greek- and the English-sounding tasks. These are most commonly used to test PSTM (Gathercole, 1999; Gathercole & Alloway, 2008; Morra & Camba, 2009).

7.7.2.1 Measures of PSTM: The forward digit span and recall test

More specifically, the forward digit span and recall test involves the presentation of spoken sequences of digits (a highly familiar verbal stimulus) for immediate serial recall. Participants must listen carefully to a series of digits which they need to repeat in the correct forward order. Items are presented at a rate of one digit per second. Following a practice session of two practice trials that are not scored, to make sure informants fully understand what the task demands, presentation begins with two digits in a series. Two trials are presented at each level of difficulty. If the sequence is reported correctly, the length of the next sequence is increased by one digit. Thus, the level of difficulty gradually increases reaching a maximum of nine digits in overall eight trials. The test is discontinued when both trials at a given level are incorrectly recalled. The total number of successful responses were taken into account as this scoring method is considered to be more accurate (Μασούρα, Gathercole, & Μπαμπλέκου, 2004). One point was allocated for every successful response and half a point for a partially given correct response (right digits, wrong order). This was considered a valid adjustment of the test, given the very young age of the participants. Maximum total score: 16 points (See Appendix 1).
**7.7.2.2 Measures of PSTM: The two nonword repetition tests**

Research with normal children has also established that nonword repetition is a reliable measure of PSTM capacity and a good predictor of a number of language skills during the early school years (Gathercole & Adams, 1994; Gathercole & Baddeley, 1993; Gathercole et al., 1992; for a review see Baddeley, Gathercole, & Papagno, 1998).

Nonword repetition tasks involve the spoken presentation (via the laptop) of nonwords, artificial words of two to five syllables, that test informants’ sensitivity at the segmental level. These words, although they conform to the phonological structure of a language are non-existent, thus void of meaning. Nonword repetition tests are often used as predictors of vocabulary learning (Gathercole et al., 1992). According to Baddeley, Thomson and Buchanan (1975), nonword repetition accuracy is a good measure of the capacity of the phonological loop. Due to the absence of any stored lexical specification of their phonological structure, informants need to rely heavily on the phonological representation of the nonword in the loop to successfully repeat each artificial word. What is documented is that nonword repetition tasks yield stronger relations to vocabulary scores than digit span tasks (Gathercole, 1999) because they involve the sequencing of highly unfamiliar items. In this sense, they are much closer to the situation the FL learner faces in the beginning, where all verbal material is not yet known (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998).

Two non-word repetition tasks were given to the informants: a) *The Children’s Test of Nonword Repetition* (CNRep) (Gathercole & Baddeley, 1996). Examples of nonwords include *woogalamic* and *skiticult* (see Appendix 2), and b) *The Test of Nonword Repetition for Greek-speaking children* (NRGreek) (Maridaki-Kassotaki, 1998). Examples of nonwords include *χνουρεσέπαδι* and *καταφάσοτα*.* Both consist of 40 nonwords (10 of each case, i.e. words that contain two, three, four, and five syllables) which are auditorily presented to the informants (see Appendix 3).
Procedure and scoring method

The nonwords presented are phonologically legal sequences of phonemes in Greek and English. Although unfamiliar, they are composed to some degree of phonological segments shared by words likely to be familiar to children, at least for the Greek-sounding test. The stress structure of each syllable within each nonword is constructed to correspond to the syllable stress structure of the syllables of Greek and English words of corresponding length respectively.

The test material was digitally recorded in the Phonetics Laboratory of the School of English. The Greek version was recorded by the researcher while the English one by a native speaker of British English, who had previously studied the phonetic transcriptions of the English nonword repetition test (see Appendix 2). During recordings he took care to articulate all sounds, including word-final consonants. The nonwords were presented to participants auditorily. They were given the following instructions in Greek: “You will soon hear some very strange words, as if coming from a fairy tale. Please repeat each word immediately after you hear it”. The recordings were listened to via Media Player and the participants were wearing earphones. The informants listened to the words once and had 5 seconds to repeat each nonword with full accuracy (all phonemes correct). Their responses were recorded with the help of a Digital Voice Recorder (Olympus DS-55) and were coded and scored objectively by the two NSs involved in this test, the researcher who scored the repetition responses to the Greek-sounding test and the English NS who rated the accuracy of the repetition responses to the English-sounding test. Both scorers followed the same scoring method: each nonword that was correctly repeated was given a point. Phonological deviations from the target form (addition, deletion, or replacement of a phoneme) were valued as incorrect and got no point (Archibald & Gathercole, 2006). In the cases where it was obvious from the spontaneous speech of some children that they tended to misarticulate certain phonemes consistently,
answers that contained these specific phonemes were scored as correct (Μασοώρα, Gathercole, & Μπαμπλέκου, 2004). Emphasis was given on the children’s ability to reproduce the sounds of the nonwords as accurately as possible and not on their ability to mimic the English pronunciation as such. The maximum total score in either test was 40 points.

7.7.2.3 Measures of CWM

CWM (or the central executive) capacity was tested by means of two measures, the *Backwards digit recall task* (Kormos & Sáfár, 2008; St. Clair-Thompson, 2010; Wechsler, 1991) and the *Listening and Recall test*, based on the listening span procedure originally developed by Daneman and Carpenter (1980). Both tasks combine processing (reversing the digit sequence or linguistic analysis of each sentence) with concurrent storage (of the digits or of the lexically unrelated word).

7.7.2.3.1 The backwards digit span and recall test

*The backwards digit span and recall test* imposes a substantial WM load on children (St. Clair-Thompson, 2010) as it involves both the central executive and the phonological loop (Gathercole, 1999). The test employs the same procedure as the forward condition in all respects except that participants are now required to re-sequence the spoken numbers by recalling them in the reverse order (Alloway & Alloway, 2010; Gathercole, 1999; Nevo & Breznitz, 2011). This transposition of order requires the involvement of executive-attention resources (St. Clair-Thompson, 2010). To give an example, the sequence 8, 5, 2 would be correctly recalled as 2, 5, 8. Two practice trials were given to ensure the child understood the concept of ‘reverse’. There is a maximum of eight digits in overall seven trials. The test is run and scored the same way as with the forward digit condition. Maximum total score: 14 points. (see Appendix 1).
7.7.2.3.2 The Listening span and Recall task

The Listening span and Recall Task (adapted) is a dual measure of WM for language, which tests both the central executive and WM capacity (Baddeley, 1996). The test requires the active maintenance of information in the face of concurrent processing and interference and thus recruits an executive attention-control mechanism to combat interference (Conway, Kane, & Engle, 2003). In this sense, performance on this task is an index of the ability to focus, divide, or maintain attention in the face of irrelevant stimuli (Dempster, 1991).

This is a modified version of the listening span test administered by Χρυσοχόου, (2006) as it requires the recall of semantically and phonologically unrelated word\(^{1}\) and not the last word of each sentence (see Appendix 4, Chapter 7). Considering the small age of the participants, this was thought a valid adjustment to make, as similarity at the phonological, orthographic, or semantic level tends to negatively affect the retrieval of the items to be recalled (Papagno & Vallar, 1992).

The input at all times was read aloud by the researcher. Informants were told they would hear a simple sentence followed by a pause and then a word that is lexically unrelated to the preceding sentence. Then, they would need to do two things: perform a semantic judgement on the sentence by replying ‘true’, when the sentence reflected an event that could happen in real life and ‘false’ otherwise. In the meantime, they needed to temporarily maintain the corresponding word in mind and provide this at the end of the set. Participants were told they would be presented with increasingly longer sets of sentences. One practice trial was given to them at the one-sentence level, consisting of a single-sentence block, to familiarise them with the task demands. None of the practice items appeared in the experimental list.

\(^{1}\) The list was designed by Tsimpli and Peristeri (personal communication).
Chapter 7: The Present Study

The test consists of six sets, i.e. difficulty levels (Daneman & Carpenter, 1980). The number of sentences in a set is incremented from level to level. More specifically, Level 1 consists of six 1-sentence trials, each followed by a lexically unrelated word. Level 2 consists of six 2-sentence trials, etc. until Level 6 is reached. On each trial, the informant listened to a series of sentences, judged their veracity and then recalled the lexically unrelated words, at the end of each set. Because of the difficulty of the test, correct serial word order was not required. All the answers provided (correct, incorrect, no answer) were recorded for further scoring. Permission to go to the next level was granted only when four (not necessarily consecutive) correct trials were given within each level (four XX responses, see the Scoring method below). Testing was terminated when informants failed to perform accurately three times within a level, either in one or in both tasks (i.e. semantic judgement or word-recall).

The processing requirements of the task are complex as it becomes more and more difficult and demanding for the informants, by placing a lot of burden onto their WM span. This can decrease the amount of additional information that can be maintained (Daneman & Carpenter, 1980).

Scoring method

Participants’ trials were codified as follows:

a) XX: the informant correctly judged the veracity of the sentence and recalled the unrelated word

b) XO: the informant correctly judged the veracity of the sentence but either failed to recall the word or provided a completely different one,

c) OX: the informant failed to judge the veracity of the sentence correctly but succeeded in the word recall,

d) OO: the informant failed in both tasks.
Then, percentages were assigned on each trial within a set. For example, if an informant was successful in the first four attempts, then the XX condition was given a 100% and (s)he would be granted permission to go to the next level. If a participant gave three XX answers, two OX answers and one XO answer, then the scoring went as follows: XX answers counted as 50%, OX answers as 33.33%, and the XO answer as 16.67%. In this last example, permission would not be granted to the next level, as the XX answers were less than four.

Scoring in complex span tasks, such as this CWM test, is complicated as the processing and the storage component jointly contribute to measurement (Conway et al., 2005). A partial credit scoring than the overall scoring method just explained has been shown to be a superior and more objective evaluation method (Juffs & Harrington, 2011). In the following chapter we examined both sources, following two different scoring methods, the one just mentioned and a second that follows. In all the correlations and the analyses of variance (ANOVAs) conducted, the ability of the participants to correctly recall the word (i.e. their storage capacity) was taken into account. Irrespective of their performance on the semantic evaluation of the sentence, participants were given a point for every correctly recalled word. With four such cases in a level, 6 points were assigned to them. It has to be noted that for levels 2, 3 and 4, participants had to correctly recall a corresponding number of 2, 3 or 4 words in each attempt, to be credited with a point.

7.7.2.4 Measures of L1 verbal intelligence

Following the Linguistic Coding Hypothesis of Sparks and Ganschow (1991), according to which native skills serve as the foundation for FLL, we also decided to measure the L1 verbal intelligence of participants. This was tested by the relevant sub-tests of the two versions of the Diagnostic Test of Verbal Intelligence (DVIQ) (Stavrakaki & Tsimpli, 2000). DVIQ I has been standardised on children between the ages of 3.5 - 6.5 (years, months). It assesses the linguistic abilities of pre-schoolers with respect to their productive and receptive skills
(vocabulary, morphology and metalinguistic concepts). The morphosyntax sub-tests were excluded as they fell outside the scope of this research. It was considered necessary to run DVIQ I because during the first year of the research some of the children fell on the borderline (6-6.5 years of age). DVIQ II assesses the linguistic performance of school-age children (ages 6.5 – 9.5 or from Grade 1 to Grade 3). Again, only the sub-tests that pertain to vocabulary and metalinguistic knowledge were administered to participants. Metalinguistic analysis is related to acceptability, synonymity and ambiguity of words and sentences and as such it falls closer to the building of vocabulary (Service, 1992). Children were credited with one point for each correctly identified or produced item. Overall score: 67 points:

a) DVIQ I (VOC) Vocabulary sub-test measures the vocabulary of pre-schoolers and their productive skills with respect to nouns and verbs. Informants were presented a series of 27 pictures and were asked to name the object or action taking place in each picture. Examples: i) What is that? (The picture shows a thermometer), ii) What are the mother and the boy doing? (The picture shows a boy with his mother buying fruit in an open market). One point was given for every correct answer. Zero point was assigned when no response was given or in the case of an incorrect answer. Score: 27 points. In Appendix 5 the interested reader will see the list of the items tested, both in Greek and in English. This also holds for the other two sub-tests that follow. Even though the test is in Greek, the English translation was done by the author of this thesis to facilitate the understanding of readers that may not speak the Greek language.

b) DVIQ I (MET) Metalinguistic concepts sub-test measures the receptive skills of pre-schoolers. The informants were presented a series of 25 pictures that examine perception of: a) time sequence: E.g. Point to the cat first and then to the dog, b) either-or relations: E.g. Point to either the cat or the dog, c) spatial relations: E.g. Point to the elephant that is next to the horse, d) size: E.g. I will point to a big duck and you will point to the small...
fox, etc. The children were asked to point to the picture that corresponded to the word spoken by the experimenter. Same assignment of scoring, as above. Total score: 25 points (see Appendix 6).

c) **DVIQ II (MET) Metalinguistic concepts** sub-test measures children’s productive and receptive skills. It involves a metalinguistic analysis and processing of the verbal material presented. The informants were presented 15 pictures that examine perception of: a) negativity: E.g. *Do not point to the square that is black*, b) quantitative markers: e.g. *Point to some squares that are small*, c) temporal markers: e.g. *Before pointing to a small square, point to a big one*, d) place adverbials: e.g. A picture is shown. The researcher uses a sentence that, for instance, best describes the place adverbial used in the picture. The child is asked to partly rephrase it by using a place adverbial that denotes the same meaning. Same assignment of scoring as above. Total score: 15 points (see Appendix 7).

### 7.7.2.5 The English vocabulary test

A vocabulary test in English was also given to the informants of the experimental school only, right before the end of their second year of English schooling. It is reminded to the reader that the control group had not had any exposure to the English language yet.

A pilot test was administered to eight students in total, to see whether the sub-tests (receptive, productive) designed were too easy, difficult, or perhaps too confusing and accordingly make the necessary modifications. Half of them had been valued as ‘high achievers’ by their English teachers, while the other half was the ‘low achievers’ group. The two pilot tests consisted of 5 items for every skill, 10 in total. None of the items of the pilot tests appeared in the experimental tests that followed. Instructions on what each test was about were given in Greek.

The reader should be reminded of the fact that the experimental school follows Asher’s (1982) TPR teaching method and Krashen and Terrell’s Natural Approach (1983). Both
methods were considered fit as they are communicatively oriented and associated with the long-term retention of new linguistic material, a better understanding of novel utterances and the development of positive attitudes in learners towards language learning (Ellis, 1994). According to the former method, in the initial stages children learn the language through play activity. They are expected to develop their comprehension skills and they only start producing language when they are well into the course. The basic premise of the latter teaching method is that learners learn to communicate in the FL by means of real communicative episodes embedded in a functional-notional syllabus that contains a series of topics that are of interest to the students, supplemented by games, tasks and other activities which all provide comprehensible input. Thus, in the earliest years of FLL the focus rests on the development of learners’ fluency, i.e. on their speaking and listening skills and their ‘oral competency in communication’ (Terrell, 1977: 326).

The test comprised two parts: one examining informants’ receptive skills, the other their productive skills. They measured the knowledge learners had gained in the course of the two years the study lasted and thus the material used was based on the material covered that far. Among the short stories or songs that learners were exposed to during the two years, were Winnie in Winter, Itsy Bitsy Spider, Goldilocks and the three bears, Hetty and the lion, Five little squirrels, Meg and Mog, etc. What follows is only some of the thematic fields covered in Grades 1 and 2 by the school curriculum that refer to children’s environment at home, outside the home and in school: family, friends, feelings, main numerals, the school, clothes, body parts, animals, the four seasons, Christmas, main colours, food, means of transportation, sports, as well as action verbs such as stand, sit, walk, turn around found in lexical rote-learned chunks of high frequency such as Point to (your ears), Sit down, or formulaic expressions such as How are you? I’m fine, thanks, etc. The two tests were designed in a way to tap on the same thematic fields and tested whether participants could recall and produce a
number of, mostly, nouns (Szpotowicz, 2010), action verbs (in the present progressive form), and adjectives they had been taught (Okalidou et al., 2011). To make it clearer, if the child was asked to identify by pointing to a means of transport (e.g. car) in the receptive test, (s)he was also asked to name a means of transport (e.g. an airplane) in the productive test.

The test was computerised (see Appendices 8 & 9 for the receptive and productive subtests, respectively). The receptive task was implemented first and comprised 15 slides. For each target word, the child was shown an array of maximum three pictures. Considering the small age of the informants, the slides were designed in a rather simple manner: the pictured objects or actions were all familiar to them, being part of the material covered. The pictures were not very ‘crowded’ as we did not wish to cause any confusion in learners with the presentation of overloaded slides. There is always the danger of some students concentrating on aspects that are irrelevant to the task demands, failing in this way to provide the proper answer even if they know it. Children were asked to point to the picture that was being named (e.g. Point to the yellow pencil) or to do exactly what they were told (e.g. Show me your fingers or Open the door). A digital stopwatch was used to record the time each participant needed to complete the test. For each individual a scoring sheet was kept in which the following information was entered: duration of the test (minutes, seconds) and responses to each item, noting whether the response was correct, incorrect or not given. Maximum score: 15 points.

The students’ productive vocabulary was tested second. 15 slides displayed a number of different objects, actions, etc. The researcher asked 15 questions and tried to elicit as many answers as possible, with the help of the pictured objects or actions included in the slides. 11 questions elicited a corresponding number of individual words, each credited with one point (e.g. What is this? A horse). Four questions were meant to elicit chunks of words (e.g. Where
is the boy sitting? On the bed), which were awarded 2 points each. The same procedure was followed with respect to the scoring sheet of each individual. Maximum score: 19 points.

The chapter that follows contains a detailed description and statistical analysis of the results compiled from the 16 tests taken for the purposes of this research.

7.8 Chapter Summary

The research of this study lasted two academic years, i.e. 2010-11 and 2011-12. The thesis seeks to explore the cognitive impact of the early introduction of English as a FL on young Greek learners. It is reminded to the reader that the experimental group which took part in the study followed a two-year intensive FL programme (5 hours per week) from the age of 6 (Grade 1), while the control group belonged to a mainstream primary school that introduces English from Grade 3.

The study attempts to explain individual differences in the acquisition of L2 English vocabulary. More specifically, it aims to explore whether the early exposure to EFL may positively affect the cognitive functioning of young learners with respect to their FL aptitude and the EFs of WM and Response Inhibition. To the best of my knowledge, the cognitive aspect of EFL has never before been examined as the relevant studies so far have primarily focused on the linguistic and affective outcome of the early teaching of a FL.
8.1 The testing waves in the two schools

As previously mentioned in Chapter 7, the experimental school was visited first for the data collection in Grade 1, while the data from the control school was collected immediately after Christmas vacation, i.e. when permission was granted by the Pedagogical Institute of Athens. Table 3 displays in detail the visiting periods of the two testing waves in both schools:

**Table 3: The testing waves in both schools**

<table>
<thead>
<tr>
<th>Testing waves</th>
<th>Experimental school</th>
<th>Control school</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From</td>
<td>To</td>
</tr>
<tr>
<td>First testing wave (Grade 1)</td>
<td>11/10/2010</td>
<td>22/11/2010</td>
</tr>
<tr>
<td>Second testing wave (Grade 2)</td>
<td>22/02/2012</td>
<td>26/03/2012</td>
</tr>
<tr>
<td>Third testing wave (exp. school, Grade 2)</td>
<td>01/05/2012</td>
<td>10/05/2012</td>
</tr>
</tbody>
</table>

The third testing wave that took place towards the end of Grade 2 concerns only the experimental school and the implementation of the English Vocabulary Test, designed by the researcher to test the productive and receptive skills of the students on the material covered during the past two academic years. Once the results of the tests were codified, analyses were conducted by using the Statistical Package for the Social Sciences, versions 18 and 21 for Windows software.
8.2 Elimination of outliers

To find the outliers in the two schools, the means and standard deviations (SDs) of all the tests were first computed separately for Grades 1 and 2. Two participants coming from the experimental group scored more than 2 SDs above or below the group mean, in more than 35% of the tasks in both grades. On this basis, their scores were excluded from all subsequent analyses. This resulted in the lowering of the initial number of the experimental group from 49 to 47. Accordingly, the mean age and SD for Grade 1 changed into Mean 6 years, 3 months, SD 3 months; range: 5 years, 10 months to 6 years, 9 months. Grade 2: Mean 7 years, 8 months, SD 3 months; range: 7 years, 2 months to 8 years, 2 months. After the exclusion of the two participants, the means and SDs of all the tests were re-computed for Grades 1 and 2. For the rest of the participants who scored more than 2SDs above or below the group mean in a couple of tests, these scores were replaced by the mean score(s) of the test(s).

In the case of the SSRT, an additional participant was excluded from the experimental school, changing the number of this group to 46. For more details, the interested reader should go to section 8.12, which discusses in great detail the exclusionary criteria followed (Huizinga, 2006) and the findings coming from this test.

8.3 Descriptive Statistics

Initially, descriptive statistics were performed to explore whether there were significant differences between the mean scores of the two schools. Tables 4 and 5 display the mean scores and SDs for each task in Grade 1 and Grade 2, respectively. These were computed with two decimal places in all cases.
### Table 4: Descriptive Statistics - Means and Standard Deviations (SDs), Grade 1

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Experimental school (n=47)</th>
<th>Control school (n=49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRADE 1</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Digit recall_Forward</td>
<td>6.13</td>
<td>1.64</td>
</tr>
<tr>
<td>Nonce_Gr.</td>
<td>34.04</td>
<td>3.61</td>
</tr>
<tr>
<td>Nonce_Eng.</td>
<td>25.78</td>
<td>4.96</td>
</tr>
<tr>
<td>Digit recall_Backwards</td>
<td>3.07</td>
<td>0.81</td>
</tr>
<tr>
<td>Listening Recall</td>
<td>5.67</td>
<td>2.72</td>
</tr>
<tr>
<td>SSRT (Stop-signal)</td>
<td>309.68</td>
<td>49.62</td>
</tr>
<tr>
<td>DVIQ I_Voc.</td>
<td>25.11</td>
<td>1.43</td>
</tr>
<tr>
<td>DVIQ I_Met.</td>
<td>22.91</td>
<td>1.17</td>
</tr>
<tr>
<td>DVIQ II_Met.</td>
<td>13.11</td>
<td>1.18</td>
</tr>
<tr>
<td>Memory</td>
<td>9.17</td>
<td>1.96</td>
</tr>
<tr>
<td>Classification</td>
<td>35.28</td>
<td>13.51</td>
</tr>
<tr>
<td>Differences</td>
<td>10.69</td>
<td>2.83</td>
</tr>
<tr>
<td>Paired Associates</td>
<td>6.79</td>
<td>3.90</td>
</tr>
<tr>
<td>Semantic Integration</td>
<td>12.99</td>
<td>1.85</td>
</tr>
<tr>
<td>Story Sequences</td>
<td>3.98</td>
<td>2.27</td>
</tr>
<tr>
<td>YLAT_Total</td>
<td>80.69</td>
<td>14.66</td>
</tr>
</tbody>
</table>
Table 5: Descriptive Statistics - Means and Standard Deviations (SDs), Grade 2

<table>
<thead>
<tr>
<th>GRADE 2</th>
<th>Experimental school (n=47)</th>
<th>Control school (n=49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tasks</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>FL_Voc._Total</td>
<td>24.69</td>
<td>5.43</td>
</tr>
<tr>
<td>FL_Receptive</td>
<td>12.82</td>
<td>1.73</td>
</tr>
<tr>
<td>FL_Productive</td>
<td>12.15</td>
<td>3.58</td>
</tr>
<tr>
<td>Digit recall_Foward</td>
<td>7.05</td>
<td>1.46</td>
</tr>
<tr>
<td>Nonce_Gr.</td>
<td>34.99</td>
<td>3.59</td>
</tr>
<tr>
<td>Nonce_Eng.</td>
<td>28.51</td>
<td>3.96</td>
</tr>
<tr>
<td>Digit recall_Backwards</td>
<td>4.31</td>
<td>1.10</td>
</tr>
<tr>
<td>Listening Recall</td>
<td>7.05</td>
<td>3.04</td>
</tr>
<tr>
<td>SSRT (Stop-signal)</td>
<td>275.19</td>
<td>55.69</td>
</tr>
<tr>
<td>DVIQ I_Voc.</td>
<td>25.50</td>
<td>1.26</td>
</tr>
<tr>
<td>DVIQ I_Met.</td>
<td>22.91</td>
<td>1.36</td>
</tr>
<tr>
<td>DVIQ II_Met.</td>
<td>12.72</td>
<td>1.25</td>
</tr>
<tr>
<td>Memory</td>
<td>11.09</td>
<td>2.50</td>
</tr>
<tr>
<td>Classification</td>
<td>42.88</td>
<td>7.08</td>
</tr>
<tr>
<td>Differences</td>
<td>12.77</td>
<td>3.02</td>
</tr>
<tr>
<td>Paired Associates</td>
<td>10.35</td>
<td>2.24</td>
</tr>
<tr>
<td>Semantic Integration</td>
<td>13.97</td>
<td>2.12</td>
</tr>
<tr>
<td>Story Sequences</td>
<td>4.45</td>
<td>1.22</td>
</tr>
<tr>
<td>YLAT_Total</td>
<td>95.72</td>
<td>9.34</td>
</tr>
</tbody>
</table>

1 The dashes indicate that the control group did not take the FL Vocabulary Test for reasons already explained.
What follows is an extensive discussion of the raw data collected in Grades 1 and 2 from both schools and the statistical tests conducted.

8.4 Research Methodology

First, we had to test the null hypothesis to examine whether the two groups displayed any marked differences before the FL intervention on the experimental group, i.e. at the beginning of Grade 1. If no differences would emerge, then the null hypothesis would be accepted. In the opposite case it would be rejected. This was considered a necessary step to take, to be able to track any differences that would emerge in the performance of the experimental group by the end of Grade 2. In other words, we needed to know what the starting point was for the experimental group before the FL intervention, to be able to compare this with the group’s performance after their two-year FL schooling. To this end, we conducted a number of independent and paired samples t-tests as well as univariate analyses of variance. Findings from the t-tests and the two-way ANOVAs would answer Research Question One that pertained to the cognitive impact of EFLL.

Answering Research Question Two would shed light on whether variables other than the often suggested PSTM can predict FL performance. First, we had to explore and define the relations between the principal measures. To this end, we conducted a number of correlation analyses. Then, to define the direction of causality we conducted a number of linear or stepwise regression analyses, the findings of which would help us answer the second research question that pertained to the predictive value of the other variables of this study, other than PSTM, that could explain FL performance.

Research Question Three investigates whether young children who are formally exposed to an intensive L2 English programme from the age of 6 enjoy any of the cognitive flexibility witnessed in bilingual children, with respect to their Inhibition mechanism. It was hypothesised that the early and intensive introduction of the FL would help young learners
have a better grasp of their inhibition mechanism. To this end, we calculated separately the correlation coefficients of the SSRT with the other variables under investigation. This was considered necessary as after the additional exclusionary criterion applied for this test one more participant was excluded from the experimental group, changing the number of the group to 46. Several regression analyses were then conducted to explore whether the index of the SSRT explained the learners’ performance in any of the other tasks. Third, we applied a number of multivariate analyses of variance (MANOVAs) to investigate whether the two dependent variables (i.e. Go_RT and SSRT) display any marked differences between the two groups across the two grades. Fourth, we applied repeated measures ANOVAs to see if any marked differences would emerge with respect to response activation and response inhibition. The paired samples t-tests would show whether there was an effect of age on the SSRT\textsubscript{mean} while the independent samples t-tests would help us answer the question whether EFLL would have any effect on the learners’ control of their response inhibition in Grade 2.

To be able to explore Research Question Four that pertains to whether EFLL is associated with a gender effect and the better performance of one sex over the other, we first applied a number of two factor independent measures ANOVA, in both grades and groups. We examined the effect of two independent variables, group-grade and group-sex, on one dependent variable, i.e. the score of each individual measure. In addition, for Grade 2 we performed several MANOVAs to see whether a school and/or gender effect would emerge in Grade 2 (after the FL intervention in the experimental group) with respect to sets of tests that examine the same skill: PSTM, the two EFs, namely CWM and Inhibition, L1 verbal intelligence, and FL aptitude. These two different kinds of statistical analyses would give us two answers. First, whether one school was better than the other in the individual tests taken. Second, whether one sex in either or both schools scored significantly higher than the other in groups of tests that tap the same skill. Based on the literature review conducted earlier, our
hypothesis was that overall, the experimental group would fare better than the control, at least in the PSTM and possibly in CWM tests. As for the gender effect, this was to be seen, as research findings suggest an overall heightened motivation of girls over boys and a more positive attitude towards FLL. Nevertheless, it should be mentioned that studies that pertain to young learners’ performance in CWM tasks, Inhibition tasks and non-verbal cognitive ones (Alexiou, 2005; for a review see Brocki & Bohlin, 2010; Monette, Bigras, & Guay, 2011) provide no such theoretical support for us to expect any gender effect to emerge.

8.5 The independent samples t-tests

8.5.1 Grade 1

At this initial stage of the research, the two groups were considered a homogeneous whole. The null hypothesis predicts that no differences were expected to emerge as all informants were of the same age and socio-economic status, while the experimental group was not yet exposed to EFL. Still, it would be interesting to see whether differences did exist in the means of the two samples before the FL intervention in the experimental group. To this end, we conducted independent samples t-tests for each test, in both grades. Interestingly enough, various significant differences emerged for the experimental group in the majority of the tests, disproving thus the null hypothesis. This was viewed as accidental since both schools, as already mentioned, were located in the same area and consequently addressed the same local residents. It should be noted that the children who attend the experimental school do not undergo a particular exclusionary procedure upon 1st grade enrollment. Taylor and Lafayette (2010) also reported such a dissimilarity between their two groups, even though every effort was made to match the treatment and the control group and thus minimise the risk of attributing differences in their performance to differences that pertain to group characteristics. Table 6 below gives a detailed picture of the performance of the two groups in the battery of tests for Grade 1:
# Data Analysis

## Chapter 8: Data Analysis

### Table 6: Independent samples t-tests, Grade 1

<table>
<thead>
<tr>
<th>Skill tested</th>
<th>Measures</th>
<th>Group</th>
<th>Mean</th>
<th>t-test</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSTM</strong></td>
<td>Digit recall_Forward</td>
<td>Experimental</td>
<td>6.13</td>
<td>(p = \text{ns})</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>5.70</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Nonce_Gr.</td>
<td>Experimental</td>
<td>34.04</td>
<td>(p = .000)</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>30.14</td>
<td></td>
<td>5.44</td>
</tr>
<tr>
<td></td>
<td>Nonce_Eng.</td>
<td>Experimental</td>
<td>25.78</td>
<td>(p = .007)</td>
<td>4.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>22.56</td>
<td></td>
<td>6.36</td>
</tr>
<tr>
<td><strong>CWM</strong></td>
<td>Digit recall_Backwards</td>
<td>Experimental</td>
<td>3.07</td>
<td>(p = .032)</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.45</td>
<td></td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Listening Recall</td>
<td>Experimental</td>
<td>5.67</td>
<td>(p = \text{ns})</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>4.73</td>
<td></td>
<td>2.32</td>
</tr>
<tr>
<td><strong>Inhibition control</strong></td>
<td>SSRT</td>
<td>Experimental</td>
<td>309.68</td>
<td>(p = \text{ns})</td>
<td>49.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>298.45</td>
<td></td>
<td>58.61</td>
</tr>
<tr>
<td><strong>L1 verbal intelligence</strong></td>
<td>DVIQ I_Voc.</td>
<td>Experimental</td>
<td>25.11</td>
<td>(p = .006)</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>24.23</td>
<td></td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>DVIQ I_Met.</td>
<td>Experimental</td>
<td>22.91</td>
<td>(p = .000)</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>21.15</td>
<td></td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>DVIQ II_Met.</td>
<td>Experimental</td>
<td>13.11</td>
<td>(p = .000)</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>12.10</td>
<td></td>
<td>1.52</td>
</tr>
<tr>
<td><strong>FL aptitude</strong></td>
<td>Memory</td>
<td>Experimental</td>
<td>9.17</td>
<td>(p = \text{ns})</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>8.94</td>
<td></td>
<td>2.54</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td>Experimental</td>
<td>35.28</td>
<td>(p = \text{ns})</td>
<td>13.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>31.69</td>
<td></td>
<td>15.51</td>
</tr>
<tr>
<td></td>
<td>Differences</td>
<td>Experimental</td>
<td>10.69</td>
<td>(p = .019)</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>9.30</td>
<td></td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>Paired Associates</td>
<td>Experimental</td>
<td>6.79</td>
<td>(p = \text{ns})</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>6.44</td>
<td></td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>Semantic Integration</td>
<td>Experimental</td>
<td>12.99</td>
<td>(p = \text{ns})</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>12.52</td>
<td></td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>Story Sequences</td>
<td>Experimental</td>
<td>3.98</td>
<td>(p = .049)</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.11</td>
<td></td>
<td>2.01</td>
</tr>
<tr>
<td><strong>FL aptitude_Total</strong></td>
<td>YLAT_Total</td>
<td>Experimental</td>
<td>80.69</td>
<td>(p = .027)</td>
<td>14.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>73.02</td>
<td></td>
<td>18.53</td>
</tr>
</tbody>
</table>
What is clear from Table 6 is that the experimental group was better than the control in half of the measures taken, with statistically significant differences emerging in 8 out of the 16 tasks. The control group ($M = 3.45$, $SD = 0.89$) scored significantly higher than the experimental ($M = 3.07$, $SD = 0.81$) only in Digit recall_Backwards, ($t(94) = -2.176$, $p < .05$). No significant differences emerged between the two groups for the following measures: Digit recall_Forward, Listening Recall, SSRT, Memory, Classification, Paired Associates, and Semantic Integration. The Levene’s test for equality of variances showed that equal variances were not assumed for Nonce_Gr., DVIQ I_Met., and YLAT_Total (see Appendix 1).

The experimental group fared better in the following measures:

1. Nonce_Gr.: Experimental group ($M = 34.04$, $SD = 3.61$), Control group ($M = 30.14$, $SD = 5.44$), $t(94) = 4.126$, $p < .001$

2. Nonce_Eng.: Experimental group ($M = 25.78$, $SD = 4.96$), Control group ($M = 22.56$, $SD = 6.36$), $t(94) = 2.761$, $p < .01$

3. DVIQ I_Voc.: Experimental group ($M = 25.11$, $SD = 1.43$), Control group ($M = 24.23$, $SD = 1.59$), $t(94) = 2.839$, $p < .01$

4. DVIQ I_Met.: Experimental group ($M = 22.91$, $SD = 1.17$), Control group ($M = 21.15$, $SD = 1.93$), $t(94) = 5.428$, $p < .001$

5. DVIQ II_Met.: Experimental group ($M = 13.11$, $SD = 1.18$), Control group ($M = 12.10$, $SD = 1.52$), $t(94) = 3.617$, $p < .001$

6. Differences: Experimental group ($M = 10.69$, $SD = 2.83$), Control group ($M = 9.30$, $SD = 2.86$), $t(94) = 2.397$, $p < .05$

7. Story Sequences: Experimental group ($M = 3.98$, $SD = 2.27$), Control group ($M = 3.11$, $SD = 2.01$), $t(94) = 1.995$, $p < .05$

8. YLAT_Total: Experimental group ($M = 80.69$, $SD = 14.66$), Control group ($M = 73.02$, $SD = 18.53$), $t(94) = 2.253$, $p < .05$. 
8.5.2 Grade 2

Independent samples $t$-tests were performed for the same battery of tests in Grade 2. Table 7 below illustrates the results:

Table 7: Independent samples $t$-tests, Grade 2

<table>
<thead>
<tr>
<th>Skill tested</th>
<th>Measures</th>
<th>Group</th>
<th>Mean</th>
<th>$t$-test</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSTM</td>
<td>Digit recall_Forward</td>
<td>Experimental</td>
<td>7.05</td>
<td>$p = .037$</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>6.45</td>
<td></td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>Nonce_Gr.</td>
<td>Experimental</td>
<td>34.99</td>
<td>$p = .026$</td>
<td>3.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>33.22</td>
<td></td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>Nonce_Eng.</td>
<td>Experimental</td>
<td>28.51</td>
<td>$p = .000$</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>23.86</td>
<td></td>
<td>4.69</td>
</tr>
<tr>
<td>CWM</td>
<td>Digit recall_Backwards</td>
<td>Experimental</td>
<td>4.31</td>
<td>$p = ns$</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>4.08</td>
<td></td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>Listening Recall</td>
<td>Experimental</td>
<td>7.05</td>
<td>$p = ns$</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>6.69</td>
<td></td>
<td>2.72</td>
</tr>
<tr>
<td>Inhibition control</td>
<td>SSRT</td>
<td>Experimental</td>
<td>275.19</td>
<td>$p = ns$</td>
<td>55.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>268.74</td>
<td></td>
<td>50.69</td>
</tr>
<tr>
<td>L1 verbal intelligence</td>
<td>DVIQ I_Voc.</td>
<td>Experimental</td>
<td>25.50</td>
<td>$p = ns$</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>25.70</td>
<td></td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>DVIQ I_Met.</td>
<td>Experimental</td>
<td>22.91</td>
<td>$p = .000$</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>20.89</td>
<td></td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>DVIQ II_Met.</td>
<td>Experimental</td>
<td>12.72</td>
<td>$p = .005$</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>11.85</td>
<td></td>
<td>1.63</td>
</tr>
<tr>
<td>FL aptitude</td>
<td>Memory</td>
<td>Experimental</td>
<td>11.09</td>
<td>$p = ns$</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>10.59</td>
<td></td>
<td>2.19</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td>Experimental</td>
<td>42.88</td>
<td>$p = ns$</td>
<td>7.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>40.93</td>
<td></td>
<td>11.21</td>
</tr>
<tr>
<td></td>
<td>Differences</td>
<td>Experimental</td>
<td>12.77</td>
<td>$p = ns$</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>12.53</td>
<td></td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>Paired Associates</td>
<td>Experimental</td>
<td>10.35</td>
<td>$p = .000$</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>8.18</td>
<td></td>
<td>3.18</td>
</tr>
<tr>
<td></td>
<td>Semantic Integration</td>
<td>Experimental</td>
<td>13.97</td>
<td>$p = ns$</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>13.39</td>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>Story Sequences</td>
<td>Experimental</td>
<td>4.45</td>
<td>$p = .000$</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.18</td>
<td></td>
<td>1.76</td>
</tr>
<tr>
<td>FL aptitude_Total</td>
<td>YLAT_Total</td>
<td>Experimental</td>
<td>95.72</td>
<td>$p = .008$</td>
<td>9.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>89.92</td>
<td></td>
<td>11.53</td>
</tr>
</tbody>
</table>
What is clear from the table above is that the experimental group maintained its initial advantage by scoring significantly higher in 8 out of the 16 measures taken. Even though in Digit recall_Backwards the control group lost its initial advantage from Grade 1 (\(M = 3.45, \ SD = 0.89\)) to Grade 2 (\(M = 4.08, \ SD = 1.06\)), no statistically significant difference emerged for the experimental group the second year, \((t(94) = 1.025, p = .308)\). Even so, it is still very important the fact that the experimental group managed to regain most of the ‘lost ground’ from Grade 1 (\(M = 3.07, \ SD = 0.81\)), by scoring impressively higher in Grade 2 (\(M = 4.31, \ SD = 1.10\)), marking thus an increase of 40.39% (see section 8.5.4 below). In addition, no significant differences were yielded between the two groups for Listening Recall, SSRT, DVIQ_I_Voc., Memory, Classification, Differences, and Semantic Integration. The Levene’s test for equality of variances showed that equal variances were not assumed for DVIQ_I_Met. and Story Sequences (see Appendix 1).

The experimental group fared significantly better in the following measures:

1. Digit recall_Forward: Experimental group (\(M = 7.05, \ SD = 1.46\)), Control group (\(M = 6.45, \ SD = 1.30\)), \(t(94) = 2.115, p < .05\)

2. Nonce_Gr.: Experimental group (\(M = 34.99, \ SD = 3.59\)), Control group (\(M = 33.22, \ SD = 4.05\)), \(t(94) = 2.259, p < .05\)

3. Nonce_Eng.: Experimental group (\(M = 28.51, \ SD = 3.96\)), Control group (\(M = 23.86, \ SD = 4.69\)), \(t(94) = 5.245, p < .001\)

4. DVIQ_I_Met.: Experimental group (\(M = 22.91, \ SD = 1.36\)), Control group (\(M = 20.89, \ SD = 2.09\)), \(t(94) = 5.619, p < .001\)

5. DVIQ_II_Met.: Experimental group (\(M = 12.72, \ SD = 1.25\)), Control group (\(M = 11.85, \ SD = 1.63\)), \(t(94) = 2.902, p < .01\)

6. Paired Associates: Experimental group (\(M = 10.35, \ SD = 2.24\)), Control group (\(M = 8.18, \ SD = 3.18\)), \(t(94) = 3.847, p < .001\).
7. Story Sequences: Experimental group ($M = 4.45$, $SD = 1.22$), Control group ($M = 3.18$, $SD = 1.76$), $t(94) = 4.107$, $p < .001$

8. YLAT_Total: Experimental group ($M = 95.72$, $SD = 9.34$), Control group ($M = 89.92$, $SD = 11.53$), $t(94) = 2.702$, $p < .01$.

8.5.3 **The performance of the two groups across the two grades**

If we compare the performance of the two groups across the two grades, it is clear from Tables 6 and 7 that the experimental group strengthened the already established significant relations from Grade 1 to Grade 2 in the following measures: Nonce_Eng. (from $p < .01$ to $p < .001$), Story Sequences (from $p < .05$ to $p < .001$), and YLAT_Total (from $p < .05$ to $p < .01$). It also managed to establish new significant relations in Digit recall_Forward (from $p = .123$ to $p < .05$) and Paired Associates (from $p = .633$ to $p < .001$), while it lost its initial advantage in DVIQ I_Voc. (from $p < .01$ to $p = .443$) and Differences (from $p < .05$ to $p = .668$).

The experimental group also maintained significant differences in Nonce_Gr. (from $p < .001$ to $p < .05$). Although it managed to slightly increase by 1.55% (see section 8.5.4 below) the mean score of DVIQ I_Voc. from one grade to the next, it lost its initial advantage (from $p = .006$ to $p = .443$) over the control group. The experimental group marked no change in DVIQ I_Met. ($p < .001$). It maintained significant differences in DVIQ II_Met. (from $p < .001$ to $p < .01$) over the control group even though it scored slightly worse the second year (-2.97%, see section 8.5.4). This is because the control group also marked lower scores in Grade 2 (-2.07%, see Appendix 2).

Table 8 below summarises the differences the independent samples $t$-tests yielded in favour of the experimental group from Grade 1 (before the English intervention) to Grade 2 (after two years of English schooling):
Table 8: Differences attested in the independent samples $t$-tests, in favour of the experimental group

<table>
<thead>
<tr>
<th>Advantage gained</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Advantage lost</th>
<th>Grade 1</th>
<th>Grade 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall_Forward</td>
<td>ns</td>
<td>.037*</td>
<td>DVIQ I_Voc.</td>
<td>.006</td>
<td>ns</td>
</tr>
<tr>
<td>Paired Associates</td>
<td>ns</td>
<td>.000*</td>
<td>Differences</td>
<td>.019</td>
<td>ns</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Advantage kept</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Advantage strengthened</th>
<th>Grade 1</th>
<th>Grade 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVIQ I_Met.</td>
<td>.000</td>
<td>.000</td>
<td>Nonce_Eng.</td>
<td>.007</td>
<td>.000</td>
</tr>
<tr>
<td>DVIQ II_Met.</td>
<td>.000</td>
<td>.005</td>
<td>Story Sequences</td>
<td>.049</td>
<td>.000</td>
</tr>
<tr>
<td>Nonce_Gr.</td>
<td>.000</td>
<td>.026</td>
<td>YLAT_Total</td>
<td>.027</td>
<td>.008</td>
</tr>
</tbody>
</table>

8.5.4 Experimental group: Paired samples $t$-tests

To be able to evaluate the magnitude of the differences attested in the mean scores of the experimental group before and after the FL intervention, a number of paired samples $t$-tests were conducted. Table 9 below provides these differences, along with the percentages in the means of the test scores between Grade 1 and Grade 2. This comparison aims at answering Research Question One. For the respective differences attested in the control group, see Appendix 2.
Table 9: Experimental group, differences (%) found in the means of all tests between Grade 1 and Grade 2

<table>
<thead>
<tr>
<th>Skill tested</th>
<th>Measures</th>
<th>Grade</th>
<th>Mean</th>
<th>t-test</th>
<th>SD</th>
<th>Means difference % (Grade 1 and Grade 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSTM</strong></td>
<td>Digit recall Forward</td>
<td>Grade 1</td>
<td>6.13</td>
<td>( p = .000 )</td>
<td>1.64</td>
<td>15.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>7.05</td>
<td></td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonce_Gr.</td>
<td>Grade 1</td>
<td>34.04</td>
<td>( p = .047 )</td>
<td>3.61</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>34.99</td>
<td></td>
<td>3.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonce_Eng.</td>
<td>Grade 1</td>
<td>25.78</td>
<td>( p = .001 )</td>
<td>4.96</td>
<td>10.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>28.51</td>
<td></td>
<td>3.96</td>
<td></td>
</tr>
<tr>
<td><strong>CWM</strong></td>
<td>Digit recall Backwards</td>
<td>Grade 1</td>
<td>3.07</td>
<td>( p = .000 )</td>
<td>0.81</td>
<td>40.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>4.31</td>
<td></td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Listening Recall</td>
<td>Grade 1</td>
<td>5.67</td>
<td>( p = .003 )</td>
<td>2.72</td>
<td>24.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>7.05</td>
<td></td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td><strong>Inhibition control</strong></td>
<td>SSRT</td>
<td>Grade 1</td>
<td>309.68</td>
<td>( p = .000 )</td>
<td>49.62</td>
<td>-11.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>275.19</td>
<td></td>
<td>55.69</td>
<td></td>
</tr>
<tr>
<td><strong>L1 verbal intelligence</strong></td>
<td>DVIQ I_Voc.</td>
<td>Grade 1</td>
<td>25.11</td>
<td>( p = ns )</td>
<td>1.43</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>25.50</td>
<td></td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DVIQ I_Met.</td>
<td>Grade 1</td>
<td>22.91</td>
<td>( p = ns )</td>
<td>1.17</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>22.91</td>
<td></td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DVIQ II_Met.</td>
<td>Grade 1</td>
<td>13.11</td>
<td>( p = ns )</td>
<td>1.18</td>
<td>-2.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>12.72</td>
<td></td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td><strong>FL aptitude</strong></td>
<td>Memory</td>
<td>Grade 1</td>
<td>9.17</td>
<td>( p = .000 )</td>
<td>1.96</td>
<td>20.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>11.09</td>
<td></td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td>Grade 1</td>
<td>35.28</td>
<td>( p = .001 )</td>
<td>13.51</td>
<td>21.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>42.88</td>
<td></td>
<td>7.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Differences</td>
<td>Grade 1</td>
<td>10.69</td>
<td>( p = .000 )</td>
<td>2.83</td>
<td>19.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>12.77</td>
<td></td>
<td>3.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paired Associates</td>
<td>Grade 1</td>
<td>6.79</td>
<td>( p = .000 )</td>
<td>3.90</td>
<td>52.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>10.35</td>
<td></td>
<td>2.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Semantic Integration</td>
<td>Grade 1</td>
<td>12.99</td>
<td>( p = .010 )</td>
<td>1.85</td>
<td>7.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>13.97</td>
<td></td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Story Sequences</td>
<td>Grade 1</td>
<td>3.98</td>
<td>( p = ns )</td>
<td>2.27</td>
<td>11.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>4.45</td>
<td></td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td><strong>FL aptitude_Total</strong></td>
<td>YLAT_Total</td>
<td>Grade 1</td>
<td>80.69</td>
<td>( p = .000 )</td>
<td>14.66</td>
<td>18.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grade 2</td>
<td>95.72</td>
<td></td>
<td>9.34</td>
<td></td>
</tr>
</tbody>
</table>
What is immediately clear from the above table is that the experimental group increased the mean scores in a number of tests, that are worthy of note. Nevertheless, we will only report the most profound ones, taking into account the fact that the control group also marked a number of equivalent higher scores in Grade 2 (see Appendix 2).

More specifically, in Grade 2 the experimental group fared significantly better in 12 out of the 16 measures taken, but the most sharp increases were recorded in the three measures that follow. The first two tap CWM, namely Digit recall_Backingwards (DRB) and Listening Recall (LR). The mean score of the first increased by 40.39%, while that of the second by 24.34%. Accordingly, the second year the control group increased the mean scores of DRB by 18.26% and of LR by 41.44%. Although quite impressive at first glance the increase of the \( LR_{\text{mean}} \) may look, no significant differences emerged between the two groups in Grade 2.

The most marked difference of all, due to its magnitude, is the one that emerged for a cognitive measure, namely Paired Associates (PA). In Grade 2 the experimental group managed to increase the mean score in this test by 52.43%. The control group also marked higher scores in PA (27.02%), which was not as high as those of the experimental group. Finally, no significant differences emerged for the experimental group across the two grades in the following measures: Nonce_Gr., DVIQ I_Voc., DVIQ I_Met., DVIQ II_Met., and Story Sequences.

Although the SSRT index that taps the inhibition mechanism of participants will be discussed in great detail in section 8.12, it suffices for now to say that the \( SSRT_{\text{mean}} \) was larger for 6-year-olds than for 7-year-olds. The general idea is that the longer estimated the SSRT the more difficult it is for participants to control their action. Thus, the lower scores attested in the performance of both groups actually indicate that the two groups were better able to control their response inhibition mechanism the second year. In particular, the experimental group lowered the mean values of the SSRT by 11.14% while the control group
Chapter 8: Data Analysis

recorded a decrease of 9.95%. In both cases, the difference between the mean scores of Grade 1 and Grade 2 was significant: a) experimental group: \( t(45) = 6.414; p > .001 \), b) control group: \( t(48) = 4.052; p > .001 \).

The section that follows discusses first the correlation matrices\(^2\) that emerged for either grade and group and the relations that were established between the variables. Second, it discusses the regression analyses conducted on the grounds of these relations. This discussion is expected to answer Research Question Two, i.e. whether variables other than PSTM can predict FL comprehension and production.

8.6 Correlation matrices: Grades 1 and 2

Pearson correlation coefficients were computed for each pair of measures obtained in Grades 1 and 2. Tables 10 and 11 provide the resulting correlation matrices that emerged for Grade 1 in the experimental and the control group, respectively. Tables 12 and 13 provide the respective correlation matrices for Grade 2. See Appendix 3 for a detailed picture of the four matrices.

\(^2\) The correlation matrices and the regression analyses that concern the SSRT are presented and discussed in depth, in sections 8.12.4 – 8.12.7. This is because in the exclusionary criteria applied for this index a third participant was excluded from the experimental school. To be able to perform the correlations with the other indices, the data coming from this participant were excluded and this changed the number of the experimental group to 46.
### Table 10: Experimental group (n=47): Correlations between the measures, Grade 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Digit recall_Foward</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Nonce_Gr.</td>
<td>.39**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Nonce_Eng.</td>
<td>.40**</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Digit recall_Backwards</td>
<td>.29*</td>
<td>.20</td>
<td>-.09</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Listening Recall</td>
<td>.40**</td>
<td>.18</td>
<td>.23</td>
<td>.36*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DVIQ I_Voc.</td>
<td>.22</td>
<td>-.08</td>
<td>.21</td>
<td>.14</td>
<td>.38**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DVIQ I_Met.</td>
<td>.16</td>
<td>.12</td>
<td>-.11</td>
<td>.20</td>
<td>.21</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Memory</td>
<td>.10</td>
<td>.08</td>
<td>.21</td>
<td>-.05</td>
<td>.21</td>
<td>-.23</td>
<td>.06</td>
<td>-.24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Classification</td>
<td>.24</td>
<td>.27</td>
<td>.14</td>
<td>.27</td>
<td>.09</td>
<td>.25</td>
<td>-.08</td>
<td>.21</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Differences</td>
<td>-.05</td>
<td>.11</td>
<td>.04</td>
<td>.20</td>
<td>.20</td>
<td>.33*</td>
<td>.30*</td>
<td>.23</td>
<td>.08</td>
<td>.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Paired Associates</td>
<td>.20</td>
<td>.12</td>
<td>-.23</td>
<td>.29*</td>
<td>.20</td>
<td>.33*</td>
<td>.27</td>
<td>.17</td>
<td>-.04</td>
<td>.11</td>
<td>.24</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Semantic Integration</td>
<td>.16</td>
<td>-.11</td>
<td>.33*</td>
<td>.21</td>
<td>.18</td>
<td>.16</td>
<td>-.10</td>
<td>-.07</td>
<td>.26</td>
<td>-.07</td>
<td>.26</td>
<td>.04</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Story Sequences</td>
<td>.01</td>
<td>.05</td>
<td>.03</td>
<td>.06</td>
<td>.23</td>
<td>.09</td>
<td>.08</td>
<td>.40**</td>
<td>-.13</td>
<td>.15</td>
<td>.18</td>
<td>-.01</td>
<td>.02</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>YLAT_Total</td>
<td>.17</td>
<td>.22</td>
<td>.08</td>
<td>.39**</td>
<td>.15</td>
<td>.38**</td>
<td>.03</td>
<td>.32*</td>
<td>.05</td>
<td>.80**</td>
<td>.28</td>
<td>.37**</td>
<td>.15</td>
<td>.26</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)
8.6.1 Experimental data: Grade 1

The indices of PSTM were highly intercorrelated as they all measure the capacity of the phonological loop, providing thus an initial indication of the internal validity of the tests. Digit recall Forward was significantly correlated with both Nonce_Gr. \((r = .39)\) and Nonce_Eng. \((r = .40)\) at the level of .01 significance. PSTM measures were also correlated with CWM and cognitive tasks. Digit recall Forward was correlated with Digit recall Backwards \((r = .29, p < .05)\), while Nonce_Eng. was correlated with Semantic Integration \((r = .33, p < .05)\).

As expected, the two CWM measures were highly correlated with each other. Digit recall Backwards was correlated with Listening Recall \((r = .36, p < .05)\). Apart from the association mentioned earlier between CWM and PSTM, CWM also shared strong relations with verbal and non-verbal tasks. In particular, Digit recall Backwards was correlated with Paired Associates \((r = .29, p < .05)\) and shared stronger relations with YLAT_Total \((r = .39, p < .01)\). The highest degree of association was obtained for Listening Recall and both Digit recall Forward \((r = .40, p < .01)\) and DVIQ_I_Voc. \((r = .38, p < .01)\).

Significant intercorrelations were obtained for the L1 verbal intelligence measures as well, with DVIQ_I_Met. being correlated with DVIQ II_Met. \((r = .37, p < .05)\). Apart from this expected relation, the L1 verbal measures also shared moderate to strong relations with non-verbal, cognitive tasks. DVIQ_I_Voc. was moderately correlated with Differences \((r = .33, p < .05)\) and Paired Associates \((r = .33, p < .05)\), while it shared stronger relations with YLAT_Total \((r = .38, p < .01)\). Likewise, DVIQ_I_Met. was moderately correlated with Differences \((r = .30, p < .05)\), and DVIQ II_Met. with YLAT_Total \((r = .32, p < .05)\). A stronger association was obtained for DVIQ II_Met. and Story Sequences \((r = .40, p < .01)\).
Finally, moderate to very strong associations emerged among cognitive skills, with Classification and Paired Associates being correlated with YLAT_Total ($r = .80$, $p < .01$ and $r = .37$, $p < .05$, respectively).

The measure that shared no correlations with any of the other skills, cognitive and non-cognitive, was Memory.
## Table 11: Control group (n=49): Correlations between the measures, Grade 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Digit recall_Forward</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Nonce_Gr.</td>
<td><strong>.32</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Nonce_Eng.</td>
<td><strong>.33</strong></td>
<td><strong>.58</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Digit recall_Backwards</td>
<td>.06</td>
<td>.17</td>
<td>.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Listening Recall</td>
<td><strong>.30</strong></td>
<td>.07</td>
<td>.15</td>
<td>.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DVIQ I_Voc.</td>
<td>.01</td>
<td>.05</td>
<td>.07</td>
<td>.07</td>
<td>.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DVIQ I_Met.</td>
<td>.09</td>
<td>.19</td>
<td><strong>.35</strong></td>
<td><strong>.31</strong></td>
<td>.14</td>
<td><strong>.38</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DVIQ II_Met.</td>
<td>-.01</td>
<td>.17</td>
<td>.16</td>
<td>-.09</td>
<td>-.09</td>
<td><strong>.40</strong></td>
<td><strong>.28</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Memory</td>
<td>.11</td>
<td>.23</td>
<td>.17</td>
<td>-.01</td>
<td>.03</td>
<td>-.08</td>
<td>.11</td>
<td>.07</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Classification</td>
<td>.08</td>
<td>.10</td>
<td>.01</td>
<td>.05</td>
<td>.17</td>
<td>.26</td>
<td>.25</td>
<td>.18</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Differences</td>
<td>.02</td>
<td>.26</td>
<td>.19</td>
<td>.15</td>
<td>.06</td>
<td>.26</td>
<td>.10</td>
<td>.20</td>
<td>.07</td>
<td>.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Paired Associates</td>
<td>.23</td>
<td>.26</td>
<td>.13</td>
<td>-.06</td>
<td>-.05</td>
<td>.16</td>
<td>.19</td>
<td><strong>.44</strong></td>
<td><strong>.12</strong></td>
<td><strong>.28</strong></td>
<td><strong>.09</strong></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Semantic Integration</td>
<td>.17</td>
<td>.13</td>
<td>.10</td>
<td>.03</td>
<td>.20</td>
<td>.25</td>
<td><strong>.34</strong></td>
<td><strong>.28</strong></td>
<td>.07</td>
<td>.16</td>
<td>-.02</td>
<td>.28</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Story Sequences</td>
<td>.01</td>
<td>-.04</td>
<td>-.16</td>
<td>-.02</td>
<td>.09</td>
<td>.05</td>
<td>-.05</td>
<td>-.11</td>
<td>.10</td>
<td>.18</td>
<td>.13</td>
<td>-.09</td>
<td><strong>.41</strong></td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>YLAT_Total</td>
<td>.14</td>
<td>.16</td>
<td>.05</td>
<td>.04</td>
<td>.20</td>
<td><strong>.34</strong></td>
<td><strong>.29</strong></td>
<td><strong>.31</strong></td>
<td>.25</td>
<td><strong>.91</strong></td>
<td><strong>.29</strong></td>
<td><strong>.38</strong></td>
<td><strong>.30</strong></td>
<td><strong>.30</strong></td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).
8.6.2 Control data: Grade 1

According to Table 11, significant correlations were obtained for a number of tasks for the control group in Grade 1. As already found for the experimental group, PSTM measures were highly intercorrelated. Digit recall_Foward was correlated with both Nonce_Gr. \((r = .32)\) and Nonce_Eng. \((r = .33)\) at the level of .05 significance, while the highest degree of association was obtained for Nonce_Gr. and Nonce_Eng. \((r = .58, p < .01)\). PSTM tasks were additionally correlated with L1 verbal intelligence, in particular metalinguistic knowledge, with Nonce_Eng. being correlated with DVIQ I_Met. \((r = .35, p < .05)\).

A number of interesting correlations emerged between CWM measures and both PSTM and L1 verbal intelligence ones. Listening Recall was correlated with Digit recall_Foward \((r = .30)\), while Digit recall_Backwards with DVIQ I_Met. \((r = .31)\), both at the level of .05 significance.

As expected, L1 verbal intelligence measures shared strong associations with one another. DVIQ I_Voc. was significantly correlated with both DVIQ I_Met. \((r = .38, p < .01)\) and DVIQ II_Met. \((r = .40, p < .01)\). Also, DVIQ I_Met. was moderately correlated with DVIQ II_Met. \((r = .28, p < .05)\). In addition to these intercorrelations, the measures were also correlated with CWM tasks (as already mentioned) as well as with cognitive ones. DVIQ I_Voc. was correlated with YLAT_Total \((r = .34)\), while DVIQ I_Met. was associated with Semantic Integration \((r = .34)\) and YLAT_Total \((r = .29)\), all at the level of .05 significance. The strongest association emerged for DVIQ II_Met. with Paired Associates \((r = .44, p < .01)\), followed by a more moderate relation of the same task with YLAT_Total \((r = .31, p < .05)\).

As cognitive tests tap a number of cognitive skills, they shared strong correlations with one another. Classification was correlated with Paired Associates \((r
CHAPTER 8: DATA ANALYSIS

= .28, p < .05) and shared an association of a higher magnitude with YLAT_Total (r = .91, p < .01). Differences (D), Paired Associates (PA) and Story Sequences (SS) were all correlated with YLAT_Total (r = .29, p < .05 (D); r = .38, p < .01 (PA); r = .30, p < .05 (SS)). Finally, Semantic Integration was significantly correlated with Story Sequences (r = .41, p < .01) and more moderately with YLAT_Total (r = .30, p < .05). As was the case of the experimental group, Memory shared no correlations with any of the other skills, cognitive and non-cognitive.
### Table 12: Experimental group (n=47): Correlations between the measures, Grade 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FL_Voc_Total</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>FL_Receptive</td>
<td>.72**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>FL_Productive</td>
<td>.96**</td>
<td>.66**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Digit recall_Foward</td>
<td>.24</td>
<td>.23</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Nonce_Gr.</td>
<td>.40**</td>
<td>.42**</td>
<td>.41**</td>
<td>.49**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Nonce_Eng.</td>
<td>.16</td>
<td>.33*</td>
<td>.15</td>
<td>.44**</td>
<td>.41**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Digit recall_Backwards</td>
<td>.13</td>
<td>.11</td>
<td>.20</td>
<td>.34*</td>
<td>.28</td>
<td>.29*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Listening Recall</td>
<td>.36*</td>
<td>.44**</td>
<td>.42**</td>
<td>.20</td>
<td>.36*</td>
<td>.34*</td>
<td>.32*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DVIQ I_Voc.</td>
<td>.21</td>
<td>.27</td>
<td>.29</td>
<td>.10</td>
<td>.29*</td>
<td>.22</td>
<td>.16</td>
<td>.34*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DVIQ I_Met.</td>
<td>.23</td>
<td>.29*</td>
<td>.24</td>
<td>.12</td>
<td>.08</td>
<td>.08</td>
<td>.45**</td>
<td>.22</td>
<td>.23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>DVIQ II_Met.</td>
<td>.20</td>
<td>.29*</td>
<td>.19</td>
<td>.01</td>
<td>.20</td>
<td>.12</td>
<td>.24</td>
<td>.38**</td>
<td>.23</td>
<td>.57**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Memory</td>
<td>.19</td>
<td>.26</td>
<td>.21</td>
<td>.10</td>
<td>.08</td>
<td>.22</td>
<td>.08</td>
<td>.32*</td>
<td>.27</td>
<td>.15</td>
<td>.00</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Classification</td>
<td>.07</td>
<td>.00</td>
<td>-0.0</td>
<td>.04</td>
<td>-0.2</td>
<td>-0.2</td>
<td>.25</td>
<td>-0.1</td>
<td>-0.3</td>
<td>.18</td>
<td>.22</td>
<td>-0.2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Differences</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td>.27</td>
<td>.25</td>
<td>.13</td>
<td>.31*</td>
<td>.27</td>
<td>.50**</td>
<td>.34*</td>
<td>.17</td>
<td>.21</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Paired Associates</td>
<td>.22</td>
<td>.20</td>
<td>.23</td>
<td>.18</td>
<td>.20</td>
<td>-0.0</td>
<td>.08</td>
<td>.28</td>
<td>.02</td>
<td>.09</td>
<td>.24</td>
<td>.06</td>
<td>.28</td>
<td>-0.0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Semantic Integration</td>
<td>.16</td>
<td>.02</td>
<td>.22</td>
<td>.27</td>
<td>.26</td>
<td>.14</td>
<td>.20</td>
<td>.44**</td>
<td>.38**</td>
<td>.10</td>
<td>.05</td>
<td>.35*</td>
<td>-.21</td>
<td>.33*</td>
<td>-.07</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Story Sequences</td>
<td>.08</td>
<td>.21</td>
<td>.02</td>
<td>.13</td>
<td>.20</td>
<td>.17</td>
<td>.08</td>
<td>.06</td>
<td>.09</td>
<td>.28</td>
<td>.27</td>
<td>-0.1</td>
<td>.18</td>
<td>.16</td>
<td>.14</td>
<td>-0.1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>YLAT_Total</td>
<td>.31*</td>
<td>.20</td>
<td>.27</td>
<td>.28</td>
<td>.15</td>
<td>.07</td>
<td>.42**</td>
<td>.29*</td>
<td>.13</td>
<td>.32*</td>
<td>.28</td>
<td>.24</td>
<td>.72**</td>
<td>.52**</td>
<td>.43**</td>
<td>.27</td>
<td>.28</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed)
8.6.3 Experimental data: Grade 2

What is evident from Table 12 is that a number of interesting correlations emerged between the tasks for the experimental group in Grade 2. All three tests that pertain to FL vocabulary knowledge correlated significantly with one another with coefficients ranging from .66 (\(p < .01\); FL_Receptive and FL_Productive) to the coefficients of the FL_Voc_Total with FL_Receptive (\(r = .72, p < .01\)) and FL_Productive (\(r = .96, p < .01\)). These findings indicate the validity and reliability of the three FL tests. In addition, the tests were also significantly associated with the rest of the tasks that tap PSTM, CWM and L1 verbal intelligence. The strongest association emerged for the aggregate score of the FL test and Nonce_Gr. (\(r = .40, p < .01\)). The aggregate score of the FL test was also, more moderately so, associated (at the level of .05 significance) with both Listening Recall (\(r = .36\)) and YLAT_Total (\(r = .31\)).

When the two skills in the FL were examined separately, the following links were established. FL_Receptive shared strong relations with two of the PSTM tasks, namely Nonce_Gr. (\(r = .42, p < .01\)) and Nonce_Eng. (\(r = .33, p < .05\)). The strongest association, though, emerged for FL_Receptive and a CWM task (Listening Recall, \(r = .44, p < .01\)). FL_Voc_Receptive was also correlated, more weakly so, with DVIQ_I_Met. (\(r = .29, p < .05\)) and DVIQ II_Met. (\(r = .29, p < .05\)). On the other hand, production in the FL (FL_Productive) was also significantly correlated with one PSTM task, namely Nonce_Gr. (\(r = .41, p < .01\)). Also, the strong correlation found between FL_Voc_Productive and Listening Recall (\(r = .42, p < .01\)).

As expected and already found earlier in this study, the PSTM measures were highly intercorrelated. Digit recall_Forward shared very strong associations with both Nonce_Gr. (\(r = .49, p < .01\)) and Nonce_Eng. (\(r = .44, p < .01\)), while Nonce_Gr. was significantly correlated with Nonce_Eng. (\(r = .41, p < .01\)). More interesting though, were the strong relations found between PSTM, CWM, and L1 vocabulary tasks.
Digit recall_Forward was correlated with Digit recall_Backwards \( (r = .34, p < .05) \), and Nonce_Gr. was associated with DVIQ I_Voc. \( (r = .29, p = .05) \). The rest of the findings establish associations between Nonce_Gr. and Listening Recall \( (r = .36, p < .05) \). Nonce_Eng. was correlated with both Listening Recall \( (r = .34, p < .05) \) and Digit recall_Backwards \( (r = .29, p = .05) \).

The two CWM measures were correlated with one another, with Listening Recall sharing moderate relations with Digit recall_Backwards \( (r = .32, p < .05) \). As already mentioned, Listening Recall was also correlated with FL performance and PSTM tasks. Stronger were the relations that emerged for CWM, L1 verbal intelligence and cognitive measures. Listening Recall was correlated with two verbal tasks, namely DVIQ I_Voc. \( (r = .34, p < .05) \) and DVIQ II_Met. \( (r = .38, p < .01) \), while it also shared moderate associations with cognitive tasks, such as Memory \( (r = .32, p < .05) \) and YLAT_Total \( (r = .29, p < .05) \). The strongest association of all emerged for Listening Recall and Semantic Integration \( (r = .44, p < .01) \). The second CWM task, Digit recall_Backwards, shared relations of a high magnitude with DVIQ I_Met. \( (r = .45, p < .01) \) and YLAT_Total \( (r = .42, p < .01) \) and a weaker one with Differences \( (r = .31, p < .05) \).

Two of the L1 verbal intelligence tasks shared strong relations with one another: DVIQ I_Met. was significantly correlated with DVIQ II_Met. \( (r = .57, p < .01) \). In addition to these expected relations, they also marked external relations with tests that tap PSTM, FL vocabulary, CWM (as already mentioned), and cognitive skills. DVIQ I_Voc. was significantly correlated with Differences \( (r = .50, p < .01) \) and Semantic Integration \( (r = .38, p < .01) \) while DVIQ I_Met. was associated with Differences \( (r = .34) \) and YLAT_Total \( (r = .32) \), both at the level of .05 significance.

Finally, cognitive tasks shared strong correlations with one another. Memory and Differences were correlated with Semantic Integration \( (r = .35, p < .05; r = .33, p \)
< .05, respectively), while Classification (C), Differences (D), and Paired Associates (PA) were all significantly associated with YLAT_Total ($r = .72, p < .01$ (C); $r = .52, p < .01$ (D), $r = .43, p < .01$ (PA)).

Story Sequences shared no association with any of the other, cognitive and non-cognitive ones.
### Table 13: Control group (n=49): Correlations between the measures, Grade 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.37**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.21</td>
<td>.64**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.35*</td>
<td>.37**</td>
<td>.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.36*</td>
<td>.36*</td>
<td>.21</td>
<td>.21</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>.36*</td>
<td>.36*</td>
<td>.21</td>
<td>.21</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.32*</td>
<td>.42**</td>
<td>.42**</td>
<td>.19</td>
<td>.31*</td>
<td>.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td>.48**</td>
<td>.34</td>
<td>.24</td>
<td>.28</td>
<td>.26</td>
<td>.57**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.22</td>
<td>.24</td>
<td>.19</td>
<td>.23</td>
<td>.28</td>
<td>.13</td>
<td>.30*</td>
<td>.14</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.06</td>
<td>-.06</td>
<td>.00</td>
<td>-.10</td>
<td>-.05</td>
<td>.04</td>
<td>-.04</td>
<td>.22</td>
<td>.14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.16</td>
<td>-.05</td>
<td>-.10</td>
<td>.02</td>
<td>.29*</td>
<td>.14</td>
<td>.27</td>
<td>-.04</td>
<td>.23</td>
<td>-.17</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td>.32*</td>
<td>.29*</td>
<td>.29*</td>
<td>.28*</td>
<td>-.16</td>
<td>.51**</td>
<td>.24</td>
<td>.27</td>
<td>.06</td>
<td>.10</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.19</td>
<td>.16</td>
<td>.11</td>
<td>.14</td>
<td>.09</td>
<td>-.01</td>
<td>.07</td>
<td>.05</td>
<td>-.08</td>
<td>-.07</td>
<td>-.17</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.12</td>
<td>.07</td>
<td>.08</td>
<td>.21</td>
<td>.24</td>
<td>.28</td>
<td>.29*</td>
<td>.22</td>
<td>.67**</td>
<td>.09</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)
8.6.4 Control data: Grade 2

What is clear from Table 13, is that DVIQ I_Voc. and Story Sequences shared no correlations with the other skills. As for the rest of the other measures, significant correlations emerged between the tasks for the control group in Grade 2.

PSTM measures shared strong correlations with one another, with Digit recall_Forward correlating with Nonce_Gr. \((r = .37, p < .05)\) and Nonce_Gr. correlating significantly high with Nonce_Eng. \((r = .64, p < .01)\). Apart from these internal relations, PSTM also shared strong external correlations with the other tasks that test CWM, L1 verbal intelligence and cognitive skills. Digit recall_Forward was correlated with Digit recall_Backwards \((r = .35, p < .05)\) and DVIQ I_Met. \((r = .32, p < .05)\). Overall, stronger associations emerged for Nonce_Gr. with DVIQ I_Met. \((r = .42, p < .01)\), DVIQ II_Met. \((r = .48, p < .01)\) and Paired Associates \((r = .32, p < .05)\).

The third PSTM task, Nonce_Eng., was significantly correlated with DVIQ I_Met. \((r = .42, p < .01)\) and more moderately so with DVIQ II_Met. \((r = .34, p < .05)\) and Paired Associates \((r = .29, p < .05)\).

The measures of CWM shared strong correlations with PSTM, L1 verbal intelligence and cognitive tests. Listening Recall was moderately correlated (at the level of .05 significance) with Digit recall_Forward \((r = .36)\), Nonce_Gr. \((r = .36)\), DVIQ I_Met. \((r = .31)\), Differences \((r = .29)\) and Paired Associates \((r = .28)\). Digit recall_Backwards, on the other hand, was significantly correlated with Nonce_Gr. \((r = .37, p < .01)\) and more moderately so with Nonce_Eng. \((r = .36)\), Paired Associates \((r = .29)\) and Semantic Integration \((r = .33)\), all at the level of .05 significance.

Two of the three L1 verbal intelligence measures, the ones that test metalinguistic knowledge, i.e. DVIQ I_Met. and DVIQ II_Met., were highly intercorrelated \((r = .57, p < .01)\), indicating in this way their validity and reliability. We already mentioned that CWM and PSTM tests shared strong relations with
Chapter 8: Data Analysis

metalinguistic knowledge. The correlation matrices provide a number of other strong correlations of these verbal tests with cognitive ones. DVIQ I_Met. was correlated with Memory ($r = .30, p < .05$) and significantly so with Paired Associates ($r = .51, p < .001$), while DVIQ II_Met. shared moderate correlations with Semantic Integration ($r = .31, p < .05$) and YLAT_Total ($r = .29, p < .05$).

Finally, cognitive skills related with one another to a weaker or stronger extent. Classification (C), Paired Associates (PA), and Semantic Integration (SI) were all correlated with YLAT_Total ($r = .67, p < .01$ (C), $r = .33, p < .05$ (PA); $r = .44, p < .01$ (SI)).

8.7 Regression analyses

Nevertheless, it would be premature to interpret statistically significant relations in terms of mere correlations. Therefore, to determine the direction of causality, a number of linear and stepwise regression analyses were performed for the experimental group in both grades. In regression analysis the impact of one or more (independent) variable(s) on another (dependent) is estimated by the proportion of variance explained by the independent variable in question. To investigate the differential effect of PSTM, CWM, as well as L1 verbal intelligence and FL aptitude on LL performance (Grade 1) and FL success (Grade 2), we carried out a number of regression analyses, the findings of which are displayed in the following sections. As already mentioned, the correlation matrices and regression analyses that concern the potential relation of SSRT with the other measures are discussed in section 8.12.

8.7.1 Grade 1

In Grade 1 we first examined whether the independent variables of PSTM (Digit recall_Forward, Nonce Gr., Nonce Eng.) and CWM (Listening Recall, Digit recall_Backwards) could explain an amount of variance in the scores of L1 verbal
CHAPTER 8: DATA ANALYSIS

intelligence (DVIQ I_Voc., DVIQ I_Met., DVIQ II_Met.: dependent variables). With respect to PSTM, none of the independent variables accounted for any variance in the scores of the dependent variables. As for the independent variables that tap CWM, only Listening Recall accounted for 12.8% of the variance in the scores of DVIQ I_Voc., $F(1,45) = 7.765; p < .01$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall</td>
<td>.128</td>
<td>7.765</td>
<td>1</td>
<td>45</td>
<td>.384</td>
<td>.008</td>
</tr>
</tbody>
</table>

We then examined whether FL aptitude (YLAT_Total, Memory, Classification, Differences, Paired Associates, Semantic Integration, and Story Sequences: independent variables) could explain an amount of variance in L1 verbal intelligence. The findings suggested that YLAT_Total accounted for 12.3% of the variance in the scores of DVIQ I_Voc., $F(1,45) = 7.475; p < .01$ and 8% of the variance in the scores of DVIQ II_Met., $F(1,45) = 5.018; p < .05$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLAT_Total (for DVIQ I_Voc.)</td>
<td>.123</td>
<td>7.475</td>
<td>1</td>
<td>45</td>
<td>.377</td>
<td>.009</td>
</tr>
<tr>
<td>YLAT_Total (for DVIQ II_Met.)</td>
<td>.080</td>
<td>5.018</td>
<td>1</td>
<td>45</td>
<td>.317</td>
<td>.030</td>
</tr>
</tbody>
</table>

Of the other independent cognitive variables (Memory, Classification, Differences, Paired Associates, Semantic Integration and Story Sequences), Paired Associates accounted for 9% of the variance in the scores of DVIQ I_Voc., $F(1,45) = 5.541; p < .05$, Differences accounted for 7.1% of the variance in the scores of DVIQ
CHAPTER 8: DATA ANALYSIS

I_Met., F(1,45) = 4.524; p < .05 and Story Sequences for 14.2% of the variance in the scores of DVIQ II_Met., F(1,45) = 8.594; p < .01.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired Associates (for DVIQ I_Voc.)</td>
<td>.090</td>
<td>5.541</td>
<td>1</td>
<td>45</td>
<td>.331</td>
<td>.023</td>
</tr>
<tr>
<td>Differences (for DVIQ I_Met.)</td>
<td>.071</td>
<td>4.524</td>
<td>1</td>
<td>45</td>
<td>.302</td>
<td>.039</td>
</tr>
<tr>
<td>Story Sequences (for DVIQ II_Met.)</td>
<td>.142</td>
<td>8.594</td>
<td>1</td>
<td>45</td>
<td>.400</td>
<td>.005</td>
</tr>
</tbody>
</table>

8.7.2 Grade 2

In the same vein, a number of linear and stepwise regressions were performed for the experimental group, in Grade 2. It should be noted that FL proficiency, the dependent variable in the models that follow, is represented both by data on the two L2 skills (FL_Receptive, FL_Productive) and a single aggregate score, FL_Voc._Total. First, we examined whether FL_Voc._Total or separately FL_Receptive and FL_Productive (dependent variables) could be predicted by PSTM, CWM, L1 verbal intelligence, or non-verbal cognitive skills.

8.7.2.1 FL Voc._Total as the dependent variable

Of the independent variables that tap CWM (Digit recall_Backwards and Listening Recall), only Listening Recall accounted for 11.3% of the variance in the scores of FL_Voc._Total, F(1,45) = 6.835; p < .05.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall</td>
<td>.113</td>
<td>6.835</td>
<td>1</td>
<td>45</td>
<td>.363</td>
<td>.012</td>
</tr>
</tbody>
</table>
Of the independent variables that tap PSTM, Nonce_Gr. accounted for 14% of the variance in the scores of the dependent variable, $F(1,45) = 8.510; p < .01$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonce_Gr.</td>
<td>.140</td>
<td>8.510</td>
<td>1</td>
<td>45</td>
<td>.399</td>
<td>.005</td>
</tr>
</tbody>
</table>

None of the independent variables that tap L1 verbal intelligence accounted for any variance in the scores of the dependent variable, while of the independent variables that tap FL aptitude, YLAT_Total accounted for 7.4% of the variance in the scores of the dependent variable FL_Voc.Total, $F(1,45) = 4.653; p < .05$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLAT_Total</td>
<td>.074</td>
<td>4.653</td>
<td>1</td>
<td>45</td>
<td>.306</td>
<td>.036</td>
</tr>
</tbody>
</table>

### 8.7.2.2 FL_Receptive as the dependent variable

Of the independent variables that tap CWM, only Listening Recall accounted for 17.7% of the variance in the scores of the dependent variable, $F(1,45) = 10.895; p < .01$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall</td>
<td>.177</td>
<td>10.895</td>
<td>1</td>
<td>45</td>
<td>.441</td>
<td>.002</td>
</tr>
</tbody>
</table>

Of the independent variables that tap PSTM, Nonce_Gr. accounted for 16% of the variance in the scores of the dependent variable, $F(1,45) = 9.788; p < .01$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonce_Gr.</td>
<td>.160</td>
<td>9.788</td>
<td>1</td>
<td>45</td>
<td>.423</td>
<td>.003</td>
</tr>
</tbody>
</table>
Of the independent variables that tap L1 verbal intelligence, DVIQ II_Met. accounted for 6.6% of the variance in the scores of the dependent variable, \( F(1,45) = 4.248; p < .05 \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVIQ II_Met.</td>
<td>.066</td>
<td>4.248</td>
<td>1</td>
<td>45</td>
<td>.294</td>
<td>.045</td>
</tr>
</tbody>
</table>

Finally, none of the independent variables that tap FL aptitude, accounted for any variance in the scores of FL_Receptive.

### 8.7.2.3 FL_Productive as the dependent variable

Of the independent variables that tap CWM, Listening Recall only accounted for 15.8% of the variance in the scores of the dependent variable, \( F(1,45) = 9.615; p < .01 \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall</td>
<td>.158</td>
<td>9.615</td>
<td>1</td>
<td>45</td>
<td>.420</td>
<td>.003</td>
</tr>
</tbody>
</table>

Of the independent variables that tap PSTM, Nonce_Gr. accounted for 15% of the variance in the scores of the dependent variable, \( F(1,45) = 9.086; p < .01 \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonce_Gr.</td>
<td>.150</td>
<td>9.086</td>
<td>1</td>
<td>45</td>
<td>.410</td>
<td>.004</td>
</tr>
</tbody>
</table>

Finally, neither the independent variables that tap L1 verbal intelligence, nor those that tap FL aptitude, accounted for any variance in the scores of the dependent variable.
**CHAPTER 8: DATA ANALYSIS**

**8.7.2.4 L1 verbal intelligence as the dependent variable**

Then, we examined whether L1 verbal intelligence could be predicted by the independent variables that tap PSTM, CWM and FL aptitude. Of the independent variables that tap PSTM, Nonce_Gr. accounted for 6.2% of the variance in the scores of DVIQ I_Voc., \(F(1,45) = 4.065; p = .05\).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>(F) change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonce_Gr.</td>
<td>.062</td>
<td>4.065</td>
<td>1</td>
<td>45</td>
<td>.288</td>
<td>.050</td>
</tr>
</tbody>
</table>

Of the independent variables that tap CWM, Listening Recall accounted for 9.5% of the variance in the scores of DVIQ I_Voc., \(F(1,45) = 5.804; p < .05\) and for 12.4% of the variance in the scores of DVIQ II_Met., \(F(1,45) = 7.520; p < .01\). Digit recall_Backwards accounted for 18.5% of the variance in the scores of DVIQ I_Met., \(F(1,45) = 11.459; p < .01\).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>(F) change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall (for DVIQ I_Voc.)</td>
<td>.095</td>
<td>5.804</td>
<td>1</td>
<td>45</td>
<td>.338</td>
<td>.020</td>
</tr>
<tr>
<td>Digit recall_Backwards (for DVIQ I_Met.)</td>
<td>.185</td>
<td>11.459</td>
<td>1</td>
<td>45</td>
<td>.451</td>
<td>.001</td>
</tr>
<tr>
<td>Listening Recall (for DVIQ II_Met.)</td>
<td>.124</td>
<td>7.520</td>
<td>1</td>
<td>45</td>
<td>.378</td>
<td>.009</td>
</tr>
</tbody>
</table>

Of the independent variables that tap FL aptitude, Differences and Classification together accounted for 29.2% of the variance in the scores of DVIQ I_Voc. More specifically, Differences accounted for 22.9% of the variance, \(F(1,45) = 14.671; p < \)
while Classification accounted for another 6.3% (negative direction), $F(1,44) = 4.988; p < .05$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>$F$ change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences</td>
<td>.229</td>
<td>14.671</td>
<td>1</td>
<td>45</td>
<td>.507</td>
<td>.000</td>
</tr>
<tr>
<td>Classification</td>
<td>.292</td>
<td>4.988</td>
<td>1</td>
<td>44</td>
<td>-.277</td>
<td>.031</td>
</tr>
</tbody>
</table>

The independent variable YLAT_Total accounted for 8.1% of the variance in the scores of DVIQ I_Met., $F(1,45) = 5.069; p < .05$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>$F$ change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLAT_Total</td>
<td>.081</td>
<td>5.069</td>
<td>1</td>
<td>45</td>
<td>.318</td>
<td>.029</td>
</tr>
</tbody>
</table>

Of the other independent cognitive variables, Differences accounted for 9.3% of the variance in the scores of DVIQ I_Met., $F(1,45) = 5.690; p < .05$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>$F$ change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences</td>
<td>.093</td>
<td>5.690</td>
<td>1</td>
<td>45</td>
<td>.335</td>
<td>.021</td>
</tr>
</tbody>
</table>

### 8.7.2.5 FL aptitude as the dependent variable

Then, we examined whether individual differences in FL aptitude, i.e. cognitive skills, could be explained by independent variables that tap CWM. The regressions performed yielded the following results. Of the independent variables that tap CWM, Digit recall_Backswards accounted for 16.2% of the variance in the scores of YLAT_Total, $F(1,45) = 9.870; p < .01$ and for 7.6% of the variance in the scores of Differences, $F(1,45) = 4.795; p < .05$. Listening Recall accounted for 17.4% of the
variance in the scores of Semantic Integration, $F(1,45) = 10.724; p < .01$ and for 8.3% of the variance in the scores of Memory, $F(1,45) = 5.175; p < .05$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall_Backwards</td>
<td>.162</td>
<td>9.870</td>
<td>1</td>
<td>45</td>
<td>.424</td>
<td>.003</td>
</tr>
<tr>
<td>(for YLAT_Total)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit recall_Backwards</td>
<td>.076</td>
<td>4.795</td>
<td>1</td>
<td>45</td>
<td>.310</td>
<td>.034</td>
</tr>
<tr>
<td>(for Differences)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Recall</td>
<td>.174</td>
<td>10.724</td>
<td>1</td>
<td>45</td>
<td>.439</td>
<td>.002</td>
</tr>
<tr>
<td>(for Semantic Integration)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Recall</td>
<td>.083</td>
<td>5.175</td>
<td>1</td>
<td>45</td>
<td>.321</td>
<td>.028</td>
</tr>
<tr>
<td>(for Memory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8.7.2.6 PSTM as the dependent variable

Then, we examined whether individual differences in PSTM could be predicted by independent variables that tap CWM. Of the independent variables that tap CWM, Digit recall_Backwards accounted for 9.9% of the variance in the scores of Digit recall_Forward, $F(1,45) = 6.043; p < .05$. Listening Recall accounted for 11% and 9.6% respectively, of the variance in the scores of Nonce_Gr., $F(1,45) = 6.694 p < .05$ and Nonce_Eng, $F(1,45) = 5.882; p < .05$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall_Backwards</td>
<td>.099</td>
<td>6.043</td>
<td>1</td>
<td>45</td>
<td>.344</td>
<td>.018</td>
</tr>
<tr>
<td>(for Digit recall_Forward)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Recall</td>
<td>.110</td>
<td>6.694</td>
<td>1</td>
<td>45</td>
<td>.360</td>
<td>.013</td>
</tr>
<tr>
<td>(for Nonce_Gr.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Recall</td>
<td>.096</td>
<td>5.882</td>
<td>1</td>
<td>45</td>
<td>.340</td>
<td>.019</td>
</tr>
<tr>
<td>(for Nonce_Eng.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8.8 Two factor analyses of variance (ANOVAs)

A number of two-way ANOVAs were performed for both grades, to examine the effect of two independent variables, i.e. the between-subjects factors (first, group and year and then, group and sex) on one dependent variable (score of each measure).

First, the two-way ANOVAs examined the combined effect of school and year on the scores of each test. Overall, the experimental group outperformed the control one, confirming the findings from the independent samples t-tests.

In particular, the experimental group was significantly better in the following measures, irrespective of grade: DVIQ_I_Met., F(1,188) = 59.873; p = .000; DVIQ II_Met., F(1,188) = 21.076; p = .000; Story Sequences, F(1,188) = 15.897; p = .000. In addition, it scored significantly higher in DVIQ_I_Voc., but only in Grade 1: F(1,188) = 7.270; p = .008. The same holds for Differences: F(1,188) = 44.713; p = .000, where in Grade 2 it lost its initial advantage.

Irrespective of school, the following measures marked significant differences in their scores, from Grade 1 to Grade 2: Listening Recall, F(1,188) = 18.245; p = .000; SSRT, F(1,186) = 20.243; p = .000; Memory, F(1,188) = 28.833; p = .000; Classification, F(1,188) = 22.621; p = .000; Semantic Integration, F(1,188) = 1.419; p = .001.

Although the control group scored higher than the experimental in Grade 1 in Digit recall_Backwards, the experimental managed to outperform the control in Grade 2: F(1,188) = 4.634; p = .033. Appendices 4 and 5 (Chapter 8) give the descriptive statistics, Profile Plots and t-tests of Digit recall_Backwards, respectively. In Grade 2 the experimental also scored significantly better than the control in Paired Associates: F(1,188) = 3.948; p = .048. Appendices 6 and 7 (Chapter 8) give the descriptive statistics, Profile Plots and t-tests of Paired Associates, respectively.
In the measures that follow, although both groups managed to score higher the second year, still the experimental group scored significantly better than the control: Digit recall_Forward (School: F(1,188) = 6.824; \( p = .010 \); Year: F(1,188) = 17.865; \( p = .000 \)); Nonce_Gr. (School: F(1,188) = 21.336; \( p = .000 \); Year: F(1,188) = 10.804; \( p = .001 \)); Nonce_Eng. (School: F(1,188) = 28.847; \( p = .000 \); Year: F(1,188) = 7.565; \( p = .007 \)); YLAT_Total (School: F(1,188) = 11.128; \( p = .001 \); Year: F(1,188) = 62.488; \( p = .000 \)).

The experimental group outperformed the control in several measures. However, the differences in the scores of two tests, namely Digit recall_Backwards and Paired Associates, were the most prominent ones. The results suggest that CWM and paired associative memory were the two skills that were heavily affected by the early FL intervention. This is a truly important finding as both abilities are closely associated with the retention of FL vocabulary.

Second, the two-way ANOVAs examined the combined effect of school and sex on the scores of each test. The only gender effect that emerged concerned Digit recall_Forward and the male participants of the experimental group in Grade 1: F(1,92) = 4.280; \( p = .033 \) (see Appendix 8, Chapter 8). However, this was of no importance as we were interested in finding whether the two-year FL intervention would have an impact on the scores of the male or female participants of the experimental group. In other words, we were looking for a potential gender effect in Grade 2. In the Listening Recall task, the girls coming from both schools scored better than the boys in Grade 2. However, this difference was marginal and did not reach statistical significance: F(1,92) = 3.827; \( p = .053 \) (ns). For the reason stage above, this finding was not considered important.

To sum up, the findings indicated no gender effect on the performance of learners in the tests taken. However, to fully answer Research Question Three, we
also conducted several MANOVAs, to see whether there was a gender effect on sets of tests that tap the same skill.

8.9 Multivariate analyses of variance (MANOVAs)

A number of MANOVAs were performed in Grade 2, where the dependent variable was more than one, to see whether there was a school and/or gender effect on batteries of tests that examine the same skill: PSTM, CWM, L1 verbal intelligence, and FL aptitude. The index of Inhibition was tested separately as the number of participants for the SSRT changed from 47 to 46 when an additional participant from the experimental group was excluded during the finding of outliers for this particular task. For the variables SSRT and YLAT_Total, two-way ANOVAs were performed. When the grouping variable was sex no gender effect was found. The MANOVA findings confirmed the outperformance of the experimental group in the majority of the tests (see Appendix 9).

8.10 The Listening span and Recall task results

8.10.1 The experimental data

As regards the experimental group, of the 49 participants that took part in the research, two students were excluded from the study. They failed to score within the range of the mean ±2 SDs in more than 35% of the tasks in either grade and, thus, their scores were excluded from any further statistical analysis.

As it is clear from Figure 2 below, of the 47 students that completed the task in Grade 1, 30 students (63.8%) passed on to the two-sentence level, 5 students (10.6%) to the three-sentence level, and 0 students (0%) to the four-sentence level. In Grade 2, of the 47 students, 38 students (80.9%) passed on to the two-sentence level, 10 students (21.3%) to the three-sentence level, and 1 student (2.1%) to the four-sentence level. In other words, in Grade 2 the number of students that moved from one level to
the next recorded a 100% increase in some cases (levels 2 and 3). Also, in Grade 2 a considerable number of students managed to progress through a greater number of levels. More specifically, 1st grade students managed to pass through level 3 and 2nd grade students through level 4.

Figure 2: Experimental group, Grades 1 and 2 (no. of students per grade & level)

The chi-square analysis revealed a statistically significant within-group difference from level 1 to level 2 for Grade 1 ($\chi^2 = 16.428, df = 1, p < .001$) and Grade 2 ($\chi^2 = 25.422, df = 1, p < .001$) (see Appendix 10).

8.10.2 The control data

Figure 3 below shows that in Grade 1, of the 49 students participating in the task, 26 students (53.1%) passed on to the two-sentence level, 3 students (6.1%) to the three-sentence level, and 0 students (0%) to the four-sentence level. In Grade 2, of the 49 students, 38 students (77.6%) passed on to the two-sentence level, 7 students (14.3%) to the three-sentence level, and 1 student (2%) to the four-sentence level. As with the experimental group, a significantly higher number of students from the control group moved from one level to the next in Grade 2 (levels 2 and 3).
**Figure 3: Control group, Grades 1 and 2 (no. of students per grade & level)**

The chi-square analyses revealed a statistically significant within-group difference a) from level 1 to level 2 for Grade 1 ($\chi^2 = 12.349$, $df = 1$, $p < .001$) and Grade 2 ($\chi^2 = 29.968$, $df = 1$, $p < .001$), and b) from Grade 1 to Grade 2 at Level 1 ($\chi^2 = 6.485$, $df = 1$, $p < .05$). See Appendix 11 for greater detail.

Chi-square analyses were also performed between groups per level and grade, that revealed no statistically significant differences (see Appendix 12).

### 8.10.3 Correct (XX+OX) vs incorrect word recall (XO+OO) (%) per group, grade, and level

Table 14 presents the scores with respect to percentage accuracy, as these were obtained by each group for XX+OX and XO+OO response-types and level of difficulty, in Grades 1 and 2. Chi-square analyses were performed within and between groups per level and grade, that revealed no statistically significant differences (see Appendix 13). When the grouping variable was sex, no significant differences emerged in either school (see Appendix 14). Descriptive statistics are displayed in the table below, as regards the correct vs incorrect recall of the lexically unrelated word:
**Table 14: Correct (XX+OX) vs Incorrect (XO+OO) word recall: response types (%) and SDs**

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Level 1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XX+OX</strong></td>
<td><strong>XX+OX</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>85.53 (22.97)</td>
<td>82.48 (26.31)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>90.11 (23.71)</td>
<td>92.86 (12.31)</td>
</tr>
<tr>
<td><strong>XO+OO</strong></td>
<td><strong>XO+OO</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>14.47 (22.96)</td>
<td>17.52 (26.31)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>9.89 (23.70)</td>
<td>7.14 (12.31)</td>
</tr>
<tr>
<td>Level 2</td>
<td>Level 2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XX+OX</strong></td>
<td><strong>XX+OX</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>35.00 (29.97)</td>
<td>33.33 (27.79)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>45.22 (32.33)</td>
<td>46.49 (31.01)</td>
</tr>
<tr>
<td><strong>XO+OO</strong></td>
<td><strong>XO+OO</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>64.99 (29.97)</td>
<td>66.66 (27.79)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>54.77 (32.34)</td>
<td>53.50 (31.01)</td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XX+OX</strong></td>
<td><strong>XX+OX</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>10.00 (13.69)</td>
<td>8.33 (14.43)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>15.00 (22.50)</td>
<td>15.00 (30.14)</td>
</tr>
<tr>
<td><strong>XO+OO</strong></td>
<td><strong>XO+OO</strong></td>
</tr>
<tr>
<td>Grade 1</td>
<td>Grade 1</td>
</tr>
<tr>
<td>90.00 (13.69)</td>
<td>91.67 (14.43)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Grade 2</td>
</tr>
<tr>
<td>85.00 (22.50)</td>
<td>85.00 (30.14)</td>
</tr>
</tbody>
</table>

What is clear from this table is that the higher the level the more difficult it was for the majority of the students to successfully recall the lexically unrelated word(s). This is natural because the task imposes increasingly higher demands from level to level upon the learners. In level 1 they had to decide on the truth value of a sentence and then immediately recall the unrelated word. In level 1, word-recall accuracy takes the lead with respect to the WM interference task, as this is the easiest of all levels in terms of memory load. In levels 2 and 3, greater interference effects were reported as participants progressed to levels of increased difficulty. Level 4 was not included in the table given above, as only one student from each group managed to get to this level, in Grade 2. Even though word-recall for second graders was found to be marginally better than that of first graders, a similar trend with the one described above was found for Grade 2.
Two-way ANOVA analyses were conducted with grade and school as the between-subjects fixed factors, and percentage scores (%) for response-types (XX+OX and XO+OO) as the dependent variables. These revealed a significant within-group difference for Grade 2 at Level 1 \((F(1,188) = 5.561, \ p = .019)\) and \((F(1,188) = 5.559, \ p = .019)\), respectively, and at Level 2 \((F(1,128) = 4.700, \ p=.032)\) and \((F(1,188)=4.698, \ p = .032)\), respectively) (see Appendix 15). No between-groups differences were found.

### 8.11 The English Vocabulary Test

Performance in FL vocabulary was assessed by the implementation of two measures, a receptive and a productive test (see appendices 8 and 9, chapter 7). Both versions of the tests were designed for the purposes of this research, and were based on the material covered in the experimental group, in Grades 1 and 2. The test was taken only by the experimental group, as the control group is not yet exposed to FL English. Of the 49 participants that took part in the research, the totality of the scores were considered for statistical analysis, i.e. no outlying scores were excluded, as we wanted to see the performance of all the students in the test. The raw scores of the students along with time duration for each test (in seconds) appears in Appendix 16, chapter 8.

#### 8.11.1 FL_Receptive: Maximum score for this test was 15. The scores of the participants were distributed in three categories: a) the first captures those that were \(\geq 7\) (below the median quartile), b) the second captures the range of scores between the median and lower limit of the upper quartile (i.e. 8-11, hereafter referred as M-Q3 range), while c) the third category displays includes the scores that ranged from 12-15 (the upper quartile). Figure 4 illustrates the performance of the informants, with respect to time duration and score. Note that each cycle represents one case.
What becomes clear from Figure 4 is that of the 49 students that took the test, 4 students (8.2%) scored below median (≥7) and completed the test in less than 172 seconds (range 151-172 secs), 11 students (22.4%) scored in the M-Q3 range (8-11) and completed the test in less than 165 seconds (range 95-161 secs), while 34 students (69.4%) scored above the upper quartile (between 12-15). There was one student who scored high but was extremely slow (198 secs), 6 students scored high and were very fast (range: 60-81), 27 students scored around the average (range: 85-165 secs).

8.11.2 FL\_Productive: Maximum score for this test was 19. As with FL\_Receptive, the scores of the participants were distributed in three categories: a) below the median quartile (≥9), b) the M-Q3 range (between 10-14), and c) above the upper quartile (between 15-19). Figure 5 below portrays the performance of the participants in this test, in terms of score and time duration:
Figure 5 shows that 15 students (30.6%) scored below the median quartile (≥9). The majority of them (i.e. 13 students) needed 161-253 seconds to complete the task, while 2 students reached or exceeded the 350 seconds (295 and 363 secs respectively), 18 students (36.7%) scored in the M-Q3 range (between 10-14). Only 1 student was rather slow (355 secs), while time duration for the other 17 students was around average (166-277 secs). 16 students (32.7%) scored above the upper quartile (between 15-19). One student was both highly accurate and extremely fast (100 secs), 12 students finished the task within the test’s average (range: 148-199 secs), while another 3 students needed 218-247 secs to complete the test.

8.11.3 The participants’ combined performance in the two FL skills

Table 15 below shows the combined performance of participants in the two tests. What appears in bold, marks the number of students whose performance did not mark any difference (i.e. they performed equally poor, average, or good) in the two tests: 4
students scored below median (≥7_Recept., ≥9_Product.), 2 students scored between 8-11 in FL_Receptive and between 10-14 in FL_Productive, while 16 students scored above the upper quartile (above 12 in FL_Receptive, above 15 in FL_Productive).

Table 15: The combined performance of the experimental group in the two FL tests

<table>
<thead>
<tr>
<th>Score</th>
<th>Productive_recoded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥9</td>
<td>10-14</td>
</tr>
<tr>
<td>≥7</td>
<td>Count</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>8.2%</td>
</tr>
<tr>
<td>8-11</td>
<td>Count</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>18.4%</td>
</tr>
<tr>
<td>12-15</td>
<td>Count</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>4.1%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>30.6%</td>
</tr>
</tbody>
</table>

As regards the number of students who scored lower in the productive test, the picture has as follows: of the 11 students (22.4%) that scored in the M-Q3 range (between 8-11) in the receptive test, 9 students (18.4%) scored below the median quartile (≥9) in the productive. Of the 34 students (69.4%) that scored above the upper quartile in the receptive test (between 12-15), 2 students (4.1%) scored below median (≥9) and 16 students (32.7%) scored in the M-Q3 range (between 10-14) in the productive test.

Overall, it seems that the FL_Productive test was more difficult than the FL_Receptive one. The overall numbers show that of the 69.4% of the students who
managed to score above the upper quartile in the receptive test, only 32.7% scored equally high in the productive one. The difficulty students experienced in the production of some language of their own in the FL, also affected the overall time they needed to complete the task. While students finished the productive test in a range of 100 to 363 seconds, they needed almost half this time for the receptive one (overall range: 60-198 seconds).

The chi-square analysis revealed a statistically significant difference in the performance of the students between the two tests, i.e. between the cases that did not manage to go to a higher quartile (the upper part of the diagonal that is marked with bold numbers, Table 15) with those that actually scored significantly worse in the productive test (the lower part of the diagonal) $\chi^2 = 33.051, df = 4, p < .001$.

8.12 Response Inhibition: The stop-signal task

8.12.1 Outlier analysis

Following Huizinga (2006), to remove the outlying trials the following criteria were taken into account. In the cases where the proportion of correct inhibits (no response to the red arrow) was lower than 20% (=10 trials) or higher than 80% (=40 trials), the results were excluded. One such case was found in the experimental data in Grade 1. Participant no. 37 correctly inhibited 6 out of the 50 stop-signal trials. This data were excluded from any further analysis. Where RT served as a dependent measure, a two-step trimming procedure was then performed. In Step One all incorrect responses [i.e. Green: conditions 2 (wrong arrow direction) and 3 (no response) and Red: condition 2 (incorrect response to the stop-signal)] were excluded. In Step Two, according to the procedure followed by van den Wildenberg, van der Molen, & Logan (2002), responses with a latency exceeding the school’s mean value of all trials (per grade) by ±2.5 SDs, were also excluded from the RT analyses. The trials that were removed
belonged to the lower end of the range in all cases. Of the 38,000 trials (both groups and grades), 8,500 trials were excluded (22.37%) in Step One. Of the remaining 29,500 (77.63%) trials, 42 additional trials (0.14%) were removed in Step Two. In total, the excluded trials amounted to 22.51% of all trials. Finally, when the SSRT\textsubscript{mean} fell below the 200 ms threshold usually found in the relevant bibliography for young adults (Band, van der Molen, & Logan, 2003; Bedard et al., 2002; Logan & Cowan, 1984, 315; Miyake et al., 2000) it was replaced by the SSRT\textsubscript{mean} of the school for the grade. In Grade 1, the experimental group recorded 6 such cases and the control only 4. In Grade 2 these numbers changed into 13 and 8, respectively.

8.12.2 The index of Response Inhibition: The stop-signal reaction time (SSRT)

The variable of Inhibition was indexed by the mean latency of the SSRT. This was estimated by the new calculation of SSRT which is easier to compute and equally valid (Logan, Schachar, & Tannock, 1997) than previous ones (Logan & Cowan, 1984). It was computed as follows: SSRT\textsubscript{mean} = Go-RT\textsubscript{mean} - SOA\textsubscript{mean}. The mean values of all variables were computed for each participant, per group and per grade.

8.12.3 Go-RT\textsubscript{mean} and SSRT\textsubscript{mean} in the two groups

The descriptive statistics of the two dependent variables (after the two-step trimming procedure) are given in Table 16 below:
## Chapter 8: Data Analysis

### Table 16: Mean and SDs Go_RT and SSRT (per group & grade)

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>Grade</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Go_RT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp.</td>
<td>1</td>
<td></td>
<td>631.33</td>
<td>84.67</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>643.14</td>
<td>89.06</td>
<td>46</td>
</tr>
<tr>
<td>Con.</td>
<td>1</td>
<td></td>
<td>653.42</td>
<td>51.83</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>648.89</td>
<td>66.03</td>
<td>49</td>
</tr>
<tr>
<td><strong>SSRT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp.</td>
<td>1</td>
<td></td>
<td>309.68</td>
<td>49.62</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>275.19</td>
<td>55.69</td>
<td>46</td>
</tr>
<tr>
<td>Con.</td>
<td>1</td>
<td></td>
<td>298.45</td>
<td>58.61</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>268.74</td>
<td>50.69</td>
<td>49</td>
</tr>
</tbody>
</table>

A graphic illustration of the overall mean values of Go_RT<sub>mean</sub> and SSRT<sub>mean</sub>, per group and grade, is given in Figures 6 and 7:

![Estimated Marginal Means of Go_RT Mean](image)

**Figure 6:** The overall Go-RT<sub>mean</sub> values (per group & grade)
The overall SSRT<sub>mean</sub> values (per group & grade)

8.12.4 The overall correlation matrices of the two groups (incl. the SSRT)

As already mentioned in section 8.4, Pearson correlation coefficients were computed separately for the index of Inhibition (SSRT) with the rest of the measures obtained in Grades 1 and 2. This was done because in the exclusionary criteria applied for this index a third participant (no. 37) was excluded from the experimental group, leaving the group with 46 participants. First, z-scores were calculated to be able to compare variables that examine different aspects in a test, e.g. accuracy and RT. Tables 17 and 18 provide the resulting correlation matrices of the two schools for Grade 1 (experimental and control), while Tables 19 and 20 give the matrices of the two schools for Grade 2. Appendix 17 gives a detailed account of the four matrices.
Table 17: Experimental group (n=46): Correlations between the totality of the measures, Grade 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall_Forward</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Nonce_Gr.</td>
<td>.40**</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Nonce_Eng.</td>
<td>.40**</td>
<td>.27</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Digit recall_Backwards</td>
<td>.30*</td>
<td>.20</td>
<td>-.08</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Listening Recall</td>
<td>.40**</td>
<td>.20</td>
<td>.22</td>
<td>.37*</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>SSRT</td>
<td>-.08</td>
<td>.14</td>
<td>-.20</td>
<td>-.06</td>
<td>-.19</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>DVIQ I_Voc.</td>
<td>.23</td>
<td>-.11</td>
<td>.25</td>
<td>.12</td>
<td>.41**</td>
<td>-.34*</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>DVIQ I_Met.</td>
<td>.16</td>
<td>.12</td>
<td>-.11</td>
<td>.20</td>
<td>.21</td>
<td>.02</td>
<td>.02</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>DVIQ II_Met.</td>
<td>.20</td>
<td>.03</td>
<td>-.10</td>
<td>.14</td>
<td>-.16</td>
<td>.28</td>
<td>.37*</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Memory</td>
<td>.10</td>
<td>.13</td>
<td>.16</td>
<td>-.03</td>
<td>-.01</td>
<td>.11</td>
<td>-.18</td>
<td>.06</td>
<td>.26</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Classification</td>
<td>.24</td>
<td>.26</td>
<td>.17</td>
<td>.26</td>
<td>.10</td>
<td>-.07</td>
<td>.23</td>
<td>-.08</td>
<td>.21</td>
<td>.07</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Differences</td>
<td>-.04</td>
<td>.10</td>
<td>.06</td>
<td>.19</td>
<td>.21</td>
<td>-.12</td>
<td>.31*</td>
<td>.30*</td>
<td>.23</td>
<td>.12</td>
<td>.01</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Paired Associates</td>
<td>.21</td>
<td>.09</td>
<td>-.19</td>
<td>.28</td>
<td>.23</td>
<td>.22</td>
<td>.29*</td>
<td>.27</td>
<td>.18</td>
<td>.06</td>
<td>.08</td>
<td>.21</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Semantic Integration</td>
<td>.16</td>
<td>-.11</td>
<td>.34*</td>
<td>.21</td>
<td>.18</td>
<td>-.27</td>
<td>.16</td>
<td>-.10</td>
<td>-.07</td>
<td>.28</td>
<td>-.07</td>
<td>.26</td>
<td>.04</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Story Sequences</td>
<td>.01</td>
<td>.03</td>
<td>.06</td>
<td>.05</td>
<td>.24</td>
<td>-.24</td>
<td>.08</td>
<td>.08</td>
<td>.41**</td>
<td>-.09</td>
<td>.13</td>
<td>.17</td>
<td>-.05</td>
<td>.02</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>YLAT_Total</td>
<td>.17</td>
<td>.20</td>
<td>.12</td>
<td>.39**</td>
<td>.17</td>
<td>-.11</td>
<td>.36*</td>
<td>.03</td>
<td>.32*</td>
<td>.11</td>
<td>.80**</td>
<td>.26</td>
<td>.34*</td>
<td>.15</td>
<td>.24</td>
<td>1</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).
Table 18: Control group (n=49): Correlations between the totality of the measures, Grade 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Digit recall_Forward</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Nonce_Gr.</td>
<td>.32*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Nonce_Eng.</td>
<td>.33*</td>
<td>.58**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Digit recall_Backwards</td>
<td>.06</td>
<td>.17</td>
<td>.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Listening Recall</td>
<td>.30*</td>
<td>.07</td>
<td>.15</td>
<td>.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>SSRT</td>
<td>-.25</td>
<td>-.20</td>
<td>-.14</td>
<td>-.12</td>
<td>-.28*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DVIQ I_Voc.</td>
<td>.01</td>
<td>.05</td>
<td>.07</td>
<td>.07</td>
<td>.16</td>
<td>.23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DVIQ I_Met.</td>
<td>.09</td>
<td>.19</td>
<td>.35*</td>
<td>.31*</td>
<td>.14</td>
<td>.09</td>
<td>.38**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DVIQ II_Met.</td>
<td>-.01</td>
<td>.17</td>
<td>.16</td>
<td>-.09</td>
<td>-.09</td>
<td>.23</td>
<td>.40**</td>
<td>.28*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Memory</td>
<td>.11</td>
<td>.23</td>
<td>.17</td>
<td>-.01</td>
<td>.03</td>
<td>-.09</td>
<td>-.08</td>
<td>.11</td>
<td>.07</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Classification</td>
<td>.08</td>
<td>.10</td>
<td>.01</td>
<td>.05</td>
<td>.17</td>
<td>-.05</td>
<td>.26</td>
<td>.25</td>
<td>.18</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Differences</td>
<td>.02</td>
<td>.26</td>
<td>.19</td>
<td>.15</td>
<td>.06</td>
<td>-.07</td>
<td>.26</td>
<td>.10</td>
<td>.20</td>
<td>.07</td>
<td>.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Paired Associates</td>
<td>.23</td>
<td>.26</td>
<td>.13</td>
<td>-.06</td>
<td>-.05</td>
<td>-.10</td>
<td>.16</td>
<td>.19</td>
<td>.44**</td>
<td>.12</td>
<td>.28*</td>
<td>.09</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Semantic Integration</td>
<td>.17</td>
<td>.13</td>
<td>.10</td>
<td>.03</td>
<td>.20</td>
<td>-.06</td>
<td>.25</td>
<td>.34*</td>
<td>.28</td>
<td>.07</td>
<td>.16</td>
<td>-.02</td>
<td>.28</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Story Sequences</td>
<td>.01</td>
<td>-.04</td>
<td>-.16</td>
<td>-.02</td>
<td>.09</td>
<td>.08</td>
<td>.05</td>
<td>-.05</td>
<td>-.11</td>
<td>.10</td>
<td>.18</td>
<td>.13</td>
<td>-.09</td>
<td>.41**</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>YLAT_Total</td>
<td>.14</td>
<td>.17</td>
<td>.05</td>
<td>.04</td>
<td>.20</td>
<td>-.06</td>
<td>.34*</td>
<td>.29*</td>
<td>.31*</td>
<td>.25</td>
<td>.91**</td>
<td>.29*</td>
<td>.38**</td>
<td>.30*</td>
<td>.30*</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01
### Table 19: Experimental group (n=46): Correlations between the totality of the measures, Grade 2

|       | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-------|---------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| FL_Voc_Total |       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| FL_Receptive  | .71**  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| FL_Productive | .96**  | .65** |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Digit recall_Foward | .24  | .22 | .25 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Nonce_Gr.     | .41**  | .43** | .42** | .49** |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Nonce_Eng.    | .16    | .34* | .16 | .44** | .41** |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Digit recall_Backwards | .14  | .12 | .21 | .35* | .28 | .29 |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Listening Recall | .37*  | .45** | .43** | .36* | .34* | .32* |    |    |    |    |    |    |    |    |    |    |    |    |    |
| SSRT         | -.14   | -.22 | -.20 | -.20 | -.09 | -.24 | -.28 | -.43** |    |    |    |    |    |    |    |    |    |    |    |
| DVIQ I_Voc.  | .23    | .29* | .31* | .10 | .29 | .22 | .16 | .34* | -.07 |    |    |    |    |    |    |    |    |    |    |
| DVIQ I_Met.  | .21    | .28 | .22 | .12 | .08 | .09 | .46** | .22 | -.33* | .24 |    |    |    |    |    |    |    |    |    |
| DVIQ II_Met. | .17    | .27 | .15 | .00 | .20 | .12 | .25 | .38** | -.18 | .24 | .57** |    |    |    |    |    |    |    |    |
| Memory       | .20    | .26 | .21 | .10 | .08 | .22 | .08 | .32* | -.49** | .27 | .15 | .00 |    |    |    |    |    |    |    |
| Classification | .06   | -.01 | -.03 | .04 | -.19 | -.16 | .25 | -.05 | .13 | -.25 | .17 | .21 | -.19 |    |    |    |    |    |    |
| Differences  | .29    | .28 | .30* | .28 | .25 | .13 | .31* | .27 | -.20 | .49** | .36* | .19 | .21 | .05 |    |    |    |    |    |
| Paired Associates | .17   | .17 | .18 | .17 | .20 | .00 | .09 | .29 | -.01 | .03 | .06 | .21 | .06 | .27 | -.01 |    |    |    |    |
| Semantic Integration | .24  | .08 | .31* | .30* | .27 | .14 | .20 | .46** | -.39** | .38* | .14 | .09 | .36* | -.20 | .30* | .00 |    |    |    |
| Story Sequences | .10   | .23 | .04 | .13 | .20 | .17 | .08 | .06 | .08 | .09 | .29 | .28 | -.12 | .18 | .16 | .16 | -.13 |    |    |
| YLAT_Total   | .32*   | .21 | .28 | .28 | .14 | .07 | .42** | .29* | -.16 | .13 | .32* | .29 | .24 | .72** | .53** | .45** | .28 | .28 |    |
# Chapter 8: Data Analysis

Table 20: Control group (n=49): Correlations between the totality of the measures, Grade 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Digit recall_Forward</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Nonce_Gr.</td>
<td>.37**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Nonce_Eng.</td>
<td>.21</td>
<td>.64**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Digit recall_Backwards</td>
<td>.35*</td>
<td>.37**</td>
<td>.36*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Listening Recall</td>
<td>.36*</td>
<td>.36*</td>
<td>.21</td>
<td>.21</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SSRT</td>
<td>-.01</td>
<td>-.24</td>
<td>-.08</td>
<td>-.29*</td>
<td>-.32*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DVIQ I_Voc.</td>
<td>-.15</td>
<td>.12</td>
<td>.07</td>
<td>-.06</td>
<td>-.04</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DVIQ I_Met.</td>
<td>.32*</td>
<td>.42**</td>
<td>.42**</td>
<td>.19</td>
<td>.31*</td>
<td>-.03</td>
<td>.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DVIQ II_Met.</td>
<td>.24</td>
<td>.48**</td>
<td>.34*</td>
<td>.24</td>
<td>.28</td>
<td>.07</td>
<td>.26</td>
<td>.57**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Memory</td>
<td>.22</td>
<td>.24</td>
<td>.19</td>
<td>.23</td>
<td>.28</td>
<td>-.30*</td>
<td>.13</td>
<td>.30*</td>
<td>.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Classification</td>
<td>-.06</td>
<td>-.06</td>
<td>.00</td>
<td>-.10</td>
<td>-.05</td>
<td>.14</td>
<td>.04</td>
<td>-.04</td>
<td>.22</td>
<td>.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Differences</td>
<td>-.16</td>
<td>-.05</td>
<td>-.10</td>
<td>.02</td>
<td>.29*</td>
<td>-.00</td>
<td>.14</td>
<td>.27</td>
<td>-.04</td>
<td>.23</td>
<td>-.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Paired Associates</td>
<td>.24</td>
<td>.32*</td>
<td>.29*</td>
<td>.29*</td>
<td>.28*</td>
<td>-.31*</td>
<td>-.16</td>
<td>.51**</td>
<td>.24</td>
<td>.27</td>
<td>.06</td>
<td>.10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Semantic Integration</td>
<td>.04</td>
<td>.13</td>
<td>.12</td>
<td>.33*</td>
<td>.13</td>
<td>-.05</td>
<td>.07</td>
<td>.24</td>
<td>.31*</td>
<td>.17</td>
<td>.26</td>
<td>.17</td>
<td>.18</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Story Sequences</td>
<td>.19</td>
<td>.16</td>
<td>.11</td>
<td>.14</td>
<td>.09</td>
<td>-.01</td>
<td>-.01</td>
<td>.07</td>
<td>.05</td>
<td>-.08</td>
<td>-.07</td>
<td>-.17</td>
<td>.01</td>
<td>.09</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>YLAT_Total</td>
<td>.08</td>
<td>.12</td>
<td>.07</td>
<td>.08</td>
<td>.21</td>
<td>-.08</td>
<td>.24</td>
<td>.28</td>
<td>.29*</td>
<td>.22</td>
<td>.67**</td>
<td>.09</td>
<td>.33*</td>
<td>.44**</td>
<td>-.04</td>
</tr>
</tbody>
</table>
8.12.5 The correlations of the SSRT with the other variables

As the rest of the measures have been discussed earlier in this chapter, we will only focus on whether the SSRT established any relations with the other measures, in either grade and school.

First of all, it has to be mentioned that all the inhibition mechanism index (the SSRT) established inversely proportional relations with the other variables because of its nature. In all other tests, a higher score reflects a better performance while in the case of the SSRT, a larger RT means that participants find it more difficult to inhibit irrelevant or distracting information from intruding into their WM with a consequent overload of their WM, often instantiated in poor performance during learning or test-taking. Thus, the shorter the SSRT the better the performance of participants in the other tests.

In Grade 1, the SSRT correlated with DVIQ I_Voc. \((z = -.34, p < .05)\) in the experimental group, while in the control group it was correlated with Listening Recall \((z = -.28, p < .05)\), the measure that taps the second EF, i.e. WM.

In Grade 2, what is clear from the experimental data is that SSRT established moderate to strong relations with indices that examine rote memory, recognition and recoding ability and L1 metalinguistic knowledge. In particular, it as correlated with Listening Recall \((z = -.43)\), Memory \((z = -.49)\) and Semantic Integration \((z = -.39)\) at the level of .01 significance, while it established a more moderate relation with DVIQ I_Met. \((z = -.33, p < .05)\).

The control data yielded similar relations between the SSRT and the other variables, which were all at the level of .05 significance. The inhibition index was correlated with the two CWM measures, namely Digit recall_Backwards \((z = -.29)\) and Listening Recall \((z = -.32)\). Also, it was correlated with two cognitive measures, namely Memory \((z = -.30)\) and Paired Associates \((z = -.31)\). The next step was to
perform regression analyses with the experimental data from Grade 2 and examine whether the measures that examine the two EFs primarily explained variance in the scores of the FL tests taken but also of the other skills studied in this thesis.

8.12.6 Regression analyses: Grade 1

In Grade 1, we first examined whether the independent variables that tap CWM (Listening Recall, Digit recall_Backwards) and Inhibition (SSRT) could explain an amount of variance in the scores of L1 verbal intelligence (DVIQ I_Voc., DVIQ I_Met., DVIQ II_Met.: dependent variables). Of the independent variables that tap the two EFs, only a CWM measure, i.e. Listening Recall accounted for 15% of the variance in the scores of DVIQ I_Voc., F(1, 44) = 8.962; \( p < .01 \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall</td>
<td>.150</td>
<td>8.962</td>
<td>1</td>
<td>44</td>
<td>.411</td>
<td>.005</td>
</tr>
</tbody>
</table>

No particular difference was detected when the SSRT index was entered into the regression analysis as one of the independent variables. It is reminded that Listening Recall had initially accounted for 12.8% of the variance in the scores of DVIQ I_Voc.

8.12.7 Regression analyses: Grade 2

In the same vein, in Grade 2 we first examined whether variance in the scores of FL_Voc._Total or separately FL_Receptive and FL_Producutive (the dependent variables) could be explained by the independent variables that tap CWM and Inhibition.

8.12.7.1 FL performance as the dependent variable

Of the independent variables that tap CWM and Inhibition, only Listening Recall accounted for a) 12% of the variance in the scores of FL_Voc._Total, F(1, 44) = 7.134; \( p < .05 \), b) 18.4% of the variance in the scores of FL_Receptive, F(1, 44) = 11.168; \( p < 
.01, and c) 16.9% of the variance in the scores of FL_Productive, F(1,44) = 10.167; p < .01:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall (for FL_Voc._Total)</td>
<td>.120</td>
<td>7.134</td>
<td>1</td>
<td>44</td>
<td>.374</td>
<td>.011</td>
</tr>
<tr>
<td>Listening Recall (for FL_Receptive)</td>
<td>.184</td>
<td>11.168</td>
<td>1</td>
<td>44</td>
<td>.450</td>
<td>.002</td>
</tr>
<tr>
<td>Listening Recall (for FL_Productive)</td>
<td>.169</td>
<td>10.167</td>
<td>1</td>
<td>44</td>
<td>.433</td>
<td>.003</td>
</tr>
</tbody>
</table>

No particular differences were detected when the independent variable of SSRT was entered into the regression analyses. Listening Recall had initially accounted for 11.3% (FL_Voc._Total), 17.7% (FL_Receptive) and 15.8% (FL_Productive) of the variance in the scores of the FL measures.

**8.12.7.2 L1 verbal intelligence as the dependent variable**

Then, we examined whether variance in the scores of L1 verbal intelligence could be predicted by any of the independent variables that tap CWM and Inhibition. Of the independent variables that tap CWM, Listening Recall accounted a) for 9.4% of the variance in the scores of DVIQ I_Voc., F(1,44) = 5.693; p < .05 and b) for 12.8% of the variance in the scores of DVIQ II_Met., F(1,44) = 7.578; p < .01. Digit recall_Backwards accounted for 19.3% of the variance in the scores of DVIQ I_Met., F(1,44) = 11.757; p < .01.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Recall (for DVIQ I_Voc.)</td>
<td>.094</td>
<td>5.693</td>
<td>1</td>
<td>44</td>
<td>.338</td>
<td>.021</td>
</tr>
<tr>
<td>Digit recall_Backwards (for DVIQ I_Met.)</td>
<td>.193</td>
<td>11.757</td>
<td>1</td>
<td>44</td>
<td>.459</td>
<td>.001</td>
</tr>
<tr>
<td>Listening Recall (for DVIQ II_Met.)</td>
<td>.128</td>
<td>7.578</td>
<td>1</td>
<td>44</td>
<td>.383</td>
<td>.009</td>
</tr>
</tbody>
</table>
No particular differences were detected when the independent variable of the SSRT entered into the regression analyses. Listening Recall had initially accounted for 9.5% of the variance in the scores of DVIQ I_Voc. and 12.4% of the variance in the scores of DVIQ II_Met. Digit recall_Backwards had accounted for 18.5% of the variance in the scores of DVIQ I_Met.

### 8.12.7.3 FL aptitude as the dependent variable

Then, we examined whether individual differences in FL aptitude, i.e. cognitive skills, could be explained by independent variables that tap CWM or Inhibition. Of the independent variables that tap CWM and Inhibition, the Inhibition index, i.e. SSRT negatively accounted for 21.7% of the variance in the scores of Memory, $F(1,44) = 13.499; p < .01$. Digit recall_Backwards accounted for 16.1% of the variance in the scores of YLAT_Total, $F(1,44) = 9.646; p < .01$ and for 7.4% of the variance in the scores of Differences, $F(1,44) = 4.602; p < .05$. Finally, Listening Recall accounted for 19% of the variance in the scores of Semantic Integration, $F(1,44) = 11.563; p < .01$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSRT (for Memory)</td>
<td>.217</td>
<td>13.499</td>
<td>1</td>
<td>44</td>
<td>-.485</td>
<td>.001</td>
</tr>
<tr>
<td>Digit recall_Backwards (for Differences)</td>
<td>.074</td>
<td>4.602</td>
<td>1</td>
<td>44</td>
<td>.308</td>
<td>.037</td>
</tr>
<tr>
<td>Listening Recall (for Semantic Integration)</td>
<td>.190</td>
<td>11.563</td>
<td>1</td>
<td>44</td>
<td>.456</td>
<td>.001</td>
</tr>
<tr>
<td>Digit recall_Backwards (for YLAT_Total)</td>
<td>.161</td>
<td>9.646</td>
<td>1</td>
<td>44</td>
<td>.424</td>
<td>.003</td>
</tr>
</tbody>
</table>
CHAPTER 8: DATA ANALYSIS

The only difference detected here when the SSRT index entered into the regression analyses was that SSRT and not Listening Recall (as was the initial finding) accounted for an amount of variance in the scores of Memory. In the initial correlations, Listening Recall had accounted for 8.3% of the variance in the scores of Memory. Apart from that, no other differences were found as Digit recall_Backwards had initially accounted for 16.2% and 7.6% of the variance in the scores of YLAT_Total and Differences respectively, while Listening Recall had accounted for 17.4% of the variance in the scores of Semantic Integration.

8.12.7.4 PSTM as the dependent variable

Then, we examined whether individual differences in PSTM could be predicted by independent variables that tap CWM and/or Inhibition. Of the independent variables that tap CWM, Digit recall_Backwards accounted for 10% of the variance in the scores of Digit recall_Forward, F(1,44) = 6.012; p < .05. Listening Recall accounted for 11% and 9.6% respectively, of the variance in the scores of Nonce_Gr., F(1,44) = 6.546 p < .05 and Nonce_Eng, F(1,44) = 5.752; p < .05.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Adjusted R Square</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>Standardised Beta</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit recall_Backwards (for Digit recall_Forward)</td>
<td>.100</td>
<td>6.012</td>
<td>1</td>
<td>44</td>
<td>.347</td>
<td>.018</td>
</tr>
<tr>
<td>Listening Recall (for Nonce_Gr.)</td>
<td>.110</td>
<td>6.546</td>
<td>1</td>
<td>44</td>
<td>.360</td>
<td>.014</td>
</tr>
<tr>
<td>Listening Recall (for Nonce_Eng.)</td>
<td>.096</td>
<td>5.752</td>
<td>1</td>
<td>44</td>
<td>.340</td>
<td>.021</td>
</tr>
</tbody>
</table>

No particular differences were detected when SSRT entered into the regression analyses. Listening Recall had initially accounted for 9.6% of the variance in the
scores of Nonce_Eng. and the same amount of the variance in the scores of Nonce_Gr., while Digit recall_Backwards had accounted for 9.9% of the variance in the scores of Digit recall_Forward.

Overall, the findings suggest the vital contribution of CWM to the performance of participants in a number of verbal tasks that relate to their L1 or to the FL as well as to the performance of non-verbal tasks. The Inhibition index explained 21.7% of the variance in the scores of a cognitive measure that taps participants’ rote memory ability.

8.12.8 Multivariate analyses of variance (MANOVAs)

The overall mean values of Go-RT\textsubscript{mean} and SSRT\textsubscript{mean} (the dependent variables), with respect to school and grade, were subjected into multivariate analyses of variance (MANOVAs, see Appendix 18, Chapter 8). The results show that the mean values of the SSRTs irrespective of school, were significantly shorter in Grade 2, $F(1,186) = 20.243, p < .001$.

8.12.9 Repeated measures analyses of variance (ANOVAs)

In order to examine differences in response activation and response inhibition in the two schools in the course of the two years of the study, the overall mean values of Go-RTs and SSRTs, were submitted to repeated measures ANOVAs for each dependent variable, with school as the between-subjects factors. As predicted by the horse-race model (Logan & Cowan, 1984), the responses on stop-signal trials that escape Inhibition were executed faster than average go-signal responses. This holds for both schools and grades, as is clear from Table 16 above.

As already mentioned, the experimental group marked increasing Go\_RTs as they performed a delay strategy (focusing more on the accuracy aspect of this task) in Grade 2 (631.33→643.14). The reverse picture emerged for the control group
CHAPTER 8: DATA ANALYSIS

(653.42–648.89). However, no significant effect was found for our repeated measures factor ‘Grade’, \( F(1,93) = .357; p = .551 \text{ (n.s.)} \), or for an interaction between grade and school, \( (1,93) = 1.798; p = .183 \text{ (n.s.)} \).

With the other dependent variable, i.e. SSRT, a significant effect was found only for the factor ‘Grade’, \( F(1,93) = 48.816; p > .001 \) but not for the interaction between grade and school, \( (1,93) = .270; p = .605 \text{ (n.s.)} \). The results confirm the previous findings from the MANOVAs as they indicate that in Grade 2 both groups marked significantly shorter SSRTs. A detailed picture of the statistics of Go_RT and SSRT is given in Appendices 19 and 20, respectively.

8.12.10 Paired samples t-tests

According to the model the stop-signal paradigm follows, the SSRT is the critical parameter. The longer estimated SSRT the more difficult it is for participants to control (Logan & Cowan, 1984: 314). The paired samples t-tests conducted revealed a main effect of age on the SSRT\text{mean} which holds for both groups. SSRT\text{mean} was larger for 6-year-olds than 7-year-olds. The speed of inhibition increased with age as the difference in the mean values of the variable reached significant levels in the t-tests of both schools: experimental, \( t = 6.41, \ DF = 45, p < .001 \); control, \( t = 4.05, \ DF = 48, p < .001 \). Table 21 displays the descriptive statistics of the two groups for SSRT\text{mean} while Table 22 displays the groups’ t-test results across the two grades:

Table 21: Descriptive statistics for SSRT\text{mean} (per group & grade)

<table>
<thead>
<tr>
<th>Groups</th>
<th>SSRT 1-2</th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>SSRT 1</td>
<td>309.68</td>
<td>46</td>
<td>46.20</td>
</tr>
<tr>
<td></td>
<td>SSRT 2</td>
<td>275.19</td>
<td>46</td>
<td>46.96</td>
</tr>
<tr>
<td>Control</td>
<td>SSRT 1</td>
<td>298.45</td>
<td>49</td>
<td>56.11</td>
</tr>
<tr>
<td></td>
<td>SSRT 2</td>
<td>268.74</td>
<td>49</td>
<td>46.28</td>
</tr>
</tbody>
</table>
Table 22: *t*-test results for SSRT\textsubscript{mean} (per group & grade)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Differences of Mean</th>
<th>SD of differences</th>
<th>95% Confidence Interval of the Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>SSRT\textsubscript{mean}</td>
<td>34.48</td>
<td>36.46</td>
<td>23.66</td>
<td>45.31</td>
<td>6.414</td>
</tr>
<tr>
<td>Control (Grade 1-2)</td>
<td>29.71</td>
<td>51.32</td>
<td>14.97</td>
<td>44.45</td>
<td>4.052</td>
<td>48</td>
</tr>
</tbody>
</table>

The range of SSRT\textsubscript{mean} in Grades 1 and 2 is wider for the experimental group (309.68 - 275.19) than for the control (298.45 - 268.74).

8.12.11 The independent samples *t*-tests: Grade 2

To examine whether EFLL has any effect on the control of Inhibition, the SSRT\textsubscript{mean} of Grade 2 from the experimental and the control data were subjected to independent samples *t*-tests. The SSRT\textsubscript{mean} of the experimental group (Mean = 275.19, SD = 46.96) failed to reach significance (*t* = .674, DF = 93, *p* = .502) when compared to that of the control group (Mean = 268.74, SD = 46.28).

Tables 23 and 24 below give a detailed picture with respect to school and grade, for Go-RT\textsubscript{mean} and SSRT\textsubscript{mean} (in ms) and the proportion of participants’ corresponding errors to these variables. Note that incorrect responses for the Go-trial (i.e. the green arrow) were conditions 2 (commission error: wrong arrow direction chosen) and 3 (omission error: no response given), while for the stop-signal trial (i.e. the red arrow) was condition 2 (commission error: response given when it should be withheld):
### CHAPTER 8: DATA ANALYSIS

Table 23: Go-RT\textsubscript{mean}, SSRT\textsubscript{mean} and their error rates (%) (Grade 1, across the groups), *CI (=Confidence Interval)

<table>
<thead>
<tr>
<th>Group</th>
<th>Go-RT\textsubscript{mean} (SD = 84.67)</th>
<th>Incorrect responses (%) on Go-task trials</th>
<th>SSRT\textsubscript{mean} (SD = 46.20)</th>
<th>Incorrect inhibits (%) on stop-signal trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (n=46)</td>
<td>631.33 (606.18 – 656.47)</td>
<td>1111 (16.10 %)</td>
<td>309.68 (295.96 – 323.40)</td>
<td>1112 (48.3 %)</td>
</tr>
<tr>
<td>Control (n=49)</td>
<td>653.42 (638.53 – 668.31)</td>
<td>912 (12.41 %)</td>
<td>298.45 (282.33 – 314.57)</td>
<td>1183 (48.3 %)</td>
</tr>
</tbody>
</table>

Table 24: Go-RT\textsubscript{mean}, SSRT\textsubscript{mean} and their error rates (%) (Grade 2, across the groups), *CI (=Confidence Interval)

<table>
<thead>
<tr>
<th>Group</th>
<th>Go-RT\textsubscript{mean} (SD = 89.06)</th>
<th>Incorrect responses (%) on Go-task trials</th>
<th>SSRT\textsubscript{mean} (SD = 46.96)</th>
<th>Incorrect inhibits (%) on stop-signal trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (n=46)</td>
<td>643.14 (616.69 – 669.59)</td>
<td>1049 (15.20 %)</td>
<td>275.19 (261.25 – 289.14)</td>
<td>1108 (48.2 %)</td>
</tr>
<tr>
<td>Control (n=49)</td>
<td>648.89 (629.92 – 667.85)</td>
<td>824 (11.21 %)</td>
<td>268.74 (255.45 – 282.03)</td>
<td>1201 (49 %)</td>
</tr>
</tbody>
</table>

This is better illustrated in Figures 8 and 9 below:
Figure 8: Go-$RT_{\text{mean}}$ across the two grades (with respect to group)

Figure 9: SSRT$_{\text{mean}}$ across the two grades (with respect to group)

The mean accuracy of responding in the primary task ranged between 88 and 84% for the two groups in Grade 1 and between 89 and 85% in Grade 2. In addition, the mean accuracy of response inhibition ranged between 48-49% in the two grades for both groups. It appears that the tracking was quite successful as both groups successfully inhibited ~ 51-52% of the strop-signal trials. Error rates with respect to Go-task and stop-signal-task trials, are depicted in Figure 10 below:
CHAPTER 8: DATA ANALYSIS

Figure 10: Error rates on Go- and Stop-signal task trials (per group & grade)

It seems that in Grade 2 the experimental group managed to slightly decrease the proportion of incorrect inhibits (48.3→48.2%) while the control group increased this proportion (48.3→49%). This could be attributed to the delay strategy exercised only by the experimental group.

8.13 Chapter Summary

Numerous statistical tests were conducted with the learner data that were compiled, during the two years this study lasted.

Overall, the experimental group performed significantly better than the control in most of the measures taken, but this was considered only accidental for reasons that have been already explained. The control group managed to score higher than the experimental only in Grade 1, in a CWM test, namely Digit recall_Backwards. However, in the second year the experimental group outperformed the control one.

A thorough comparison of the performance of the two groups across the two grades, demonstrates that in Grade 2 the experimental group: a) strengthened the already established significant relations from Grade 1 in Nonce_Eng., Story Sequences, and YLAT_Total, b) established new significant relations in Digit
CHAPTER 8: DATA ANALYSIS

recall_Forward and Paired Associates, c) lost its initial advantage in DVIQ I_Voc. and Differences, d) maintained significant differences in Nonce_Gr. (see Table 8).

The paired samples t-tests that were conducted show that overall, the experimental group increased the mean scores in a number of tests, the most profound of which were recorded for Digit recall_Backwards (40.39%), Listening Recall (24.34%), and Paired Associates (52.43%).

As a number of interesting correlations emerged between the tasks, we conducted numerous regressions in Grade 1. These demonstrated that, unlike PSTM, CWM (Listening Recall), the aptitude total (YLAT_Total) and associative memory (Paired Associates) predict 12.8%, 12.3%, and 9% respectively, of the variance in the scores of L1 vocabulary achievement. Also, cognitive skills such as inductive reasoning (Story Sequences) and visual perception (Differences) predict differently (14.2% and 7.1%, respectively) the learners’ performance in L1 metalinguistic knowledge tests.

Regarding the regressions that were performed in Grade 2, the findings suggest that L1 verbal intelligence plays no predictive role in overall FL vocabulary performance at this early age. Also, the aptitude total is a moderate predictor (7.4%), while PSTM (Nonce_Gr., 14%) is a stronger predictor than CWM (Listening Recall, 11.3%), with respect to overall FL vocabulary achievement. In a finer analysis of the data, i.e. when the two FL skills (comprehension and production) were examined separately, CWM (Listening Recall) was a stronger predictor (17.7%) of FL comprehension than PSTM (Nonce_Gr., 16%). Metalinguistic knowledge (DVIQ II_Met.) only moderately predicted FL comprehension (6.6%), whereas the aptitude qualities played no predictive role. In the same vein, CWM (Listening Recall) was a stronger predictor (15.8%) than PSTM (Nonce_Gr., 15%) as far as FL production is concerned. The rest of the regression analyses that were performed, showed that
CWM crucially explains the performance of learners in the rest of the other tasks, verbal and non-verbal ones.

The two-way ANOVAs as well as the MANOVAs that were conducted, examined the combined effect of school and sex on the scores of each test or sets of tests that tap the same skill. Both failed to demonstrate a gender effect with respect to EFLL.

As regards the Listening span and Recall task, the findings suggest that both groups performed in a similar manner. The majority of the students faced great difficulty in recalling the lexically unrelated words in the higher levels of the test, although in Grade 2, both groups scored slightly better.

The FL vocabulary test was taken only by the experimental group, as the control group was not exposed to FL English at the time of testing. The scores gathered, demonstrate that the informants found FL_Productive more difficult than FL_Receptive. Of the 69.4% of the students who managed to score above the upper quartile in the receptive test, only 32.7% scored equally high in the productive one. This difficulty was also reflected in the respective finishing times, which were overall longer for FL_Productive (range: 100-363 seconds) than for FL_Receptive (range: 60-198 seconds).

Finally, as far as the Response Inhibition index is concerned (SSRT), the MANOVAs and the paired samples t-tests revealed that the mean values of the SSRTs, irrespective of school, were significantly shorter in Grade 2. Also, the repeated measures ANOVAs that were performed in both groups and grades, confirmed the basic assumption of the horse-race model (Logan & Cowan, 1984), as the responses of both groups on stop-signal trials that escaped Inhibition were executed faster than average go-signal responses. The only interesting difference that emerged between the two groups was that the experimental group alone, performed a
delay strategy in Grade 2, trading thus speed for accuracy. The experimental informants marked increasing Go_RTs in order to provide more accurate responses.

Bearing in mind the four research questions that were posed at the beginning of this thesis, we will thoroughly discuss these in relation to the study findings, to see whether the initially formulated hypotheses were borne or not.
CHAPTER 9
DISCUSSION OF THE FINDINGS

9.1 Introduction

All the findings from the battery of tests will be thoroughly analysed in this chapter. First, the research questions are discussed along with the results the statistical analyses yielded to see whether our hypotheses were borne out or not. Some reflection and comparison with previous research findings also takes place.

The present study is not directly comparable to any other as it views EFLL from a cognitive and not a purely linguistic perspective. We have included a variety of tests that tap a number of different skills which have not been examined together in earlier studies, such as PSTM, CWM, Response Inhibition, FL aptitude and L1 verbal intelligence. Therefore, our study is not directly comparable to any previous study. Although this is the case, we will look at the results obtained in other studies and contexts, bearing in mind the peculiarities of this study.

The study examines the factors of starting age and intensity of exposure in a formal FLL environment. The main issue investigated is whether an earlier starting age for studying EFL yields any cognitive benefits to Greek learners, when the learning context provides a substantial amount of input (5 hours per week, 360 hours in the course of two years) and optimal conditions for FLL to occur.

If the results indicate that EFLL has a positive cognitive impact on young learners, then this would be a very interesting finding on its own. If the early introduction of a FL can indeed sharpen and boost certain cognitive skills in young learners, then this will implicate a chain reaction of events as the earlier activation of learners’ explicit learning mechanisms may result in a faster, better and more efficient cognitive functioning and consequently learning, be this L1 acquisition that is still in progress, L2 or L3 learning. This thesis is
looking for evidence that supports a short-term advantage of an earlier start, and by extension a long-term beneficial effect too. So far, the relevant EFL research, with few exceptions only, has focused on the EFL impact on affective factors such as heightened motivation and the development of positive attitudes in young learners. However, Piaget’s prediction (1966/1974, as cited in Cole & Cole, 2001: 36), i.e. that the schooling experience (and for that matter the early FL schooling experience) may accelerate the cognitive development in children, was never before investigated.

As already mentioned, to be able to reach some firm conclusions, we first need to revisit the Research Questions and the relevant hypotheses posed at the outset of the thesis, to check whether these were confirmed or not.

9.2 Research Question One Revisited: Does EFL enhance the aptitude skills associated with young learners (Alexiou, 2005)? If yes, which are these? On the basis of the evidence provided, it was predicted that the experimental group would benefit from the EFL experience and that both their memory (paired associative memory) and analytic abilities (inductive reasoning, visual perception, classification ability) would be boosted. To be able to answer the first research question we conducted various statistical analyses such as paired and independent samples $t$-tests and two-way ANOVAs, that examined participants’ performance, within and between groups, across the two years of the study.

9.2.1 Descriptive statistics

Tables 4 and 5 in Chapter 8 show that the scores obtained in the majority of the tests from both groups increase with age, showing a clear developmental trend. Participants were overall more accurate with the repetition of Greek-sounding and English-sounding nonwords than others in previous studies (Maridaki-Kassotaki, 2002; Μασούρα, Gathercole, & Μπαμπλέκου, 2004). Also, in line with Μασούρα, Gathercole and Μπαμπλέκου (ibid), they were better at the Greek- than in the English-sounding nonword test. It seems that in their effort to support
nonword repetition, when there was a lack of fit between their L1 phonological structure and the FL one, participants resorted to their LTM lexical knowledge to find close phonological neighbours that might be available in their L1 lexicon.

The same picture emerged for the tasks that tap WM (PSTM and CWM) as well as Response Inhibition. The findings from both groups suggest a gradual development of the two EFs, which is in line with relevant studies mentioned earlier in the thesis. In Grade 2 both groups performed significantly better in the stop-signal task that taps Response Inhibition, confirming previous research (Brocki & Bohlin, 2010). They lowered the mean scores of the SSRT, the critical variable of this task, showing in this way that the second year they were better able to exercise firmer control over their response inhibition mechanism.

Both groups scored the same or slightly worse the second year in the two tests that examine their metalinguistic knowledge in the L1 (DVIQ I_Met., DVIQ II_Met.). In addition, the means obtained from each group in Grade 1 for the two subtests of the standardised test of DVIQ I, i.e. DVIQ I_Voc. (experimental 25.11, control 24.23) and DVIQ I_Met. (experimental 22.91, control 21.15), are well beyond the means reported in 6- to 6.5-year-old-children with normal development (15.3 and 17, respectively) (Tsimipli, personal communication).

The mean scores of the experimental group in the cognitive tasks of Memory (Grade 1: 9.17; Grade 2: 11.09), Classification (Grade 1: 35.28; Grade 2: 42.88), Differences (Grade 1: 10.69; Grade 2: 12.77), Paired Associates (Grade 1: 6.79; Grade 2: 10.35), and Semantic Integration (Grade 1: 12.99; Grade 2: 13.97) were much higher than those reported in Alexiou (2005: 181), while with Story sequences the reverse picture emerged (Grade 1: 3.95; Grade 2: 4.45). In Alexiou (ibid), though, the mean scores reported were those obtained from an age-group of 5- to 9-year olds as a whole. In any case, the fact that the second year the experimental group scored significantly better in the aptitude total score (the aggregate
cognitive score) is suggestive of the dynamic nature of FL aptitude that develops over time (Alexiou, 2005). The results suggest that it is a set of cognitive skills that are affected by experience and develops spontaneously during LL.

9.2.2 The independent and paired samples t-tests and the two-way ANOVAS

To be able to answer the first research question, we first tested the null hypothesis by performing a number of independent samples t-tests. In this way, we set a point of reference for the experimental group before the two-year-FL intervention. The findings of the t-tests established a number of differences in the performance of the two schools in the beginning of Grade 1. To their great majority, they favoured the experimental group. The finding was considered accidental as all participants were local residents of the same neighbourhood and shared the same socio-economic background, while the experimental school applied no special exclusionary criteria upon first grade enrollment.

More specifically, with the exception of the higher scores obtained by the control group in the SSRT task (Inhibition index) (in both grades), the Digit recall_Backwards (in Grade 1) and the DVIQ I_Voc. (in Grade 2), the experimental group fared overall better in 8 out of the 16 measures taken, both in Grades 1 and 2. Thus, the null hypothesis was rejected which is not an unusual thing to happen (Taylor & Lafayette, 2010).

Table 6 in Chapter 8 shows that in Grade 1 the experimental group fared significantly better than the control in 8 out of the 16 tasks taken: a) Nonce_Gr. and Nonce_Eng that both tap PSTM, b) DVIQ I_Voc., DVIQ I_Met., and DVIQ II_Met. which all examine L1 verbal intelligence, and c) Differences, Story Sequences, and YLAT_Total which respectively index visual perception, inductive reasoning and the aptitude total. The performance of the two groups was no different in: a) Digit recall_Forward (PSTM), b) Listening Recall and SSRT (the two EFs that respectively tap CWM and Inhibition), c) Memory, Classification, Paired
CHAPTER 9: DISCUSSION OF THE FINDINGS

Associates, and Semantic Integration (all cognitive skills). The control group outscored the experimental group only in one measure that taps CWM, namely in Digit recall_Backwards.

The same pairs of \( t \)-tests were conducted towards the end of Grade 2, after the two-year intensive FL intervention in the experimental group. Table 7 in Chapter 8 shows that it fared significantly higher in all three PSTM measures, outscored the control group in the two L1 metalinguistic knowledge measures, and recorded significantly higher scores in three cognitive tasks, namely Paired Associates, Story Sequences, and YLAT_Total. No significant differences were found between the two groups for Digit recall_Backwards, Listening Recall, SSRT, DVIQ_I_Voc., Memory, Classification, Differences, and Semantic Integration.

The paired samples \( t \)-tests conducted for each group separately, show that in Grade 2 both schools fared significantly better in 12 out of the 16 measures taken. However, the sharpest increase was recorded in some of the mean scores coming from the experimental group. In particular, in Grade 2 the experimental group scored very high in the two CWM tasks, namely Digit recall_Backwards (DRB) and Listening Recall (LR). The difference between the mean scores of DRB across the two years reached the 40.39%, while that of LR was 24.34% higher. One could argue that this is natural as children of this age group (6- to 8-year-olds) are in a state of transition on their path of cognitive development, moving from one Piagetian stage (the pre-operational) to the next (the concrete operational stage). This is true and partly supported by the respective increase of the control group scores in the same tasks. In particular, the second year it increased the mean scores of DRB by 18.26% and of LR by 41.44% (which however failed to reach significance in the independent samples \( t \)-tests). Thus, it is clear that cognitive maturity alone cannot account for the major differences that emerged in the performance of the two groups in these tasks.

The case of the DRB task is rather special and deserves some discussion because it was the only task the control group was better at in Grade 1. The fact that the experimental group
covered much of the lost distance from Grade 1 in a CWM task, by increasing the mean score of the test by more than 40%, is truly significant. The difference in the mean scores reached statistical significance \((p = .000)\), as the paired samples \(t\)-test and the two-way ANOVA \([F(1,188) = 4.634; p = .033]\) in Chapter 8 show (see also Appendices 4 and 5, Chapter 8).

A further comparison of the independent samples \(t\)-tests across the two years and the two-way ANOVAs shows that the experimental group maintained its initial advantage over the control in Nonce_Gr. (PSTM) as well as DVIQ_I_Met. and DVIQ_II_Met. (metalinguistic knowledge in the L1). This last finding confirms previous others (see Bialystok, 2006; Mihaljevic Djigunovic & Krevelj, 2009) which suggest that the early mastery of more than one language enhances children’s metalinguistic awareness, i.e. their understanding of how language is structured and how it works, as they gradually adopt a more analytic view towards language and LL and become more sensitive to the formal aspects of language. However, the experimental group lost their advantage in Differences (visual perception) and DVIQ_I_Voc., which can perhaps be explained by the additional load the FL places upon young learners (Bialystok & Feng, 2009). Although the experimental group slightly increased the mean score of DVIQ_I_Voc. (by 1.55%) in Grade 2, the control group which had no such additional load scored even higher (6.07%) (see Appendix 2, Chapter 8).

What is perhaps the most important finding of all is the fact that the experimental group strengthened the already established significant relations in Nonce_Eng. (PSTM), Story Sequences (inductive reasoning) and YLAT_Total (the aptitude total), while it managed to establish new ones for Digit recall_Forward (PSTM) and Paired Associates (associative memory), which was the most marked difference of all because of its magnitude. In Paired Associates, the experimental school increased the mean scores by 52.43%. The difference in the mean scores across the two grades reached statistical significance \((p = .000)\), as the paired samples \(t\)-test and the two-way ANOVA \([F(1,188) = 3.948; p = .048]\) in Chapter 8 show (see
also Appendices 6 and 7). The difference is more impressive if one compares it to the one attested by the control group, i.e. 27.02%.

Paired Associates is a test that indexes participants’ ability to form new links in mind and retain sign pairs between unrelated things, in our case shapes with pictures. The finding is extremely significant if one considers that FLL, especially in the earliest stages, during which the learner needs to ‘build’ a new vocabulary with the new FL ‘building blocks’, is an associative process. By the age of 6, learners have established a rich conceptual map in their L1 that will be further enriched by the FL experience. Novel bonds will need to be formed between the L1 referents in the learners’ LTM and the new FL words that have different phonological realisations (Papagno & Vallar, 1992). The IP paradigm assumes that when surrounding information is attended to (via the central executive), it then becomes intake to be further processed. Once this stage is ensured, then the ability to form links in memory can only facilitate the retention of FL vocabulary (Alexiou, 2005). The finding is in line with Alexiou (ibid), who found that 7-year-olds enjoy an advantage at memory and phonetic skills, when compared to the age group of 5- to 9-year-olds examined.

Story Sequences tests children’s inductive reasoning ability and rational thought with the help of situational clues, an activity which requires the perceptual and conceptual skills of participants. The fact that the experimental group strengthened the already established difference with the control group in Grade 2 indicates that the early FL exposure contributed to the further boosting of this skill. The finding corroborates Alexiou’s (ibid) ones and demonstrates that young learners display analytic abilities much earlier than Piaget previously suggested. It is reminded that the test requires participants to put a number of jumbled pictures in the correct order, for these to make sense. Considering the young age of the participants, this makes it even more difficult. The task is demanding in terms of cognitive resources, especially for such young learners, as they have to process the information coming
from the pictures to make up a story that is in line with the information (‘scripts’) that is already stored in their LTM. The whole task follows an analysis by synthesis procedure that is rather demanding and presupposes rational thinking. A very important finding on its own, as the development of higher-order processes such as inductive reasoning crucially determines further success in a wide range of everyday routines that belong to domains which are not limited to the linguistic one (Csapó & Nikolov, 2009).

The first hypothesis is not fully and clearly supported. Truly strange remains the fact that no significant differences emerged in the independent samples $t$-tests and the two-way ANOVAs, as two of the cognitive skills that were expected to be positively affected by EFLL were left unaffected. These are the ability to classify and categorise pictures along a dimension (groups of things), which also tested learners’ ability to establish a connection between the colour and the group this represents, as well as their visual perception and their ability to detect differences in the information given. This is rather peculiar for two reasons. First, during the two years of the study the experimental group was exposed to vocabulary that was structured along a number of familiar thematic fields: the family, friends, the school, clothes, body parts, etc. In theory, this experience should have further exercised their classification ability which is in the process of emergence in their L1, now that they move from Piaget’s pre-operational to the concrete operational stage. Second, the close relations established between both Classification and the Paired Associates task and the YLAT_Total (see sections 8.6.1, 8.6.3, and 8.6.4), suggest otherwise. Had the experimental group also been literate in English, these abilities might have developed more as receiving the same kind of input through many different channels (the ear and the eye) would probably prove to be more helpful to this end. This is only a speculation, though.

The findings that emerged from both the independent and the paired samples $t$-tests and the two-way ANOVAS answer Research Question One and confirm only partially the initially
CHAPTER 9: DISCUSSION OF THE FINDINGS

formulated hypothesis. By the end of Grade 2, after two years of intensive exposure to the FL, the experimental group fared significantly higher than the control in three out of the five cognitive tests predicted, namely Paired Associates, Story Sequences and the Aptitude Total. It is really peculiar that their classification ability or their ability to spot differences were not affected by their early FL experience. However, this was partly attributed to the fact that the teaching methodology followed focuses on the development of oral skills in children.

To conclude, the results obtained so far led credence to the idea that early FL performance is directly related to the enhancement of certain aptitude components, namely memory, analytic and phonetic skills. This may serve as a powerful argument in the relevant discussion of whether learners should be exposed to FLL from early on.

9.3 Research Question Two Revisited: Is PSTM the sole critical factor of FL vocabulary acquisition at this early stage of FLL? The formulated hypothesis was that CWM, i.e. the central attentional controller of WM, will also play a significant role in FL performance that will be more evident in Grade 2.

The cognitive constructs that are held to be important predictors of success in LL as well as L2 learning are FL aptitude (in more advanced stages according to Carroll, 1990) and WM, in particular the phonological loop component of WM (Baddeley, Gathercole, & Papagno, 1998; for a review see Sawyer & Ranta, 2001). The FL process, though, is quite different from L1 acquisition because it requires a conscious and directed effort on the part of FL learners before language patterns become automatic (Randall, 2007). On this ground, it was hypothesised that the participants’ PSTM would take the lead in Grade 1 while their central executive would be a stronger predictor of FL performance in Grade 2.

To further explore the predictive role of variables other than the often suggested PSTM, we first defined the relations between the principal measures by means of correlation analyses. Pearson correlation coefficients were obtained for both groups and grades. Then, we
performed various regression analyses, the implementation of which determined the direction of causality between the investigated variables. As the experimental group’s FL performance was only tested in Grade 2, in Grade 1 we examined whether PSTM was the only predictor of the learners’ L1 verbal intelligence.

9.3.1 Correlation analyses: Grade 1

What becomes clear from the correlation matrices of the two groups in Grade 1 (Tables 10 & 11 and Appendix 3, Chapter 8), is that significant intercorrelations emerged between the tests that tap the same skill. The strong internal relations that were established among the PSTM tasks indicate their internal validity and are in line with previous findings (Adams & Gathercole, 1996; Gathercole & Adams, 1994; Gathercole, Pickering, Ambridge, & Wearing, 2004; Μασούρα, Gathercole, & Μπαμπλέκου, 2004).

Apart from the intercorrelations established, external relations (at the levels of .05, .01, and .001 significance) also emerged in both group data between PSTM measures and the rest of the other measures that tap CWM, cognitive skills or L1 verbal intelligence, confirming previous studies on the predictive role of PSTM in scholastic achievement (Alloway & Alloway, 2010).

More specifically, in the experimental data Nonce_Eng. was correlated with Semantic Integration ($r = .33$) while Digit recall_Forward was correlated in both grades with a CWM measure, i.e. Digit recall_Backwards. This finding reflects the dependence of both verbal and non-verbal tasks on WM, PSTM and CWM, and confirms previous studies (Gathercole, Pickering, Ambridge, & Wearing, 2004; Gathercole, Pickering, Knight, & Stegmann, 2004; Lehto, 1996; St. Clair-Thompson, 2010). Alloway et al. (2004) argue that the close link between the phonological loop and the executive-attention resources reflects that verbal complex span tasks require both storage and processing functions for their execution. The involvement of the two processes, St. Clair-Thompson (2010) claims, is more evident in
Chapter 9: Discussion of the Findings

Children than adults, as the former find the transposition of order in the Digit recall_Backwards task a more attentionally demanding process. As regards the control data, PSTM measures were correlated with L1 verbal intelligence (Nonce_Eng. with DVIQ I_Met., \( r = .35 \)), reflecting a close link between PSTM and metalinguistic knowledge.

As far as the two CWM measures are concerned, these were highly intercorrelated in both grades but, quite strangely, only in the experimental data. The two correlation coefficients of Grade 1 and 2 (\( r = .36 \) and \( r = .32 \)) fall very close to the coefficient (\( r = .34 \)) found in Gathercole and Pickering (2000b).

In addition, CWM established external relations with other verbal and non-verbal variables. In the experimental data Digit recall_Backwards was correlated both with Paired Associates (\( r = .29 \)) and YLAT_Total (\( r = .39 \)) while Listening Recall exhibited the highest degrees of association with both Digit recall_Forward (\( r = .40 \)) and DVIQ I_Voc. (\( r = .38 \)).

The above-mentioned associations may suggest two things. First, in their effort to better execute the non-verbal tasks (e.g. in the display of pictures), learners recoded the visual stimulus into a phonological code (Brocki & Bohlin, 2010; Luciana & Nelson, 1998). It is possible that in the paired associates task they silently repeated the picture-shape pair to keep the association alive in their WM longer. In this sense, associative memory could be regarded as an analytic component, as children use memorising techniques that themselves require some degree of analysis (Alexiou, 2005). Second, to perform better in the task of Digit recall_Forward learners seem to have rehearsed subvocally (by repeating the digits) to refresh the contents of their PSTM.

In the control data similar correlations were established. Digit recall_Backwards was associated with DVIQ I_Met. (\( r = .31 \)). The correlation was smaller in magnitude than the ones reported by Gathercole et al. (1992) and Μαυρά, Gathercole and Μπαμπλέκου (2004), probably because the authors examined the relation between CWM and a pure L1 vocabulary.
measure and not metalinguistic knowledge as such (i.e. concepts that denote time sequence, either-or relations, spatial relations, synonymity), which is the case in this study. Still, the strong association found, indicates that the central executive facilitates the execution of L1 verbal IQ tasks. In addition, Listening Recall was correlated with Digit recall_Forward \((r = .30)\), indicating the learners’ dependence on their PSTM for the storage of the lexically unrelated word during the execution of a CWM task, confirming in this way previous findings (Gathercole, Pickering, Knight, & Stegmann, 2004).

As expected, L1 verbal intelligence measures shared strong associations with one another with DVIQ I_Met. correlating with DVIQ II_Met. in both schools and grades. More important, though, were the strong correlations that emerged in the control data between DVIQ I_Voc. and both DVIQ I_Met. \((r = .38)\) and DVIQ II_Met. \((r = .40)\), which indicate that vocabulary knowledge can serve as the vehicle towards metalinguistic knowledge. Regarding the external relations of the L1 verbal intelligence measures, in the experimental data, DVIQ I_Voc. was correlated with Differences \((r = .33)\), Paired Associates \((r = .33)\) and YLAT_Total \((r = .38)\). The associations indicate that vocabulary acquisition is facilitated by one’s visual perception and ability to detect differences in all surrounding information, the ability to form pairs in mind and one’s general language aptitude. DVIQ I_Met. was correlated with Differences \((r = .30)\), while DVIQ II_Met. with Story Sequences \((r = .40)\) and YLAT_Total \((r = .32)\). It seems that the cognitive skills of inductive reasoning, and the ability to spot differences in the information given or the aptitude total support the acquisition of metalinguistic knowledge. Similar were the associations that emerged in the control data, as DVIQ I_Voc. was correlated with YLAT_Total \((r = .34)\), DVIQ I_Met. with both Semantic Integration \((r = .34)\) and YLAT_Total \((r = .29)\), and DVIQ II_Met. with both Paired Associates \((r = .44)\) and YLAT_Total \((r = .31)\).
Cognitive skills related with one another to a weaker or stronger extent. In the experimental data Classification and Paired Associates were both correlated with YLAT_Total ($r = .80$ and $r = .37$, respectively). This is an interesting finding, as both tasks require storage and processing, as already mentioned, for their better execution. Similarly, in the control data, Classification was correlated with both Paired Associates ($r = .28$) and YLAT_Total ($r = .91$). Differences (D), Paired Associates (PA) and Story Sequences (SS) were all correlated with YLAT_Total ($r = .29$ (D); $r = .38$ (PA); $r = .30$ (SS)). Semantic Integration was correlated with Story Sequences ($r = .41$) and YLAT_Total ($r = .30$).

Rote Memory (i.e. the immediate recall of previously shown pictures) was the one measure that failed to establish any relations in Grade 1 with the other skills, cognitive and verbal, in both group data. This is in accord with previous findings (Alexiou, 2005). The task is the least cognitively demanding one of all the aptitude qualities tested. The finding may suggest two things. First, that young learners find the activity the least challenging of all and thus boring that asks for only a minimum of their cognitive resources, when compared, for instance, to the Paired Associates task. Secondly, it seems that they are more than good mimics and seem capable (even from this age) of engaging much more in analytic thinking and top-down processes than was previously suggested (Harley & Hart, 1997). This is interesting, as rational thought presupposes the execution of more cognitively demanding processes. Third, the finding may indicate that children as young as these rely more on their PSTM by generating phonological codes for visual items than on their visuo-spatial sketchpad to support immediate memory for visual material (Gathercole, 1998). The strong associations that were previously found between CWM tasks and PSTM support this interpretation.

### 9.3.2 Correlation analyses: Grade 2

Apart from the established intercorrelations among the tests that tap the same skill that were already mentioned, what becomes clear from the correlation matrices of the two schools in
CHAPTER 9: DISCUSSION OF THE FINDINGS

Grade 2 (Tables 12 & 13 and Appendix 3, Chapter 8), is that a number of similar to Grade 1 external relations emerged between the variables under investigation.

In the experimental data PSTM measures were correlated with the rest of the other variables. Nonce_Gr. was correlated with DVIQ I_Voc. \((r = .29)\). Gathercole et al. (1992) argue that the link between PSTM and L1 vocabulary is strong up to the age of 5. The close relationship found in this study in Grade 2 (~7 years of age) between the two skills is more in line with the claim of Service (1992), according to which, this association may still be evident in middle childhood years in the case of FLL. As regards the control data, Digit recall_F was correlated with both Digit recall_B and DVIQ I_Met. \((r = .32)\). Nonce_Gr. shared strong associations with both DVIQ I_Met. \((r = .42)\) and DVIQ II_Met. \((r = .48)\), and a weaker one with Paired Associates \((r = .32)\). Finally, Nonce_Eng. was more strongly associated with DVIQ I_Met. \((r = .42)\) than it was with DVIQ II_Met. \((r = .34)\) and Paired Associates \((r = .29)\).

As regards CWM measures and the experimental data, similar links to the first year associations emerged between Listening Recall with both Nonce_Gr. \((r = .36)\) and Nonce_Eng. \((r = .34)\), and between Digit recall_B and Nonce_Eng. \((r = .29)\). This last finding falls very close to the correlation coefficient \((r = .31)\) of Gathercole, Brown, and Pickering (2003), who examined the performance of 4- to 5-year-olds in these measures. Listening Recall was also correlated with verbal tasks such as DVIQ I_Voc. \((r = .34)\) and DVIQ II_Met. \((r = .38)\) and with cognitive tasks such as Memory \((r = .32)\), Semantic Integration \((r = .44)\), and YLAT_Total \((r = .29)\). Digit recall_B shared relations of a higher magnitude with DVIQ I_Met. \((r = .45)\) and YLAT_Total \((r = .42)\) and a weaker one with Differences \((r = .31)\). The relation between Digit recall_B and YLAT_Total, though, is stronger than the one \((r = .34)\) observed by Sáfár and Kormos (2008). In the control data the picture is almost identical. The second year a stronger relation emerged between
Chapter 9: Discussion of the Findings

Listening Recall and Digit recall Forward \((r = .36)\). In addition, Listening Recall was correlated with Nonce_Gr. \((r = .36)\), DVIQ I_Met. \((r = .31)\), Differences \((r = .29)\) and Paired Associates \((r = .28)\). Digit recall Backwards was correlated with Nonce_Gr. \((r = .37)\), Nonce_Eng. \((r = .36)\), Paired Associates \((r = .29)\) and Semantic Integration \((r = .33)\). Alexiou (2005: 206, 2009) argues that recognition memory (Semantic Integration) is especially important during FLL as it may facilitate the acquisition of features such as word endings.

With regard to L1 verbal intelligence, in the experimental data DVIQ I_Voc. was correlated with Differences \((r = .50)\) and Semantic Integration \((r = .38)\) while DVIQ I_Met. was associated with Differences \((r = .34)\) and YLAT_Total \((r = .32)\). In the control data DVIQ I_Met. was correlated with Memory \((r = .30)\) and Paired Associates \((r = .51)\), while DVIQ II_Met. shared correlations with Semantic Integration \((r = .31)\) and YLAT_Total \((r = .29)\). Paired Associates yielded the strongest association of all with DVIQ I_Met. The results corroborate previous findings (Kormos & Sáfár, 2008) and indicates that cognitive qualities are directly linked to L1 achievement and to the emergence of metalinguistic knowledge.

As far as cognitive skills are concerned, in the experimental data Memory and Differences were correlated with Semantic Integration \((r = .35\) and \(r = .33\), respectively), while Classification (C), Differences (D), and Paired Associates (PA) were all significantly associated with YLAT_Total \((r = .72\) (C); \(r = .52\) (D); \(r = .43\) (PA)). In the control data Classification (C), Paired Associates (PA), and Semantic Integration (SI) were all correlated with YLAT_Total \((r = .67\) (C), \(r = .33\) (PA); \(r = .44\) (SI)).

So far, the findings suggest that CWM is involved in L1 achievement, while it plays a critical role in the execution of non-verbal tasks too. To perform well in cognitive tasks, participants made use of their central attentional controller, which facilitated and supported the cognitive processes of noticing, spotting the differences, engaging in inductive reasoning, solving problems. This is a truly important finding because it suggests that a common pool of
cognitive resources subserves language processing. These cognitive skills may transfer between the L1 that is still in the process of being acquired and the additional FL that is being learned. Irrespective of the language children operate in, they are able to process information, engage in rational thought and eventually learn, using their central underlying ‘engine’. Kecskes and Papp (2003) speak of a Common Underlying Conceptual Base (CUCB) which is formed only when the learners’ L2 proficiency reaches a necessary threshold. Once this is established, skills acquired through one language may be used by the other. This, however, takes time and develops with the L2 and for that matter, the L1 proficiency (De Bot, 2004).

Seen under this perspective, the result is more than encouraging.

9.3.2.1 The correlations of the FL vocabulary tests: Grade 2

As it is clear from Table 12 in Chapter 8, the same highly significant intercorrelations were found between measures that examine the same skill. In addition, significant were the internal relations that were established in the experimental data between the FL_Total (aggregate score), FL_Receptive, and FL_productive, with coefficients ranging from .66, to .96, indicating the internal validity of the tests.

The FL_Total was correlated with Nonce_Gr. \((r = .40)\), confirming previous findings which demonstrate the relationship between the phonological loop and FL vocabulary acquisition (Alexiou, 2005; Masoura & Gathercole, 1999; Μασούρα, Gathercole, & Μπαμπλέκου, 2004). The FL aggregate score was also associated with Listening Recall \((r = .36)\), indicating the direct relation between CWM and FL vocabulary acquisition and confirming previous research (St. Clair-Thompson & Gathercole, 2006). It was also correlated with YLAT_Total \((r = .31)\) with a weaker correlation coefficient than the ones found a) in Alexiou (2005), where the correlation coefficient was .66, and b) in other studies where the coefficients ranged from .40 to .60 (see reviews in Carroll, 1981; Dörneyi & Skehan, 2003; Sawyer & Ranta, 2001). This is not without explanation. In Alexiou (2005) the coefficient
covered a wider age spectrum, i.e. from 5- to 9-year-olds, while the coefficients reported in the other studies concerned adult data.

FL_Receptive was correlated with both Nonce_Gr. \((r = .42)\) and Nonce_Eng. \((r = .33)\) as well as with a CWM task, i.e. Listening Recall \((r = .44)\). This last finding is suggestive of the direct involvement of CWM for the regulation of attention during FL comprehension (Kormos & Sáfár, 2008). FL_Receptive was also correlated with metalinguistic knowledge in the L1, i.e. DVIQ I_Met. \((r = .29)\) and DVIQ II_Met. \((r = .29)\), suggesting that the already existing knowledge of how a language works, facilitates the learning of an additional one (Mihaljevic Djigunovic, 2010). On these grounds, Mihaljevic Djigunovic (ibid) hypothesised that an earlier start and thus a longer exposure to the L2 might, in the long run, activate the transfer of what Cummins (1980) called cognitive/academic language knowledge between the two languages, which may work in both ways. Goorhuis-Brouwer and De Bot (2010) found stronger correlation coefficients between Dutch and English comprehension \((r = .49)\) as well as between Dutch comprehension and English production \((r = .41)\) after the first year of FL exposure. Their data came from a partial immersion setting, where English was taught for 1-4 hours per week, usually by bilingual teachers with a certificate for Dutch primary education. The authors, though, report a non-linear FL development over time, which goes fast the first year to slow down in the second.

Finally, FL production was correlated with both Nonce_Gr. \((r = .41)\) and Listening Recall \((r = .42)\), indicating that both PSTM and CWM are at work when the child attempts to produce language in the FL. This is natural as this task poses very high attentional demands on learners at this very early stage of the FLL process and the limited linguistic resources they have at their disposal (Kormos & Sáfár, 2008). According to Service (1992), in the initial stages of FLL all new foreign words sound like foreign-sounding nonwords.
The fact that both FL comprehension and production shared strong relations with the Greek-sounding nonword repetition task, is in line with previous findings (Μασούρα, Gathercole, & Μπαμπέκου, 2004). This could be interpreted in two ways. First, that FL performance in young learners is supported by a language-general phonological loop process (Andersson, 2010; Μασούρα, Gathercole, & Μπαμπέκου, 2004). Seen under this perspective, PSTM serves as the foundation for FLL (Cheung, 1996; Μασούρα, Gathercole, & Μπαμπέκου, 2004; Sparks & Ganschow, 1991, 2001), supports memory performance and facilitates the ease with which learners acquire new lexical material (Baddeley, 2002; Gathercole & Adams, 1994; Masoura & Gathercole, 1999; Μασούρα, Gathercole, & Μπαμπέκου, 2004). As Baddeley, Gathercole, and Papagno put it (1998):

The long-term learning of the sound structures of novel, phonologically unfamiliar words depends on the availability of adequate representations of the sound pattern in the phonological loop. Thus, this appears to provide a critical input to the construction of the more permanent phonological structures that are stored in the mental lexicon. Learning of associations that require the production of familiar lexical items, on the other hand, is achieved typically either without any reliance on the phonological loop or with reduced loop support and is presumably mediated instead by the use of existing knowledge of the native language. (p. 163)

Second, a plausible explanation is that participants, by means of their central executive, resort to their LTM phonological representations of their L1, i.e. the language to which they are phonologically more sensitive to, that has better-quality or longer-lasting phonological traces and as such are consequently easier to repeat (Gathercole, 1998; Gathercole et al., 1997; Μασούρα, Gathercole, & Μπαμπέκου, 2004; Morra & Camba, 2009). This, according to Masoura and Gathercole (1999) is particularly evident in the initial stages of FLL.
Chapter 9: Discussion of the Findings

The findings are in accord with previous ones, which have established that STM and LTM enter into a reciprocal relation between the ages of 5 and 8 (Gathercole et al., 1992), where the capacity of the phonological loop promotes the learning of the new phonological patterns of the foreign words, while the already stored phonological knowledge of the L1 supplements the loop and influences the ease with which learners retain both lexical and non-lexical sequences in the phonological store.

9.3.2.1.1 Concluding remarks

Overall, performance in the FL was found to share moderate to stronger relations with PSTM, CWM, L1 verbal intelligence and the aptitude total. FL comprehension shared slightly stronger relations with CWM than it did with PSTM and was also linked with metalinguistic knowledge in the L1. This last finding did not emerge for FL production, suggesting thus a clear developmental trend. This is an important finding, which confirms our prediction of the vital role the central attentional controller plays in EFL.

Interesting were the strong relations found (all at the level of .05 significance) between PSTM, CWM, and L1 vocabulary. In particular, the relation found between Nonce_Gr. and DVIQ_I_Voc. confirms previous results (Gathercole et al., 1992; Μασούρα, Gathercole, & Μπαμπλέκου, 2004) and indicates the direct link between the phonological loop and L1 vocabulary knowledge.

The findings also indicate the vital role of CWM in the execution of more complex tasks, verbal or cognitive, that require both storage and processing processes. CWM shares strong relations with PSTM and supports the learning of native- as well as foreign-sounding nonwords. According to Gathercole and Pickering (2000b), this association between the phonological loop and the central executive could reflect the contribution of a higher-order construct to both systems that may possibly correspond to general intelligence.
The strong correlations found between CWM and L1 verbal intelligence indicate the contribution of the former to the development of the latter. Finally, the findings suggest that children heavily rely on their central executive to perform well on complex cognitive tasks (e.g. Paired Associates, Story Sequences, Semantic Integration, and YLAT_Total) that require the best of their cognitive resources.

Also, significant correlations emerged between L1 verbal intelligence and cognitive tasks such as Differences, Semantic Integration, and the aptitude total. These associations are in line with Kormos and Sáfár (2008) and suggest that the cognitive qualities implicated in these tasks, i.e. the ability to detect differences as well as similarities in all surrounding information, the ability to recode information which expands the brain’s storage capacity, and one’s general language aptitude skill facilitate the growth and development of one’s vocabulary and metalinguistic knowledge.

9.3.3 Regression analyses

The findings indicate that PSTM and the central executive share equally strong relations with the learners’ FL performance. This is very important as performance on CWM tasks during the kindergarten years largely predict reading achievements a year later (Nevo & Breznitz, 2011).

To investigate the differential effect of WM, FL aptitude and the rest of the other variables on FL vocabulary acquisition, we carried out several regression analyses for the experimental group in Grade 2. However, we will briefly mention the most important findings that come from the regression analyses of the experimental group in Grade 1. Interestingly so, unlike PSTM, CWM (Listening Recall), the cognitive total (YLAT_Total) and associative memory (Paired Associates) predict 12.8%, 12.3% and 9% respectively, of the variance in the scores of L1 vocabulary achievement. Various cognitive skills such as inductive reasoning (Story Sequences) and visual perception (Differences) predict differently (14.2% and 7.1%
respectively) the learners’ performance in L1 metalinguistic knowledge tests. This is a truly important finding as inductive reasoning was found to predict a large portion of the variance (12%) in the scores of reading tests at the age of 12 (Csapó & Nikolov, 2009). The findings so far support the strong relations found in the correlation analyses conducted and further explain the predictive roles of CWM, the aptitude total and associative memory to L1 vocabulary achievement as well as those between visual perception and inductive reasoning with metalinguistic knowledge.

In Grade 2, L1 verbal intelligence plays no predictive role while the aptitude total is a moderate predictor (7.4%) of the overall FL performance at this early age. This is not surprising as LL aptitude is assumed to play a more crucial role in adult FLL (DeKeyser, 2000; Harley & Hart, 1997) and only a small one in child FLL (Abrahamsson & Hyltenstam, 2008). Carroll (1990: 24-5) himself had predicted that FL aptitude abilities “would possibly be relevant in the later stages of foreign language attainment”, while Skehan (1989) argued that aptitude might be more important in naturalistic SLA than in an instructional setting, a view also adopted by Krashen (1981a), who questioned the relevance of the traditional concept of language aptitude, as this was conceived by Carroll, in the communicative classroom. Csapó and Nikolov (2009) as well as Kiss and Nikolov (2005) report similar but of a higher magnitude findings, where the cognitive total explained 22% of the FL performance in both cases. This was due to the older age of their target groups (6th-8th graders) and their consequent advanced cognitive maturity.

More importantly, PSTM (Nonce_Gr., 14%) is a stronger predictor than CWM (Listening Recall, 11.3%) of the overall FL vocabulary acquisition performance. To perform well in the overall FL vocabulary test, the learners seem to rely heavily on both the phonological loop for the temporary storage of the new phonological representations of the FL words and on their central executive for the coordination and execution of the multiple
processing required during the FLL process. The findings are generally in line with previous ones. Andersson (2010) found an overall significant contribution of the two constructs (27%) to FL comprehension in 9- to 10-year-olds Swedish learners of EFL. A similar significant level of prediction (19%) was also found between nonword repetition scores and either L1 vocabulary acquisition (Gathercole & Adams, 1994) or FL vocabulary learning (Service, 1992). Kormos and Sáfár (2008) found an even stronger association (30.25%) between CWM and FL proficiency, but with 15- to 16-year-olds.

To conclude, the results support the hierarchical nature of cognitive abilities, as this was proposed by Robinson (2002a, 2005). The findings suggest that WM (PSTM and CWM) as well as inductive reasoning are cognitive qualities of primary importance and can better describe (when compared to the traditional aptitude components) the abilities students need to draw on, when studying in FL communicative classrooms. Also, the findings are in accord with Sáfár and Kormos (2008), who found that WM is a key underlying cognitive variable that affects both language aptitude and LL success.

When the two FL skills (comprehension and production) were examined separately, CWM (Listening Recall) was a stronger predictor (17.7%) of FL comprehension than PSTM (Nonce_Gr., 16%). Metalinguistic knowledge (DVIQ II_Met.) only moderately predicted FL comprehension (6.6%), whereas the aptitude qualities played no predictive role. In the same vein, CWM (Listening Recall) was a stronger predictor (15.8%) than PSTM (Nonce_Gr., 15%) with respect to FL production, suggesting that speaking in the FL poses very high attentional demands on learners (Kormos & Sáfár, 2008). This is reasonable, as attention is at the core of noticing (Sawyer & Ranta, 2001). The encoding of new pieces of information and of regularities in LTM constitutes the basic mechanism responsible for learning words and rules of FL grammar. Since the participating learners of this study had only rather basic skills in English, it is reasonable to assume that they relied heavily on their WM resources, PSTM
and the central executive, to encode the incoming FL information in the loop. Nevertheless, according to the IP paradigm and the connectionist approach, with experience and practice, this reliance is expected to gradually decrease over the years, when the FL processing will become more automatic and less effortful (Andersson, 2010). However, Andersson (ibid) holds that even then, the established L2 system is still less efficient than the L1 phonological system and thus still affected by it. According to the author, the L2 system is less familiar than the L1, incomplete, segmented, which results in the slower and the less accurate processing of the FL. Hummel (2009) found an association between PSTM and FL proficiency (13%) but with non-novice L2 learners (18-30 years old), which indicates that PSTM still plays a role in later FLL. Finally, neither L1 verbal IQ nor any cognitive qualities explained FL production.

Apart from these analyses, we conducted a number of others, to further explore the predictive role of CWM in the learners’ performance with respect to PSTM, L1 verbal intelligence and FL aptitude and those of FL aptitude and PSTM to L1 achievement. A number of interesting causal relations emerged, which are perhaps worthy of note.

With regard to L1 verbal intelligence measures, the findings suggest that CWM (Listening Recall) is a stronger predictor (9.5%) than PSTM (Nonce_Gr., 6.2%) of the variance in the scores of DVIQ I_Voc., while it becomes a powerful predictor of metalinguistic knowledge: Digit recall_Backwards explains 18.5% of the variance in the scores of DVIQ I_Met. and Listening Recall predicts 12.4% of the variance in the scores of DVIQ II_Met. Certain cognitive skills such as visual perception (Differences) seem to be the vehicles for L1 vocabulary acquisition, as the task accounted for 22.9% of the variance in the scores of DVIQ I_Voc. In the same vein, it seems that metalinguistic knowledge is also subserved by FL aptitude with Differences (9.3%) and the aptitude total (8.1%) predicting performance in DVIQ I_Met. Thus, the findings suggest that CWM is a powerful predictor of
native verbal intelligence, while performance in L1 IQ verbal tests is strongly predicted by various cognitive skills.

It seems that CWM subserves a number of cognitive processes. Digit recall_Backwards predicts 16.2% of the variance in the scores of YLAT_Total, while it also moderately explains (7.6%) performance in a visual perception task (Differences). Listening Recall is a strong predictor (17.4%) of the ability to recognise information and recode it (Semantic Integration) and a moderate one (8.3%) of short-term rote memory (Memory). CWM is also at work with PSTM tasks. Digit recall_Backwards explains 9.9% of Digit recall_Forward, while Listening Recall is a slightly stronger predictor of performance in the two nonword repetition tests, Nonce_Gr (11%) and Nonce_Eng. (9.6%). The findings are in accord with those of Robinson (2002b) as well as of Sáfár and Kormos (2008). The latter found a moderate relationship (13%) between backward digit span scores and the total aptitude score both at Time 1 and at Time 2 which indicates that these two constructs are related but are not interchangeable. In the regression analyses that he performed, Robinson (2002b) found that WM was a better predictor of FLL success than the traditional construct of language aptitude.

At this point we need to open a parenthesis to discuss some of the findings that pertain to the FL vocabulary test and the Listening span and Recall task. With regard to the FL vocabulary test, our findings confirm previous studies (Lundberg & Lindgren, 2008, as cited in Nikolov & Mihaljevic Djigunovic, 2011: 99) on early FL vocabulary learning. With meaning-focused interaction, spontaneous production emerges very slowly in the first two years and substantial time is needed before learners develop creative and fluent speech with reasonable accuracy and breadth (Blondin et al., 1998). Fall-backs have been reported for both comprehension and production after the first year of the FL schooling, with first-year development being faster than the second-year one. These have been attributed to the small amount of time spent on English (1 to 4 hours per week), the different didactic approaches
adopted (Goorhuis-Brouwer & De Bot, 2010) and/or the teachers’ frequent code-switching between the native language and the FL (Lundberg, 2010). Larson-Hall (2008: 36-7) holds that in FLL contexts where exposure to the FL is limited, small amounts of input are not enough ‘to trigger the formation of a morphological, syntactic or phonological system’.

As far as the Listening span and Recall task is concerned, the chi-square analyses performed in Chapter 8 yielded no between-groups differences but only a within-group one, with respect to level and grade (see also Appendix 12, Chapter 8). Table 14 in Chapter 8 shows that the higher the level, the more difficult it was for the majority of the students to successfully recall the lexically unrelated word(s). The findings corroborate previous ones (Tsimpli & Peristeri, personal communication).

The informants’ performance can be explained on the grounds of the WM model of Baddeley and Hitch (1974), and the fragility of the system which, apart from being limited in capacity, is also highly susceptible to disruption by interference. Our findings suggest that the learners’ performance was modulated by the task difficulty. The time spent for the semantic evaluation of the sentence exceeded by far the estimated time (1.5-2 seconds) acoustic information stays alive in the phonological store before it gets forever lost. Apparently, the semantic processing of the sentence caused a major interference and did not let participants engage in subvocal rehearsal. As a consequence, they faced great difficulty in the word recall task. Nevertheless, previous findings in this thesis indicate that in Grade 1 (see section 8.6.1) and in Grade 2 (see sections 8.6.3. and 8.6.4) participants did rehearse subvocally. This corroborates previous findings which suggest the emergence of subvocal rehearsal after the age of 7 (Gathercole & Pickering, 2000a, 2000b; Gathercole, Pickering, Ambridge, & Wearing, 2004).

Considering the young age of our participants, it is probable that the demands posed on their central executive through the concurrent semantic processing of the sentences and the
word recall were by far too high. Therefore, they exhibited a trade-off relationship while handling this complex cognitive task. As the resource devoted to ongoing processing (semantic evaluation) was greater, the one dedicated to the storage of the phonological material (lexical retrieval) was less (Baddeley, 2007; Daneman & Carpenter, 1980). Conway, Kane and Engle (2003) and Kane and Engle (2002, as cited in Conway et al., 2005: 771) hold that WM span tasks which require the active maintenance of information in the face of concurrent processing and interference, involve a domain-general executive attentional control mechanism that is recruited to combat interference.

The fact that the second year both groups performed significantly better in the listening span and recall task shows a steady developmental improvement and comes to confirm previous findings which suggest a process of maturation for CWM, that begins around 4 to reach a peak at 11-15 years of age (Gathercole, 1999; Gathercole & Alloway, 2008; Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006). This has been directly related to the maturation of the frontal lobes, the principal brain region associated with CWM capacities.

Based on all the above-mentioned findings we can now answer the second research question. When the FL aggregate score was entered into the regression analyses, PSTM was the strongest predictor of overall FL performance, followed by CWM that explained an equally important proportion, and finally by FL aptitude that predicted a 7.4%. This last finding is in line with previous studies which suggest that FL aptitude may predict performance, moderately so, from the early stages of FLL (for a review see Sparks & Ganschow, 2001). In a finer analysis of the data, when the two FL skills were examined separately, CWM was a stronger predictor of performance in the FL than PSTM. This finding confirms Kormos and Sáfár (2008), who argued that PSTM and CWM play a different role in instructed FLL, with CWM playing a more decisive role than PSTM at the beginner’s level and aids explicit learning mechanisms. With beginners, WM acts as a bottleneck in the
acquisition of L2 skills while with more advanced FL learners, WM capacity allows for the storage of verbal material in WM and the acquisition of an even wider repertoire of words and expressions. In addition, the findings suggest that CWM also sub-serves the execution of several other tasks, verbal as well as cognitive in nature.

To conclude, the findings corroborate previous ones with respect to PSTM and its close link to FLL (Alexiou, 2005; Cheung, 1996), while they establish a new, interesting and more powerful relation between the central executive and FL vocabulary learning. Our hypothesis was partially confirmed as CWM was a stronger predictor of L1 vocabulary acquisition (in Grade 1) and FL comprehension and production (in Grade 2) than PSTM. The large proportion of the variance in FL performance that was left unidentified suggests that FLL is a rather multidimensional process and remains highly subject to multiple other variables that were discussed in Chapter 4.

WM is an important part of our cognitive system that has close links with the acquisition of knowledge and skills in key domains over the school years. Its central executive function is capable of performing a range of high-level activities which include the coordination of the flow of information through WM, the retrieval of information from more permanent LTM stores, the application of retrieval strategies, logical reasoning, etc. (Baddeley, 1996; Baddeley & Hitch, 1974; Gathercole, 1998). In this sense, our finding is very important in that individual differences in WM capacity have an important impact on our ability to comprehend and manipulate language, be this one’s mother tongue or a foreign language (Gathercole, ibid).

9.4 Research Question Three Revisited: Can the early introduction of the FL, implemented in an intensive manner, be associated with a firmer control of one’s response inhibition mechanism?
CHAPTER 9: DISCUSSION OF THE FINDINGS

What is clear from Table 16 (Chapter 8) is that, unlike the control group, the experimental group slowed down their Go-RTs in Grade 2. This indicates the group’s tendency to pull response to increase the probability of correct inhibits and thus ensure maximum response accuracy at the expense of response latency, which has also been witnessed in other studies (Logan & Burkell, 1986). At first sight this could threaten the assumption of the horse-race model that underlies the stop-signal task, i.e. that go and stop processes are independent and compete for the first finishing time (Logan & Cowan, 1984). Their delay strategy increased the SOA\textsubscript{mean}, yielding thus a number of shorter SSRT\textsubscript{mean} (Logan & Burkell, 1986; Logan & Cowan, 1984).

Response Inhibition and the stop-signal task demand both a fast and an accurate stop process. Participants are required to simultaneously perform multiple functions. First, they have to keep alive in their WM that the green arrow equals to a go process and the red to a stop one, and then they have to accordingly sequence their motor responses. Interestingly enough, the performance of both groups did not deteriorate in this complex task, as predicted by Luciana and Nelson (1998). What is clear from Tables 23 and 24 (Chapter 8) and the proportion of participants’ corresponding errors to Go-RT\textsubscript{mean} and SSRT\textsubscript{mean} is that the test worked well in both groups across the two years. The mean accuracy of responding in the primary task ranged between 88% and 84% in Grade 1 and between 89% and 85% in Grade 2. Both groups successfully inhibited 51%-52% of the stop-signal trials, which confirms previous results and indicates that the test managed to engage participants in a satisfactory manner (van den Wildenberg, van der Molen, & Logan, 2002). The error rates on the stop-signal trials fall relatively close to the ones observed in Bedard et al. (2002) and in Huizinga, Dolan and van der Molen (2006), which are 49% and 44% respectively.

According to Logan and Burkell (1986), delay strategies are often applied by informants who tend to wait for the stop-signal before responding. The stop signal reduces their readiness
to respond and could increase the Go-RTs when they wish to avoid false alarms (van den Wildenberg, van der Molen, & Logan, 2002). As already mentioned, in Grade 2 the experimental group traded speed in the go-task in the name of accuracy (successful inhibits) in the stop-signal one, whereas the control group displayed no such tendency. Nevertheless, the outcome of this effort was only modest, as the experimental group marginally lowered the percentage of incorrect inhibits (from 48.3% to 48.2%), while the control group marked a very small increase in the error rates (from 48.3% to 49%).

The fact that a delay, which is a meta-cognitive strategy, was adopted by the experimental group only, can perhaps be explained on the grounds of the group’s intensive exposure to an additional language for two years, during which time a gradual mapping was taking place in the learners’ minds between their already established concepts and the new labels in the FL, i.e. English. With time and sufficient exposure to the TL, the learners come to realise that the concepts they hold in mind can be expressed in two alternative ways (i.e. words), one in their mother tongue and the other in the FL. This arbitrariness in language is something that bilingual children are better at when compared to monolingual ones, as they learn very early in life that a word is only an arbitrary name for something and can be realised in many ways, depending on the language one chooses to use (Hakuta & Diaz, 1985). With respect to the stop-signal task, then, it is perhaps due to their FL experience that the experimental group marked increasing RTs to the go-signal. In other words, it is possible that the early introduction of the FL had sensitised them enough to anticipate the occasional occurrence (on 25% of the total go-signal trials) of the alternative signal, (i.e. the red arrow), upon which they had to refrain from responding. Hence, the delay recorded in their effort to increase the likelihood of successful inhibits (Band, van der Molen, & Logan, 2003). It seems that the control group, on the other hand, which had had no such FL exposure to the time of
testing, was less aware of the random display of the stop-signal, as their responses to the Go- signal suggest.

To see whether the SSRT index predicts any of the FL performance, separate z-scores were first computed and then the correlation coefficients, for reasons that have been already mentioned (see section 8.12.4, Chapter 8). In Grade 1, the SSRT correlated with DVIQ I_Voc. \((p < .05)\) in the experimental data and with Listening Recall in the control ones \((p < .05)\). Both findings make sense. An efficient and properly WM is supported by a mechanism that filters all surrounding information, to select for further processing only those stimuli that relate to the demands of the task at hand, which is the task of L1 vocabulary in this case. On the other hand, the findings confirm previous studies where the two EFs, although separable processes, often work in collaboration (Brocki & Bohlin, 2010; Lehto, 1996; Miyake et al., 2000).

In Grade 2 and in the experimental data the SSRT established moderate to strong relations with other indices that examine rote memory, recognition and recoding ability as well as L1 metalinguistic knowledge. In particular, it correlated with Memory and Semantic Integration at the level of .01 significance, while it established a more moderate relation with DVIQ I_Met. \((p < .05)\). The significant relation between the two EFs is in line with previous findings. The control data yielded similar relations between the SSRT and the other variables, which were all significant at the level of .05 significance.

Regarding the several regression analyses performed in Grade 1, no particular differences were detected when the SSRT index was entered into the equation as the independent variable. The proportion of Listening Recall (CWM index) that explained variance in the scores of performance in the L1 vocabulary test (in Grade 1) and the FL tests (in Grade 2) fell very close to those conducted in sections 8.7.2.1 - 8.7.2.3 (Chapter 8). In the additional regression analyses that were performed to explore the predictive role of the two
EFs with respect to PSTM, FL aptitude and L1 verbal intelligence in Grade 2, only one significant difference emerged. The SSRT index this time, and not Listening Recall (as was the initial finding), accounted for a considerable amount of variance (21.7%) in the scores of (rote) Memory. It is reminded that participants in this task had to focus their attention on a number of pictures to be able to correctly recall them immediately after. This is a lower-order cognitive process which poses demands only upon the storage function of their non-verbal WM. Therefore, the finding makes sense as the efficiency of the inhibition mechanism has clear consequences on the amount of material that managed to be temporarily stored for immediate recall.

Compared to CMW, the SSRT was not a good predictor of FL performance, at least at this early stage of FLL. Apparently, there is a gap between motor control and cognitive control. According to Logan and Cowan (1984), control is understood in terms of the interaction between an executive system that forms intentions and issues commands to realise them, and a subordinate system that interprets these and carries them out. The inhibition mechanism may be important in both types of control, i.e. motor control (i.e. stopping ourselves from pressing the button for the green arrow on the presentation of a stop signal) and cognitive control (suppressing a perceptually salient but otherwise irrelevant cue). However, the case may be that they entail different processes of the central executive and of the subordinate systems that issue the command (Logan & Cowan, ibid). Additionally, CWM is also involved in the stop-signal task, for the sub-goal of accuracy is particularly dependent on WM capacity. Learners had to keep the task instructions in mind (green=go process, red=stop process) and use this information to guide their immediate behavior (the pressing of the button or the withholding of the response). This has been found with other Inhibition tasks such as the Tower of Hanoi and the Tower of London which both tap planning (Lehto et al., 2003) but also load on shifting. In this sense, the stop-signal task is as multifaceted as the
CHAPTER 9: DISCUSSION OF THE FINDINGS

previously mentioned tasks. The ‘impurity’ of EF tasks has been a frequently discussed issue addressed in theoretical discussions (Lehto et al., 2003).

The MANOVAs conducted (section 8.12.8, Chapter 8) show that the SSRT\textsubscript{mean}, irrespective of group, was significantly shorter in Grade 2, confirming previous studies which report that this index gets faster throughout childhood (Bedard et al., 2002). Most importantly, though, it indicates that the school factor and, consequently, the early introduction of English as a FL did not give the experimental group any additional advantage while handling this task.

The differences between response activation and response inhibition in the two groups across the two years, were detected with the application of repeated measures ANOVAs. As predicted by the horse-race model (Logan & Cowan, 1984), the responses on stop-signal trials that escape Inhibition were executed faster than average go-signal responses. This holds for both groups and grades (Table 16, Chapter 8). The results are aligned with previous findings on the independence of the two processes (Huizinga, 2006; Logan & Burkell, 1986). It seems that even though the experimental group applied a delay strategy the second year, this did not threaten the basic assumption of the horse-race model (Logan & Cowan, 1984). To conclude, the independent samples \textit{t}-tests (section 8.12.11, Chapter 8) indicate that although shorter SSRTs were recorded by the two groups in Grade 2, these failed to reach significance. The repeated measures ANOVAs confirm this finding as a significant effect was only found for the grade factor but not for the interaction between grade and school.

Central to the horse-race model is the critical parameter of the SSRT. The paired samples \textit{t}-tests from both schools indicate a clear developmental trend with the SSRT\textsubscript{mean} being larger in 6- than 7-year-olds. The result corroborates previous research findings on the increasing speed of inhibition from 6 years of age up to young adulthood (Bedard et al., 2002; Brocki & Bohlin, 2010; Huizinga, 2006). Bedard et al. (2002) found that response execution (Go-RT) accuracy increases throughout the life span, with children 6-12 years old being
significantly less accurate than adolescents and adults. Both Go_RT and SSRT improve throughout childhood and diminish throughout adulthood, but follow different developmental trends. Faster SSRTs are recorded in children aged 6-8 and young adults 18-29 as compared to the oldest group of 60-82, while Go_RTs tend to be faster throughout childhood and then gradually decrease across the adult years.

Also, the values of this index from both group data in Grade 2 are close to the SSRT\textsubscript{mean} (289 ms) reported in previous studies in 7-year-olds (Huizinga, 2006). The range of SSRT\textsubscript{mean} in both schools and grades suggests that both groups were capable of exercising close control over their actions and of stopping the current course of action, which became more evident in Grade 2. This is in line with Bedard et al.’s finding (2002), in that selective inhibition is one of the earliest emerging EFs. However, the difference between the overall mean values of the SSRT of the two groups failed to reach significance, suggesting again that the early exposure of the experimental group to the English language did not have any additional effect on their control over their response inhibition, at least not in the first two years of their FL exposure. If an advantage for the experimental group was to emerge, this would perhaps relate to a significantly faster response on the red arrow (Martin-Rhee & Bialystok, 2008).

Overall, the findings suggest that the EFLL did not have any additional effect on the already established control of participants on their inhibition mechanism. Thus, our initially formulated hypothesis is confirmed. Nevertheless, another important finding emerged out of the analyses, i.e. the fact that only the experimental group applied a delay strategy to perform well in this task. Perhaps the FL intervention contributed to the earlier emergence of this strategy which did not threaten the independence of the two processes, namely response activation and response inhibition.

Our results corroborate previous studies on bilingual children: the bilingual advantage is lost in tasks that require control over motor responses (response inhibition) (Martin-Rhee &
Bialystok, 2008), while it is maintained in others that require control over cognitive or conceptual ones (i.e. to keep task-irrelevant but perceptual information de-activated). This ability is part of the EF that is developing gradually in preschool children to reach a peak around young adulthood (Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006).

9.5 Research Question Four Revisited: Is there a gender effect associated with EFLL?

The last research question investigates whether such an effect will emerge in Grade 2 (after the two-year FL intervention) with respect to the skills under investigation: PSTM, CWM, Inhibition, L1 verbal intelligence, and FL aptitude. It was predicted that no gender effect would be found, as it is perhaps too early in the FLL process for any differences between the two genders to come about. Kiss and Nikolov (2005) found an outperformance of girls over boys in terms of language aptitude and English language proficiency, but that was found at the age of 12.

To explore the last research question, we first applied a number of two-way ANOVAs, in both grades and examined the effect of two independent variables, school and sex, on one dependent variable, i.e. the score of each individual measure (see section 8.8, Chapter 8). Only in the Listening Recall task, a CWM task, the girls coming from both schools scored marginally better than the boys in Grade 2, without this difference reaching any significance: F(1,92) = 3.827; p = .053. The two-way ANOVAs confirmed the findings of the independent samples t-tests, i.e., the significant outperformance of the experimental group as a whole over the control, in the majority of the tests.

Then, we performed several MANOVAs to examine whether a group and/or gender effect would emerge in Grade 2 with respect to sets of tests that examine the same skill: PSTM, L1 verbal intelligence, and FL aptitude. All in all, no gender effect was found in

---

1 CWM and Inhibition were examined by means of two-way ANOVAs.
either group. Again, the MANOVAs confirmed the outperformance of the experimental group in the majority of the tests (see Appendix 9, Chapter 8).

Even though girls are generally reported to be more motivated towards FLL, this is not a sufficient condition on its own that may predetermine a better performance in various tasks. Our findings indicate that this is not so, confirming in this way various other previous studies (Alexiou, 2005; Brocki & Bohlin, 2010). Luciana and Nelson (1998), who studied the development of WM functions in 4- to 8-year-olds, also found that neither gender exhibited a clear advantage for tasks that required complex reasoning, strategy use, or planning. Only Klenberg, Korkman and Lahti-Nuuttila (2001) found a gender effect that favoured girls in some of this study’s variables (phonemic fluency, visual attention, auditory response) but the effect was significant within the age range of 3 to 5. From 6 years onwards the boys performed equally well with girls. At this early age, no gender effect has been mentioned in the relevant bibliography, at least not in the one I am aware of.

As predicted, no gender effect was associated with EFLL. It should again be noted that during these two years of English exposure, the experimental group was principally exposed to aural input and performed numerous oral activities. Perhaps the emergence of literacy could have caused the earlier emergence of differences between the two genders.

9.6 Limitations of the study

The above-mentioned findings should not be generalised but instead be taken as suggestive for several reasons that deserve some discussion. First, it is well-established that vocabulary learning is a multidetermined phenomenon. Certainly, there are other variables that could account for larger variance in the learners’ performance, which were not investigated in this study. These would include student, family (e.g. parental education, family income), classroom and school variables (resources available to children) that contribute to general school success as well as L2 success (Csapó & Nikolov, 2009; Johnstone, 2009b).
The size of the sample was relatively small. Therefore, it may not be representative of the population as a whole with respect to social status, educational advantages or parental interest. It would be interesting to see whether differences would emerge in the performance of two groups that would be of different socio-economic and family backgrounds.

The participants of this study were all monolingual speakers of Greek. Based on studies that concern bilingual children and the general positive effect of bilingualism on children’s cognitive functioning and on the learning of an additional language, it would be very interesting to compare the performance of L2 and L3 speakers (within the experimental school) against that of students who only learn English as a FL.

The nature of the FL test was in a sense limited as no reading or writing skills were examined since the study participants were not yet literate in the FL at the time the study took place.

Another limitation is the age itself. The results are valid for these particular ages and could not be generalised unless they are further trialed in older students.

Finally, the scope of the study was restricted to the two most frequently studied EFs, namely WM and Inhibition, in particular Response Inhibition. Although cognitive flexibility (the ability to shift between tasks) has been identified as another critical executive component (Miyake et al., 2000), this was not examined in this study.

9.7 Chapter Summary

This study mainly investigates whether an earlier starting age for studying EFL in the Greek context can be associated with an enhancement of young learners’ cognitive abilities, when these are formally and intensively exposed to the FL. To help the reader form a more comprehensive picture of the study findings, we will summarise these in relation to the initially posed research questions and our corresponding predictions.
The first research question examined whether EFL can be related directly to the enhancement of young learners’ aptitude components. Based on the relevant literature, it was predicted that the experimental group’s memory (paired associative memory) and analytic abilities (inductive reasoning, visual perception, classification ability) would be boosted by the early FL intervention. Our initial hypothesis was partially confirmed as the informants’ classification ability and their ability to spot differences were left unaffected. Quite the contrary, their cognitive total ability and associative memory skill as well as their inductive reasoning ability were positively and significantly affected by the two-year FL intervention. The importance of the finding lies in the fact that these cognitive operations have been found to determine the retention of FL vocabulary (Alexiou, 2005) and future LL success (Csapó & Nikolov, 2009).

Research Question Two examined whether, apart from PSTM, CWM can explain FL vocabulary acquisition at this early stage of FLL. It was hypothesised that the attentional controller of WM would play a more significant role in FL performance. This was eventually confirmed, as CWM proved to be a stronger predictor of FL comprehension and production than PSTM. This corroborates Kormos and Sáfár’s (2008) view, in that with instructed FLL, i.e. an explicit and conscious LL process, CWM plays a more decisive role than PSTM at the beginner’s level and aids explicit learning mechanisms. In addition, CWM was found to play a vital role in the execution of complex verbal as well as cognitive tasks, that require storage and processing processes. This suggests the existence of a higher-order cognitive construct which involves the central executive to combat interference, and which may possibly correspond to general intelligence (Gathercole & Pickering, 2000b).

Research Question Three examined whether the experimental group, who were formally and intensively exposed to EFL from the age of 6, would exhibit a more efficient response inhibition control than the control group. As predicted, the two groups were equally able to
exercise close control over their actions and stop the current course of action. Thus, EFLL was found to have no additional effect on the already established control of participants on their response inhibition mechanism. However, the fact that only the experimental group applied a delay strategy to perform well on the stop-signal task is important. This was explained on the grounds of the early FL intervention, which may have sensitised the group to the existence of a random occurrence of an alternative signal (i.e. the red arrow). The experimental group alone was clearly anticipating for it to occur and thus delayed their go-responses to mark more successful inhibits.

Research Question Four examined whether EFLL is associated with a gender effect regarding the investigated skills: PSTM, CWM, Response Inhibition, L1 verbal intelligence, and FL aptitude. No such effect was expected to emerge because the informants were in a very early FLL stage. At the time of testing they had been exposed to aural English for two years. As predicted, no gender effect was associated with EFLL.

In Chapter 10 that follows we will consider the implications of this study and, finally, we will provide some suggestions for further research.
10.1 The significance of the present study

The present study, by virtue of its longitudinal design, has shed valuable light on a number of issues that relate to the cognitive impact of EFLL. A sample of young learners’ data was compiled in the period of two academic years, from two groups, the experimental and the control, the statistical manipulation of which helped us explore and answer the research questions posed at the beginning of this thesis.

The study has many strong suits, one of which is the multitude of the tests performed, verbal and non-verbal, and the wealth of the data compiled. In addition, the contribution of this thesis rests on the fact that, among other things, it examined the role of PSTM in a language other than English (Cheung, 1996; Maridaki-Kassotaki, 2002; Μαριδάκη-Κασσωτάκη, 1998; Μασούρα, Gathercole, & Μπαμπέλεκου, 2004; Papagno, Valentine, & Baddeley, 1991; Service, 1992). The number of such studies is still limited.

Developmental improvements in task performance were seen on virtually every performance of the skills examined (PSTM, CWM, Response Inhibition, FL aptitude and L1 verbal intelligence) and are thus in accord with previous research findings (Alexiou, 2005; Band et al., 2000; Bedard et al., 2002; Brocki & Bohlin, 2010; Huizinga, 2006; Huizinga, Dolan, & van der Molen, 2006; Lehto et al., 2003; Luciana & Nelson, 1998; Sáfár & Kormos, 2008; van den Wildenberg & van der Molen, 2004).

The present study expands our knowledge on the variables that may determine variation in FL performance. Our findings demonstrate that the learning and use of a FL draws on a range of cognitive processes. One such process that has received considerable attention is WM, in particular PSTM, which has been implicated to be a core component of L2 aptitude (Dörneyi & Skehan, 2003; Hummel, 2009; Kormos & Sáfár, 2008; Robinson, 2005; Sawyer
& Ranta, 2001). An efficient PSTM provides an advantage in EFLL, as this study has demonstrated. This is especially important in today’s communicative classrooms because FL learners are required to make sense of large amounts of aural, authentic data that places heavy processing demands on their PSTM (Hummel & French, 2010; Robinson, 2002b). The phonological store of Baddeley and Hitch (1974) that was not measured by the MLAT, shares a lot in common with Carroll’s (1981) phonemic coding ability (Sparks & Ganschow, 2001; Wen, 2012), i.e. the ability to analyse sound in order for this to be retained (Carroll, 1965, as cited in Chan, Skehan, & Gong, 2011: 60). To this end, it is suggested that the development and integration of nonword repetition tasks in the current YLAT task would constitute a future research aptitude agenda, an issue that has already been touched by many scholars interested in aptitude batteries (Chan, Skehan, & Gong, 2011; Harley & Hart, 1997; Kiss & Nikolov, 2005; Robinson, 2001; Sawyer & Ranta, 2001; Skehan, 1989, 1998; Sparks & Ganschow, 2001). This could possibly serve as a diagnostic tool of phonological dyslexia as well, which according to Harley (2001: 191), “is just one aspect of a general impairment to phonological processing”, manifested in the performance of PSTM tasks.

Apart from the vital role PSTM plays in FL vocabulary performance, it also crucially contributes to L1 vocabulary acquisition. In both language processes, its contribution is still evident in middle childhood years. Even though the study informants relied on their PSTM for the storage of verbal as well as visual stimuli, they heavily depended on their CWM to perform well on almost the totality of the measures and consequently the skills investigated in this thesis. More importantly, in a more refined statistical analysis, when FL comprehension and production were examined separately, the central attentional controller of WM proved to be a stronger predictor than PSTM in FL vocabulary performance. CWM was also directly linked to more complex cognitive skills (e.g. associative memory, reasoning ability, recognition and recoding ability and the aptitude total) which required the best of the learners’
cognitive resources. These findings strongly support an early and intensive FL intervention and may serve as a powerful argument in favour of EFLL which, among other things, does not only facilitate the development of students’ linguistic resources but also sharpens their cognitive skills and capacities.

Our findings correspond well with previous ones as to the dynamic nature of aptitude (Alexiou, 2005). They demonstrate that cognitive skills are plastic at this early age and can be enhanced with early practice and training through adequate exposure to the FL, while they are directly linked to L1 achievement, the development of metalinguistic knowledge and the overall FL vocabulary performance. The findings suggest a beneficial effect of EFLL on children’s associative memory, analytic and inductive thought. Also, they indicate that a better top-down processing of incoming information is achieved by means of the central executive, during the execution of verbal as well as cognitive tasks.

The findings also support the often suggested idea that Piaget’s developmental stages are going down (Das Gupta & Richardson, 1995; Pinter, 2011). The study participants displayed a heightened performance in phonetic, analytic, and memory skills which have been implicated (Skehan, 2002) in the different stages of SLA (input, central and output processing) and in the cognitive processes that take place within these stages (e.g. noticing, pattern recognition, automaticity in retrieval) in the mental workspace of WM. This can be attributed to the wide variety of stimuli children receive nowadays, the formal schooling and the directed training this provides as well as to the early introduction of the FL (as is our case). Piaget himself (1966/1974, as cited in Cole & Cole, 2001: 36) had acknowledged that formal schooling could accelerate children’s cognitive development.

The vital role of the central executor has clearly been established in this thesis. CWM seems to play a far greater role than the one initially suspected: in L1 achievement, in the development of metalinguistic thought, in FL success and during the execution of non-verbal
tasks. The ‘chief executive officer’ (Kimberg, D'Esposito, & Farah, 1997), which is a domain-general system (Baddeley, 2003; Lehto, 1996) performs a range of high-level regulatory EFs, such as directing attention, planning actions, solving problems, reasoning logically (Baddeley, 1986; Kimberg, D'Esposito, & Farah, 1997). Neo-Piagetian theorists (Pascual-Leone (1970) call this resource the ‘M capacity’ (Morra & Camba, 2009), while within an IP framework this has been named as the Supervisory Attention System (Randall, 2007) or complex/verbal WM (Baddeley & Hitch, 1974). Whatever the name of this cognitive construct, its proper functioning determines, among other things, the conscious and directed attention to certain aspects of the data with important and direct consequences for (F)LL.

This is a truly important and encouraging finding, because during language processing, be this L1, L2 or L3, the common pool of cognitive resources that is formed serves equally well the processing of every language. As Mihaljevic Djigunovic (2010) suggests, an earlier start and thus a longer overall exposure to the L2 will, in the long run, activate the transfer of what Cummins (1980) called cognitive/academic language knowledge between the two languages, which may work in both ways, displaying thus a boomerang effect. Mihaljevic Djigunovic (2010) examined the interaction between L1 and L2 achievement. Her findings suggest that the longer the exposure to and the experience in the L2, the easier it is for learners to transfer language use behaviours (at the levels of reception and production) from one language to the other. She argues that early FL beginners (i.e. when EFLL starts before the age of 10), due to the longer overall period of exposure to the L2, reach the necessary competence levels in the two languages sooner to allow transfer in both directions (reflected in more refined vocabulary choice, sentence construction, text-writing skills, etc.). She cites Bartolovic (1993), who demonstrated a higher achievement of the experimental group in their L1 as well as Kecskes and Papp (2003), who talked of a Common Underlying Conceptual Base (CUCB) which makes possible the influence of the L2 on the L1, manages the two
language channels and is dependent on and develops with L2 proficiency. Once CUCB is established, skills acquired through one language may readily be used by the other. De Bot (2004) also views that only when learners reach a high enough level in each language are they able to make use of their knowledge of both languages. Cummins called this a Common Underlying Proficiency, in other words a central operating system, a common area which both languages can have access to and use. Viewed under this perspective, the results are more than encouraging.

The FL intervention was not associated with any gender effect, a finding which is in accord with previous studies (Alexiou, 2005; Luna et al., 2004). The performance of the male and female experimental informants across a number of tasks did not differ in any significant way after the two-year-exposure to the FL.

Of the two EFs studied in this thesis, the central executive of WM was associated with higher-order cognitive processes and with verbal and cognitive skills while Response Inhibition was associated only with bottom-up processes implicated in a non-verbal, less cognitively demanding task, i.e. Memory. Unlike CWM, Response Inhibition did not predict performance in the FL vocabulary test, at least not at this early stage of EFLL. Both groups performed the same in the stop-signal task and were capable of exercising firm control over their actions, which became more evident in Grade 2. However, it is very significant the fact that only the experimental group adopted a meta-cognitive strategy in this task, possibly due to their early FL exposure that sensitised them to the occasional occurrence of an alternative signal (i.e. the red arrow) to which they were not supposed to respond.

Apart from this strategy, the experimental group also rehearsed subvocally, which is a strategy that increases PSTM capacity (Gathercole & Pickering, 2000a, 2000b; Gathercole, Pickering, Ambridge, & Wearing, 2004) and also has immediate consequences on children’s performance, be this L1 vocabulary acquisition or L2 one. Assuming that the total processing
space (Pascual-Leone, 1970) remains stable in one’s life span, it is very important the fact that
the early and intensive exposure to the FL offers learners the time, the experience and the
practice they need to facilitate the earlier emergence of efficiency and automaticity in
cognitive processing, which will set free much earlier the attentional resources of WM for a
more in-depth processing of other aspects in the input, such as syntactic patterns and semantic
content (Case, Kurland and Goldberg, 1982; Hummel & French, 2010).

As far as the FL performance of the experimental group is concerned, the current study
revealed a significant difference in the informants’ combined performance in the two FL tests
that examined comprehension and production. The latter was more difficult than the former
and this difficulty was reflected in the performance of informants, who traded speed for
accuracy in their eventual output (Baddeley, 2007). This was expected, however, for a number
of reasons.

First, this is a very early stage of FLL and learners still have very limited FL linguistic
resources. Second, it is well-documented that language production emerges at a later stage
because it is a more cognitively demanding process than language comprehension. Third, FL
processing in general is slower, less automated and more effortful than L1 processing,
especially in the earliest stages (Andersson, 2010; Wattendorf & Festman, 2008).

Also, although the FL programme followed by the experimental school is intensive, it
should be kept in mind that our participants were not yet literate in the FL at the time of
testing. Early literacy (from the age of 7-8) in the FL has been reported to accelerate the
overall FL performance (Johnstone, 2009a). Sparks et al. (1997) claim that the direct teaching
of the new FL sound-symbol system and the exposure to print early in the FL course may
benefit FL students, especially those with relatively weak L1 skills.

Additionally, children as young as these, have not yet developed the full and complex
range of learning strategies (Cole & Cole, 2001). Clearly, more time is needed before any
noteworthy gains start to emerge. Webb (2007a) discusses the favourable effect of repetition in incidental vocabulary learning, where learners need to encounter unknown words as many as ten times in context for any sizable gains to occur.

From what been reported so far and from the literature review conducted earlier in this thesis, it is clear that EFLL is a time-consuming process whose valuable gains show in the long run (Muñoz, 2006, 2010). The linguistic outcome of the EFLL experience may be modest in the short-term (Nikolov, 2009b) but its cognitive impact can be great as this study has demonstrated and Nikolov and Mihaljevic Djigunovic (2006) had predicted.

While an extensive amount of research has been conducted with respect to EFLL, its cognitive impact has not been explored so far. EFLL has been investigated in relation to the linguistic impact an early exposure might have on learners over the years (García Lecumberri & Gallardo, 2003; García Mayo, 2003; García Mayo & García Lecumberri, 2003; Muñoz, 2003, 2006, 2010; Nikolov, 2009b). Larson-Hall (2008) argues that in minimal input situations, such as an instructed FL setting is, perceivable linguistic effects (phonological and morphosyntactic) emerge only after learners receive a substantial amount of input, i.e. around 1600-2000 hours distributed in the course of at least 6 years of FLL, that would amount to 6-8 hours per week for 44 weeks per year. A younger starting age makes a modest difference (3-14% of the variation in scores) to both phonological and basic morphosyntactic abilities. This means that age seems to play a non-negligible role in improving FL aptitude, provided that language learners are sufficiently exposed to the FL. One way to ensure larger amounts of language input is to start studying a FL at a younger age, i.e. from the first year of primary school.

The present study commenced while in the beginning of the experimental group’s exposure to the FL. After a two-year-period of FL exposure, which is a rather short period, the results primarily reflect not the FL competence of young learners that is still developing,
but the considerable cognitive changes that have taken place during this period. In this sense, an earlier start offers a longer overall period of FLL, during which time their cognitive growth can be positively influenced, even enhanced, when they are still at a highly developmental stage.

For all the above-mentioned reasons, the importance of this study is far reaching. Its major aim was to explore whether a connection could be established between the early teaching of a FL and a positive impact on children’s cognitive skills. The findings demonstrate that this goal was more than sufficiently met.

10.2 An early start is compatible with all FLL and child cognitive theories

An early start provides optimal learning conditions (e.g. early learning through playing) and the triggering data in the input for FLL to take place. Young learners are given the chance to engage in FL interaction that is meaning-driven (VanPatten & Williams, 2007) in a fun and pleasant way and because of this overall longer FL exposure they have plenty opportunities to notice crucial aspects in the input (Schmidt, 1990). The feedback, the assistance and the scaffolding (Wood, Bruner, & Ross, 1976) they receive during the FL interaction with their teacher, a more experienced and knowledgeable speaker of the L2 than they are, all facilitate the awakening of their ZPDs (Vygotsky, 1978) and their becoming independent users of the FL as they gradually reach higher levels of knowledge (Das Gupta & Richardson, 1995; van Geert, 1998).

In addition, as Krashen has suggested, at this early stage of FLL the affective filter in children is still low. As a consequence, young learners are more open to the input flood they receive, while their motivation is high and their attitudes towards FLL are positive.

EFLL gives young learners ample time to experiment with the new language, attend to and select the new forms for further processing to finally discover (along with the more explicit learning that will take place in the years to come) the rules of the FL. The thesis has
clearly demonstrated that the early FL start can have a boosting effect on aptitude qualities, general problem-solving skills which facilitate the earlier development of strategies in learners and the earlier activation of their explicit learning mechanisms.

Current SLA research has placed great emphasis on the role of social and environmental factors (Piaget, as cited in Rogoff, 1998: 684; Wertsch, 1985; Bruner, 1983), interaction-driven FL development (Gass & Mackey, 2007; Long, 1996; Mackey et al., 2002; Pica 1998; Swain, 1985), innate factors in the learners (Chomsky, 1976, 1986, 2001), WM (Baddeley & Hitch, 1974; Randall, 2007; Robinson 1995a, 1995b, 2001, 2002b), as well as attention and noticing (Schmidt, 2001). The thesis has demonstrated that these last two factors are very important. It is well-known that today’s children are constantly exposed to and overwhelmed by all kinds of external stimuli. Because of the limited nature of the IP capacity (McLaughlin, 1990, as cited in Ellis, 2003, p. 390), they need to employ their attentional resources to the maximum to ‘sort through’ the massive amounts of input they receive, ‘tune in’ only certain stimuli and ‘tune out’ the rest (Gass & Mackey, 2007: 186). For any FLL to occur, the new forms must be first noticed to and then be registered in STM to subsequently be transferred to LTM. In this sense, Gathercole and Alloway (2008: 23) were right in viewing WM as a ‘bottleneck for learning’.

One should not underestimate the beneficial effect of the school that can speed up children’s cognitive development (Das Gupta & Richardson, 1995; Piaget, 1966/1974, as cited in Cole & Cole, 2001: 36; Pinter, 2011) and the supporting environment and optimal conditions the experimental school provides to its pupils from very early on, with respect to the teaching of English. On the one hand, the deliberate learning and training efforts play a vital role in the expansion of children’s knowledge base and the development and further enhancement of their memory strategies and certain cognitive skills, such as logical problem-solving, memory and meta-cognition. On the other hand, through play-like activities, story-
telling, crafts and art activities, the foundations of the FL are laid in a playful, tension-free atmosphere, where the child is left free to innovate and experiment with the language while interacting with others (Bruner, 1977, 1981, 1983; Ratner & Bruner, 1977; Veraksa, 2011).

The informants of this study are in a state of cognitive flux, leaving the Pre-Operational stage to enter the Concrete Operational one, a period in one’s life that Wood (1998: 23) rightly called the age of ‘intellectual revolution’. The study informants have just entered early middle childhood (Cole & Cole, 2001), a period which witnesses the rapid development of metalinguistic thought, the ability to perform various mental operations (e.g. forming combinations, ordering, and transforming information in a logical manner), the great increase of memory abilities and the enhanced ability to store and retrieve information in a systematic way, the expansion of linguistic abilities and vocabulary and the development of effective strategies (Pinter, 2011). Cole and Cole (2001) believe that the maturation of the brain that takes places from 6 to 8 years of age plays an important role in the further development of children’s thinking that becomes more organised, flexible and rational. Vygotsky (1934, as cited in Wertsch, 1985: 184) also believed that the interesting interplay between individual cognitive processes such as perception, attention, memory and thinking, are to be held responsible for the huge quantitative and qualitative changes that are witnessed in children’s mental functioning and cognitive development during middle childhood. The findings of this thesis demonstrate that the early introduction of the FL can sharpen even further the mental functioning of children and make it work more efficiently.

The IP processing paradigm assumes that knowledge acquisition is an accumulative process while automaticity develops through practice (Atkinson & Shiffrin, 1968; Eysenck & Keane, 1995; Neves & Anderson, 1980; Randall, 2007) when declarative knowledge converts into procedural one. In its turn, the connectionist account assumes that automatic learning results from the strengthening of some neural connections due to the highly frequent exposure
to certain linguistic features or the inhibition of others (in the case of less frequent items) that all build an associative network (Eysenck & Keane, 1995). Whatever the perspective one wishes to adopt, the FLL process makes use of and benefits from the knowledge that is acquired accumulatively over the years, often in a playful manner.

Finally, the literature suggests an interesting interplay between language and cognition. Research has shown that learning an additional language raises the metalinguistic awareness in learners, which in turn helps them to be more aware of their L1 skills as well as of the LL process in general (Edelenbos, Johnstone, & Kubanek, 2006; Johnstone, 2009a, 2009b; Nikolov & Mihaljevic Djigunovic, 2006). Also, Mihaljevic Djigunovic (2010) raises the issue of bi-directionality by reporting the outperformance of early beginners (before the age of 10) on all L1 and L2 tests over later beginners. Therefore, it is very important that this study has demonstrated the EFL experience can further enhance children’s cognitive skills and their executive functioning as this enhancement, in its turn, is expected to have a beneficial effect on their overall LL process (Bialystok, 2006; Taylor & Lafayette, 2010).

10.3 Future directions of research

This research suggests a number of avenues for further investigation. The primary suggestion for future research is to build on the present study by replicating parts of it in the future (perhaps combining the quantitative with the qualitative design) to examine both the linguistic (in the L1 and the FL) and the cognitive performance of the same cohort of students, before they graduate from the experimental school at the end of Grade 6. This would allow for the examination of an even greater cumulative effect of FL study, revealing whether the effects found in the present study are maintained and/or further boosted.

Tracking the development of these children over a longer time frame is viewed necessary for an additional reason. At the time of the study the experimental group was not literate in English as reading and writing skills were not yet developed. As of Grade 3,
though, these students will learn how to read and write in English and by the time they finish primary school they will already have a 3-year-experience in CLIL. It would be interesting to see what the impact of CLIL would be on the variables investigated by this study. In addition, the present data could also be compared against those coming from a higher grade (Grade 4 or possibly 6 to allow adequate time for any potential CLIL effect to emerge) in order to investigate the potential interaction between the L1 and the FL along different language areas, such as grammar, vocabulary, pragmatics, and so on.

As a result of global travel and migration, more and more classrooms that were traditionally monolingual now have many L2 and L3 children. It would be interesting to compare the linguistic, cognitive and executive functioning of monolingual learners against those of bilingual or trilingual young learners.

Future research may also be directed towards exploring whether the same or different aptitude components are implicated in higher levels of FLL, as suggested by a number of researchers (Carroll, 1981; DeKeyser, 2000; Harley & Hart, 1997; Johnson & Newport, 1989).

The study carried out was restricted to intensive classwork conducted in an experimental school for the teaching of English as a FL. The supporting environment and the optimal conditions (qualified teachers, input flood, small size of FL classes), provided by the experimental school, might have facilitated the FLL process. As of September 2010 the Ministry of Education and Lifelong Learning runs a pilot program in 800 primary schools all over Greece where English is introduced with 2 hours per week from Grade 1. Comparing the executive functioning of learners that come from these two different FLL contexts would be the goal of a future study.

Research on EFLL is still scarce and sparse as this is a rather new practice. As a consequence, the general understanding of this process remains piecemeal. However, because
of its numerous intricacies, it deserves adequate and systematic attention. Given its ‘hybrid’ nature (EFLL stands somewhere between child L1 acquisition, child and adult SLA and FLL in general), EFLL is a unique test case of theories pertaining to the mechanism, process and outcome of L1 acquisition as well as FLL.

My sincere hope is that this study will provide enough food for further scientific thought and EFLL research because the young FL learners of this world deserve it.
REFERENCES


REFERENCES


Bartolovic, B. (1993). Young learners' cognitive abilities in learning foreign languages. In M. Vilke, & Y. Vrhovac (Eds.), Children and foreign languages (pp. 27-44). Zagreb: Faculty of Philosophy.


REFERENCES


García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. P. García Mayo, & M. L. García Lecumberri (Eds.), *Age and the


REFERENCES


REFERENCES


References


REFERENCES


REFERENCES


REFERENCES


